UNIVERSITY OF THE WESTERN CAPE

Faculty of Community and Health Sciences

MINI THESIS

Title: Evaluation of the data quality of routine health information system in the public health facilities in the Eastern Cape Province, South Africa

Student Name: Sibusiso Sifundo Thabethe

Student Number: 4002960

Type of Thesis: Mini-thesis

Degree: Master of Public Health

Department/School: Public Health

Supervisor: Dr. Verona Mathews

Date: November 2022

10 Keywords: Health Information, Data Quality, Routine Health Information System, District Health Information Software, Assessment, Completeness, Consistency, Accuracy, Data Elements, Health Facilities.

Declaration

I declare that the *Evaluation of the data quality of routine health information system in the public health facilities in the Eastern Cape Province, South Africa* is my own work, that it has not been submitted for any degree or examination in any other university, and that all the sources I have used or quoted have been indicated and acknowledged by complete references.

Full name: Sibusiso Sifundo Thabethe .... Date: 19/11/2022 ....

Signed:  ....

Acknowledgments

I wish to express my deepest appreciation to the people that provided support, guidance, and contribution to the completion of this study. Above all, I wish to give thanks to the All Mighty God who gave me health, strength, and wisdom because, without His Mercy and Grace, I would not have been able to successfully complete this treatise.

My sincere appreciation goes to a number of amazing people who have made this study successful through their support and encouragement:

- I am highly grateful to my Supervisor, Dr V Mathews, for her support and professional guidance. Without your consistent teaching and valuable advice, I would not have been able to complete the study.
- My classmates; Dr Simphiwe Phindile Khumalo, Mr Motlatsi Letsika, and Mr Kefiloe Pokela for your valuable support and motivation throughout the MPH journey.
- My colleagues Dr Matandela, Ms Sandisa Nkwintyi, Ms Thozama Sigudu, and Mr Mdliva with all their support and assistance as and when I needed their input.
- Finally, a special thanks go to the Eastern Cape Department of Health for allowing me to conduct the study, and granting access to their data. I also would like to thank the ECDoH Epidemiology and Research Team, and HIMS Team for assisting me.

Abstract

Background: The significant role of health information has over the years improved, and many countries adopted routine health information systems to effectively and efficiently manage data collection. The information system helps to improve the quality of health information and enables healthcare workers and authorities to take informed decisions and appropriately monitor and evaluate the implemented healthcare programs. However, the evidence shows that routinely collected health information is often underutilized due to poor quality, especially in Low- and Middle-income Countries.

Objective: To assess the quality of data reported in the routine health information system from health facilities of the Eastern Cape Department of Health.

Method: A retrospective descriptive study design was used to assess the quality of data reported in the WebDHIS software for the period from April 2017 to March 2020. A total of 265 health facilities and 77 data elements were selected using a multistage sampling procedure. Data was extracted from WebDHIS software using a standardized report functionality of the system. Data quality was assessed using data completeness, data consistency, and data cross-checking dimensions. Extracted data were all exported to Microsoft Excel version 2013 wherein the descriptive statistical analysis was performed by calculating frequencies and percentages. A score grading was used, which had three levels to rate the outcome of each data quality dimension. A score of less than 75 percent was rated as poor, between 75 to 89 percent was rated as good, and 90 percent and above was rated as very good.

Result: A total of 322 532 data element values were reported, 102 836 missing data values and 6 098 data errors were identified. The data completeness was rated at 92.7 percent, data consistency achieved a rate of 86.6 percent, and data accuracy was rated at 95.2 percent. The overall study result for data quality in the WebDHIS software was very good at 91.5 percent for the period under review.

Conclusion: The good quality of data was maintained since the implementation of WebDHIS software in Eastern Cape Province. However, the high number of missing data and data inconsistent over time were identified across all districts. The study emphasized the important role of WebDHIS software in the management of routine health data. Regular data quality checking and timely correcting of data errors in the system, and continuous health information management training for healthcare workers are required to further improve data quality.

Abbreviations and Acronyms

AIDS	:	Acquired Immunodeficiency Syndrome
ART	:	Antiretroviral therapy
CCMDD	:	Central Chronic Medicines Dispensing and Distribution
DE	:	Data Element
DHIS	:	District Health Information System
DHIS2	:	District Health Information Software version 2
ECDoH	:	Eastern Cape Department of Health
EPI	:	Expanded Programme on Immunisation
HF	:	Health Facility
HIS	:	Health Information System
HISP	:	Health Information System Program
HIV	:	Human Immunodeficiency Virus
HMIS	:	Health Management Information System
HMN	:	Health Metrics Network
LMICs	:	Low Middle-Income Countries
NDoH	:	National Department of Health
NIDS	:	National Indicator Dataset
PHC	:	Primary Health Care
PMTCT	:	Prevention of the Mother-to-Child Transmission
POPIA	:	Protection of Personal Information Act
PRISM	:	Performance of Routine Information Systems Management
RHIS	:	Routine Health Information System
RSA	:	Republic of South Africa
Stats SA	:	Statistics South Africa
TB	:	Tuberculosis
Tier.Net	:	Three Interlinked Electronic Register
USAID	:	United States Agency for International Development
UWC	:	University of Western Cape
Web-DHIS	:	Web-based District Health Information System
WHO	:	World Health Organization

Table of Contents

Figures

Tables

Graphs

Appendices

Chapter 1: Introduction

1.1 Background to the Study

The role of health information over the years has gradually become a fundamental feature for health systems, in particular, planning and general management of healthcare services. In 2008, an international framework for strengthening the health systems was adopted, wherein the framework is made up of six building blocks which included health information and the other five blocks are service delivery; health workforce; medical products, vaccines, and technologies; financing; and leadership and governance (World Health Organisation (WHO), 2007). Health information plays a critical role not only in the management of health programs but to ensure that implemented health policies are properly measured (Blödt et al., 2018). The process of collecting health information in South Africa begins with collecting information from registers at each service point, followed by monthly collation into a monthly data input form for the health facility, and then get captured into the health information system (Mphatswe et al., 2012).

The introduction of the Health Information System (HIS) emanates from the desire to improve data collection, availability, and utilization. Teklegiorgis et al. (2016:1) describe HIS as a "system that integrates data collection, processing, reporting, and use of the information necessary for improving health service effectiveness and efficiency through better management at all levels of health services". The HIS serves as a fundamental element to ensure that collected health information from health facilities is appropriately organized and easily available at all tier levels of the health system to ensure that health managers can make informed decisions, manage day-to-day health activities and improve the efficacy of public health (Tshabalala & Taylor, 2016).

Although many countries have been implementing the HIS to provide comprehensive health information at all levels of health system service delivery, these rich sources of information are regularly overlooked for evaluating the causal effects of health programs due to data quality issues (Wagenaar et al., 2016). The impact of poor-quality data is significant to the efficiency and effectiveness of an organization, while good quality data are crucial to the success of the organization (Haug et al., 2011). According to Chen et al. (2014: 5172), data quality is "recognized as a multi-dimensional concept across public health and other sectors". Furthermore, Lee et al. (2021) found that routine HIS plays a fundamental role; however, bemoan that the quality of data produced from these systems in many low-income and middle-income countries remains inadequate.

In South Africa, the Ministry of Health has been using District Health Information System (DHIS), a standalone system since 1996, first in the Western Cape Province, and subsequently extended to the entire country in the year 2001 (Garrib et al., 2008). Since 2001, the system has undergone developments and is currently into a more modernized platform which is an online version that is commonly known as WebDHIS. The National Department of Health in South Africa officially migrated to WebDHIS in the year 2017 (NDoH, 2017). Since then, the Eastern Cape Province has been using WebDHIS as the source of routine and non-routine data. WebDHIS software has its strengths in technical and functional, however, these strengths have particular challenges and concerns that need consideration (Dehnavieh et al., 2019). A great trust has been placed in the DHIS for data management and supports the provision of the information needed to inform decision-making at all levels of the health system. It is therefore important that the DHIS produces good quality information. However, with the increasing implementation of health information systems in low-middle-income countries most health information systems are plagued with poor data quality (Manya & Nielsen, 2016).

Begum et al. (2020) alluded that the implementation of WebDHIS has the potential to assist countries to improve data quality and usability over time. Kiberu et al. (2014) found that the implementation of WebDHIS software improves data quality in particular timeliness and completeness from the district to the national level. A similar sentiment found that the implementation of WebDHIS software has a positive impact on improving the timeliness and completeness of data reporting over time (Begum et al., 2020). Notwithstanding the improvements in the role of health information and routine health information system in the global health system, the quality of health data is still very poor in low and middle-income countries as expected (Shama et al., 2021). Data quality is a cause for concern in particular the accuracy of data captured in Routine Health Information Systems, which is attributed partly to insufficient competencies of health information personnel (Nicol et al., 2016).

The WHO in collaboration with its partners developed a data quality review framework that is recommended to be used by countries to measure the quality of Routine Health Information Systems data (WHO, 2017). The WHO data quality framework provides a comprehensive and holistic approach to reviewing data collected from health facilities. Furthermore, the framework’s scope includes “routine and regular reviews of data quality built into a system of checks of the HIS data, an annual independent assessment of a core set of tracer indicators to identify data errors and gaps in reporting health facility data, and periodic in-depth program-specific reviews of data quality that focus on a single disease or program area" (WHO, 2017). The main goal of the data quality review framework is to contribute to the improvement of the quality of data used by healthcare workers and authorities to review the progress and performance of the health system (WHO, 2017). Furthermore, the WHO framework emphasizes the importance of quality of data across four dimensions: completeness, internal consistency, external comparisons, and external consistency of population data. This study followed the WHO framework approach as a guide to assess the data quality in the WebDHIS. The WHO framework metrics are built-in to WebDHIS for data quality assessment, which is typically utilized by Information Cadres at all tier levels in the health system. Using data from WebDHIS will enable the researcher to apply the framework metrics to assess the data quality without any difficulty and/or inconvenience of data collection processes.

1.2 The Problem Statement of the Study

The significance of health information is determined by its quality and utilization in decision-making. Furthermore, maintaining good health information is an essential part of strengthening a health system (Tilahun et al., 2018). In South Africa, data quality remains a challenge, particularly at primary health care facilities and district levels, and this is attributed to insufficient competencies in the routine health information system and also the lack of training for personnel (Nicol et al., 2016).

In South Africa, limited research studies have been conducted on the assessment of data quality in the DHIS software, considering their compliance with built-in data quality checks as per recommendations from the WHO to monitor data quality. The Eastern Cape Province started the implementation of WebDHIS software in the year 2017, with all public health facilities expected to use the system for data collection. The province has not conducted a comprehensive review of the quality of data collected from the public health facilities as suggested by WHO as part of the process for data consolidation (WHO, 2017a).

1.3 Research Question

What is the quality of WebDHIS data in terms of completeness; consistency; and accuracy?

1.4 Aims and objectives

1.4.1 Aim

This study aims to determine the quality of the data collected and reported in public health facilities as part of the routine health information system by assessing the data quality dimensions in the collected routine health information.

1.4.2 Objectives

- To determine data completeness in the routine health information system
- To determine data consistency within the routine health information system
- To determine data accuracy within the routine health information system

1.5 Significance of the Study

The Department of Health has been using DHIS as an official routine health information system since the late 1990s (Garrib et al., 2008). The DHIS exists to collect aggregated routine health data from all public health facilities in the country and is intended to support the monitoring and evaluation of the performance of the healthcare system and to provide timely and accurate information to support decision-making and health management (Lutge et al., 2016). An assessment of data quality is crucial to describe the status of the WebDHIS software in providing good quality data for supporting decision-making and health management.

The WebDHIS software in the Eastern Cape Department of Health (ECDoH) has not been evaluated to determine whether it produces quality health information to help decision-makers to be informed and take appropriate actions. The researcher believed that it was important to conduct the study to know the level of data quality in the WebDHIS software as reported by the public health facilities in the Eastern Cape Province. Understanding the level of data quality by healthcare workers and health managers requires them to commit to working together to ensure that appropriate steps are taken to manage health information.

1.6 Structure of the Thesis

Chapter 1 outlined the study's background, then followed by the problem statement, research questions, aim, and objectives were briefly discussed. In addition, the chapter discussed the significance of the study. Finally, the chapter discusses the structure of the thesis. The next chapter**,** which is **chapter 2,** explores the literature review on health information systems and related topics. The literature reviewed includes routine health information systems and the adoption of the District Health Information System in Developing Countries. In addition, the chapter review literature on the framework for assessing the quality of data in the HIS. This chapter concludes with literature on the factors influencing data quality, and the quality of data collected through the routine health information system. **Chapter 3,** presents an overview of the methodology followed in this study. This chapter starts with a discussion on the study research design, study population, sample methods, and sample size, and followed by a description of the characteristics of the study sites, and the unit of analysis which included health facilities and data elements. In addition, the chapter addresses the procedure for data collection, the technique used for data analysis, the definition of key data elements, and the study ethics considerations.

The last three chapters, which include **chapter 4**, deal with data presentation and summarise the data collected from the routine health information systems. The chapter starts with an introduction, followed by the presentation and analyses of data reporting and completeness, then followed by the presentation of data consistency wherein both data elements outlier report and data missing report are shown, and finally, the data accuracy is presented wherein the data validation report, and data marked for follow report are summarised. Then followed by **chapter 5,** where the researcher discusses, and analyses in detail the data collected from the routine health information systems. The chapter starts with an introduction, followed by the presentation and analyses of data reporting and completeness, then followed by data consistency wherein both data elements outlier report and data missing report discussed, and finally, the data accuracy is measured by data validation errors, and data marked for follow reports are presented. The final chapter, which is **chapter 6**, draws conclusions from all the chapters, especially the chapter that discusses the results. The chapter starts with a brief introduction, answers the research questions, and provides recommendations that should be implemented to improve and/or maintain the data quality in routine health information systems.

1.7 Conclusion

In summary, this chapter discussed the background of the study, the research problem statement, and the research question. This was followed by the aim and objective of the study, the significance of the study, and end with the layout of the study. The next chapter will focus on the literature review in relation to routine health information systems, the assessment frameworks for data quality in routine health information systems, and other related topics.

Chapter 2: Literature Review

2.1 Introduction

This chapter reviews the literature starting with an empirical review of the routine health information system (RHIS) and its role in the health system. In addition, it explores the adoption of the District Health Information System in Developing Countries. Then followed by the data quality assessment frameworks for routine health information systems. Lastly, the quality of the routine health information system data was discussed as an important element in improving health data reliability and validity.

2.2 Routine Health Information Systems

Saigí-Rubió et al. (2021) described a routine health information system as “any system of data collection, distribution, and use that provides information at regular intervals that is produced through routine mechanisms to address predictable health information needs”. Maïga et al. 2019) stated that RHIS has the potential to serve as a source of data to generate health statistics and indicators to track the progress of the implementation of health programs toward universal health coverage and to inform planning. A lot of countries have adopted RHIS as the preferred data source to provide routinely collected health activities on all levels of health system service delivery (Wagenaar et al., 2016). A well-functioning RHIS can be achieved through effective and efficient management and clear processes as well as accountability in the healthcare system (Cheburet & Odhiambo-Otieno, 2016). In public health, the concept of a health information system (HIS) attracted attention when it was named one of the six building blocks for strengthening the health system (WHO, 2007).

The primary aim of the RHIS is to provide quality health information to health managers and authorities to strengthen the utilization of information to improve the decision-making and performance management of health programs. Hung et al. (2020) nicely describe RHIS as the information system that collects and provides information at regular intervals on services and activities delivered in health facilities which are crucial for management decision-making as well as strategy development in all tier levels of the health system. RHIS helps to generate data collected from various health establishments that provide and/or manage healthcare services (MEASURE Evaluation, 2016). This statement was echoed by Shiferaw et al. (2017) who stated that the main objective of the RHIS is to produce good quality routine health information; and to strengthen the effective utilization of routine health information for decision-making. As a management tool, RHIS enables the timely availability and utilization of health information within the health system at all levels but poor information support is identified as the biggest challenge (Leon et al., 2020).

Though Routine Health Information Systems (RHISs) are available in most countries in particular the Low Middle-Income Countries (LMICs) they are often overlooked as the source of health data to evaluate the effectiveness of health programs due to concerns regarding data quality level (Wagenaar et al., 2016). Hoxha et al. (2020) concurred that despite the significant role of RHIS data in improving health system functioning, numerous challenges continue to impede their use in practice. The study conducted by Kebede et al. (2020) found that the implementation of health information systems is poorly coordinated in primary health facilities and alluding to a lack of accountability and support mechanisms. In contrast, the study conducted in Uganda found that using of health information system for data collection improved data completeness and timeliness as well as usage throughout the health system level (Kiberu et al., 2014).

However, a study conducted by Wude et al. (2020) found that there is a high rate of utilisation of routine health information among healthcare workers. Furthermore, the study identified training on health information, supportive supervision, perceived culture of health information, having a standard set of indicators, and competence in routine health information as factors that help to improve routine health information utilization. Lee et al. (2021) affirmed that training of staff and the utilization of electronic health management information systems as useful to improve RHIS data. Nguefack-Tsague et al. (2020) conducted a cross-sectional study in Cameroon and found the factors that are associated with the performance of the RHIS include supportive supervision, provision of feedback from the hierarchy, and training on health information management. The next section will discuss the implementation of a District Health Information System in developing countries.

2.3 District Health Information System in the Developing Countries

In many developing countries, District Health Information System (DHIS) has evolved from being a research program into a prominent and largest health management information system for data collection for information management and decision-making (Begum et al., 2020). Manoj et al. (2012: 109) described DHIS as a "tool for collection, validation, analysis, and presentation of aggregate statistical data tailored (but not limited) to integrated health information management activities". The DHIS was developed for the collection of aggregated routine data from all of the public health facilities in a country, to facilitate analysis of health services provided in the country throughout the levels of the health system, to forecast required services for future planning purposes, and to enable evaluate the performance of healthcare workers (Garrib et al., 2008). DHIS is an open-source software that was developed and managed by the Health Information Systems Programme in collaboration with the University of Oslo and the University of Western Cape.

The introduction of DHIS dates back to the year 1996 when it was initially implemented in the Western Cape Province, South Africa, and subsequently, in the year 2001, the implementation was extended to the other eight provinces (Garrib et al., 2008). The conceptual drive for developing DHIS stems from the effort to rebuild and strengthen the healthcare system and the desire to improve the health information system in post-apartheid South Africa in the mid-1990s (Karuri et al., 2014). The initial purpose of the DHIS was to manage routine health information from primary health care (PHC) facilities, and since then, the system has grown from being a standalone for Microsoft Office Professional which included Visual Basic to online open-source software that is now used by many countries around the world, in particular, the developing nations (Braa & Sahay, 2017). In addition, the online version of DHIS software was first implemented in Kerala in India in the year 2006, and since then the software footprint expanded into many countries (Faujdar et al., 2019). Currently, DHIS2 is the world's largest health information management platform and is implemented in more than 100 countries (DHIS2, 2022).

The evolution of the DHIS has led many countries in particular the LMICs to adopt the software as the source of routine health data. (Githinji et al., 2017) affirmed that the introduction of DHIS improves the completeness, availability, and consistency of health data. These dimensions form part of measuring data quality (Adane et al., 2021). Ehsani-Moghaddam et al. (2021:88) describe data quality as "the degree to which a given dataset meets a user's requirements". The WHO considers the dimensions to examine data quality in the RHIS as completeness, internal consistency, external comparisons, and external consistency of population data (WHO, 2017). The data quality dimensions guide the activities and processes undertaken to evaluate the RHIS data. The good quality of RHIS data is important to guide the decision-making by health professionals and/or managers to monitor the health care service utilization, and performance of the health system against the set targets. Highlighting the critical role of RHIS data in improving the health system, is not only important for management but to strengthen the management of the health information systems (O'Hagan et al., 2017).

Over the past years, many countries have opted to use DHIS as a tool for health data management (Hagel et al., 2020). Several studies have been conducted and the researchers affirmed the DHIS implementation and the positive impact it has in these countries. Manoj et al. (2012) alluded that in Sri Lanka, the implementation of DHIS2 was initiated as a project for graduate research projects who were completing their master's degrees, and secondly, the Ministry of Health adopted DHIS2 for several projects both centrally and in the provinces for health data management. In the Lao People's Democratic Republic, the implementation of DHIS2 was initiated as part of the National Health Sector Reform Strategy adopted by the country's Ministry of Health (Choudrie et al., 2017). In Bangladesh, Begum et al. (2020) conducted a study in two districts to understand the facilitators and barriers to implementing DHIS as a tool for health data collection and utilization. The findings highlight DHIS as the data repository for different health data and it is used by multiple stakeholders. The study also noted the exclusion of data from Bangladesh's large private health sector which is a limitation to complete the picture of the country's health system.

In Africa, several countries have implemented DHIS as the data source. Kiberu et al. (2014) stated that Uganda adopted DHIS2 early in the year 2011 as a pilot project in four districts, and subsequently rolled all districts in the country. The study findings revealed that the implementation of DHIS2 resulted in improved timeliness and completeness in reporting routine data from the district to the national level (Kiberu et al., 2014). According to Hagel et al. (2020), Kenya is one of the first countries to implement DHIS on a national scale. The study revealed that DHIS2 as a tool has the potential to improve health information management, which is crucial for supporting decision-making (Hagel et al., 2020). In Tanzania, the implementation of DHIS was kick-started at the piloted projects between November 2008 and August 2009 in the Kibaha and Bagamoyo districts, thereafter the national rollout was completed in December 2013 (Kiwanuka et al., 2015). In a study conducted in Ethiopia, Kanfe et al. (2021) findings showed that the country has a good utilization of DHIS among health professionals. In addition, skills, training, supportive supervision, feedback, and motivation as the determinant factors for DHIS utilization.

According to Shuaib et al. (2020) found that in Nigeria, the DHIS was adopted as the platform of the National Health Management Information System for real-time data reporting and to promote government ownership and accountability. The study affirms that the implementation of DHIS has helped to strengthen the accuracy and completeness of routine immunization information (Shuaib et al., 2020). By the same token, the study conducted by Githinji et al. (2017) confirmed that Kenya adopted DHIS as the information system to report routine health data, which is used by health facilities assigned to 299 sub-counties and 47 counties. In addition, the study affirmed that since the adoption of the DHIS, the country has sustained improvements in the data completeness from health facilities (Githinji et al., 2017).

Although DHIS has been widely adopted by countries, some ministry of health has opted to use the system for specific programs. A study was conducted in Senegal which assessed the quality of the malaria data reported in the first four years of implementation of the DHIS (Muhoza et al., 2022). The study findings reveal that the quality of reporting malaria indicators in the country improved over time as well as data accuracy to enable appropriate monitoring progress in the malaria programs (Muhoza et al., 2022). In Guinea, the adoption of DHIS as the health information system of choice to capture surveillance data started in the year 2015, whereby the ministry of health established and formulated a strategic plan to improve health information (Reynolds et al., 2022). The actual implementation DHIS was piloted in two regions, and the findings revealed that there has was an increase in the national average timeliness of reporting at 72.2%, and data completeness at 98.5% (Reynolds et al., 2022).

In the study conducted by Dehnavieh et al. (2019) involved 11 countries that are using DHIS, and the primary aim of the study was to examine the strengths and operational challenges of DHIS as an instrument for decision-makers to evaluate the performance of the healthcare system. The findings affirmed that the DHIS is technically and functionally sound as the software, which enables the decision-makers and policymakers to take informed decisions.

The implementation of the DHIS software is largely dependent on the availability of resources and capabilities of the countries to support and manage the software. The literature has revealed that the adoption of DHIS by countries has yielded in improving data quality, and in some instances, it did not make much impact. Following the topic of DHIS software implementation in developing countries, the research will now discuss frameworks that are usually used to assess the data quality in the Routine Health Information System in more detail.

2.4 Routine Health Information Systems Assessment Frameworks

The approaches to RHIS assessment are developed based on one or more domains such as technical, sociological, economic, human, and, organizational. Over the years there are various frameworks have been adopted for the assessment of RHIS. Andargoli et al. (2017) state that the framework aims to provide a set of guidelines and procedures for the evaluation of the adequacy of healthcare information systems. The evaluation criteria are important for an undertaking assessment of the health information systems to ensure that parameters are demarcated. In this section, frameworks for the evaluation of RHIS are reviewed which are deemed to be suitable for data assessment in the healthcare system setting.

2.4.1 The Health Metrics Network Framework

The Health Metrics Network framework (HMN) as an evaluation tool for health information systems was developed by WHO in collaboration with its partners to measure the performance of the implemented country HIS (Barro et al., 2020). The HMN framework focuses on two core requirements of health system strengthening in LMICs. Firstly, the need to enhance entire health information and statistical systems, and secondly, to concentrate efforts on strengthening country leadership for health information production and use (WHO, 2008). The framework describes health information system components and standards in terms of HIS resources, indicators, data sources, data management, information products, and dissemination and use (WHO, 2008). Furthermore, the framework outlines the principles, processes, and tools that should be implemented to strengthen the country's HISs and development thereof.

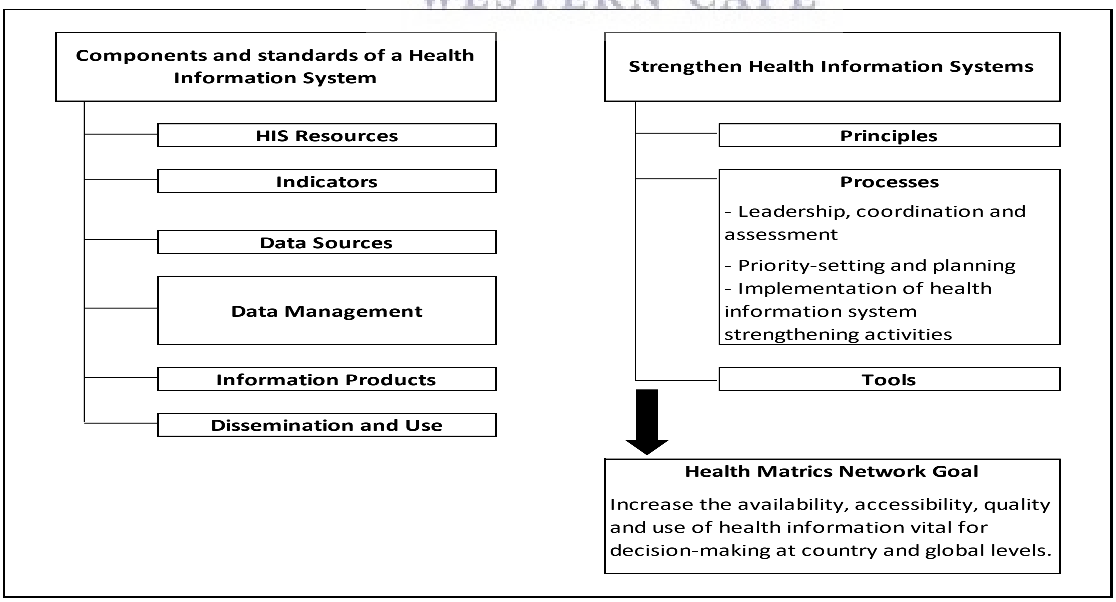


Figure 2.1: Health Metrics Network Framework
Source: WHO (2008)

As shown in figure 2.1, the HMN framework has two parts. The first part shows the components and standards of a health information system which describes the six components of health information systems and provides standards for each component. The components address different aspects of the information systems such as input which include HIS available resources, a process that is performance indicators, data sources and data management, and lastly the output which comprise information products, and data dissemination and use (WHO, 2008). The second part shows the key elements for strengthening a country's health information systems which should serve as the guiding principles, processes, and tools that should be taken together to outline a roadmap for strengthening health information systems (WHO, 2008).

The HMN framework's primary goal is to increase the availability, quality, value, and use of timely and accurate health information through continuous improvement and support for the development of country health information systems. The HMN framework enables the comprehensive evaluation of the components of the country HIS. The framework includes a wide range of stakeholders in the evaluation process as an important accomplishment. However, the long list of indicators and the time required to complete the evaluation are limitations to its consistent use (WHO, 2008).

2.4.2 The Performance of Routine Information System Management

The Performance of Routine Information System Management (PRISM) framework was developed by MEASURE Evaluation together with John Snow Inc. as a mechanism to measure the performance of HIS (MEASURE Evaluation, 2011). This framework broadens the analysis of routine health information systems to include three key factors for success: firstly, the behavioural determinants which cover knowledge, skills, attitudes, values, and motivation of the people who collect and use data; secondly, the technical determinants which include data collection processes, systems, forms, and methods; and thirdly, the organizational/environmental determinants which encompass information culture, structure, resources, roles, and responsibilities of the health system and key contributors at each level (Aqil, Lippeveld & Hozumi, 2009).

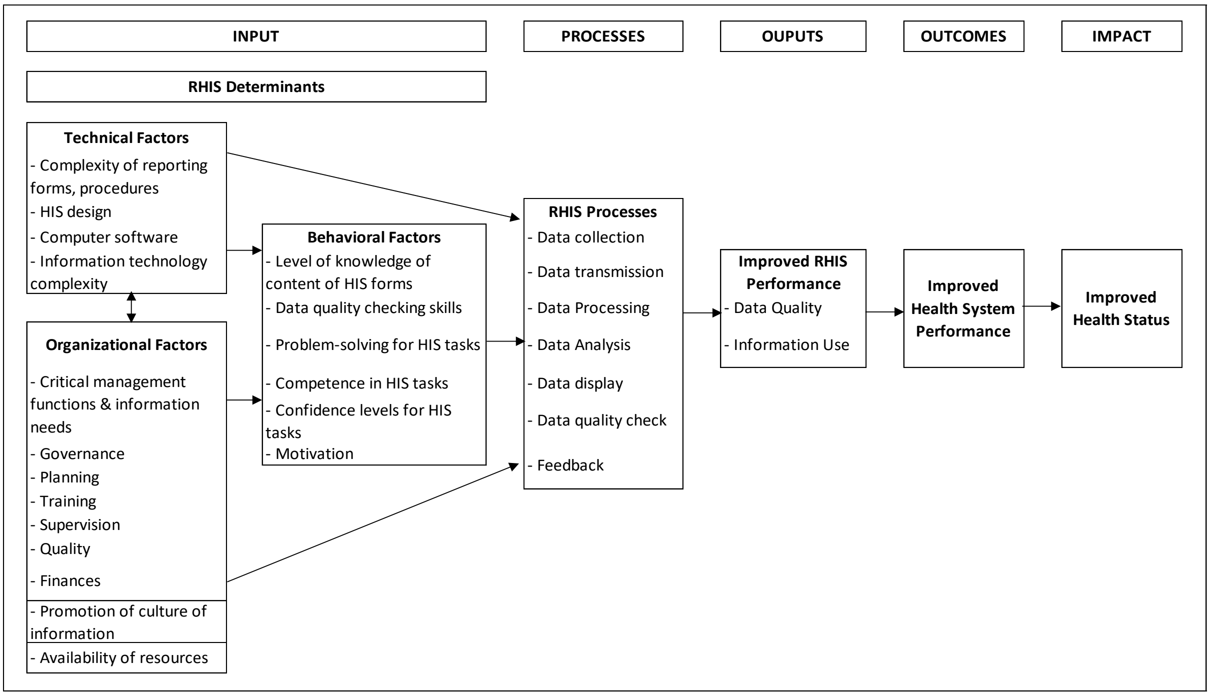


Figure 2.2: PRISM Framework
Source: Aqil, Lippeveld, and Hozumi (2009)

As shown in figure 2.2, the PRISM framework outlines the RHIS determinants that impact the performance of HIS. The PRISM framework explores how much the HIS processes influence the performance of HIS by testing the links between the activities. The framework tests the relationships among technical, behavioral, and organizational determinants of the RHIS process and performance. The framework illustrates that technical and organizational determinants have a direct influence on RHIS processes and performance. Behavioral determinants can also affect RHIS processes and performance directly, and they also can be influenced by both technical and organizational determinants (Aqil, Lippeveld & Hozumi, 2009). The framework aims at improving data quality and continuous use of information for decision-making (Aqil, Lippeveld & Hozumi, 2009). The PRISM framework can play an important role for HIS policymakers and practitioners to assess the RHIS and evaluate RHIS strengthening interventions to improve data quality and utilization (Hotchkiss et al., 2010).

2.4.3 The WHO Toolkit for Data Quality

This WHO toolkit was developed through collaboration between WHO, The Global Fund to Fight AIDS, Tuberculosis, and Malaria (The Global Fund), Gavi, the Vaccine Alliance (Gavi), and the United States Agency for International Development (USAID)/MEASURE Evaluation (WHO, 2017). The toolkit provided a unified approach to data quality. The toolkit integrates and builds upon existing tools and methods designed to assess data quality at the facility level, taking into account best practices and lessons learned from different countries (WHO, 2017).

Adane et al. (2021) describe the WHO toolkit as a framework for data quality that provides a method for analysing routine HIS data using four dimensions of data quality. The WHO data quality dimensions include completeness and timeliness of data, followed by the internal consistency of the data compared over time and between indicators, then followed by external consistency of data when compared with data from other sources, and lastly the external comparisons of population data (WHO, 2017). The WHO toolkit uses a methodology that comprises two separate processes that can be used jointly or separately, namely; desk review of the data, and health facility assessment (WHO, 2017). The desk review examines data quality across four dimensions that have been captured into RHIS. Further, the desk review examines a core set of indicators or data elements that are selected across program areas concerning these dimensions (WHO, 2017).

The desk review utilizes the monthly routine data as reported by health facilities, sub-districts, and/or districts in the WebDHIS software. The toolkit also makes provision for the assessment of data quality at the health facility level through verification of indicator or data element values for specific reporting periods, as submitted by health facilities to the next reporting level, as well as an evaluation of the completeness of reporting and required data collection (WHO, 2017). The data verification process encompasses data checking from source documents (registers and tally sheets) and comparing data reported through the RHIS to determine the discrepancies between the data sources.

The different assessment frameworks focus on different aspects of the HIS data quality and undertake different activities. The WHO toolkit framework was adopted for data quality assessment and the focus was on the data quality dimensions of completeness and internal consistency (WHO, 2017). The two dimensions included are completeness and internal consistency which will allow the researcher to use WebDHIS software data and reports only to assess the quality of RHIS data in the Eastern Cape Province. The excluded dimensions are external consistency of data when compared with data from other sources, and lastly the external comparisons of population data which require access to different source documents and permissions. Following the discussion of the frameworks for the assessment of the RHIS, the study will now provide an overview of the data quality in the RHIS.

2.5 Quality of Data Collected Through Routine Health Information System

Good quality data and effective data quality assessment are required to ensure that accurate monitoring and evaluation of the health programs are undertaken in order to adopt interventions and measurements to improve the efficacy of the health system (Chen et al., 2014). In many countries, the RHISs are becoming a standard for the management of health data. A well-functioning RHIS is important not only for good quality health data but to support and strengthen the health system. Tamfon et al. (2020) stressed the importance of strengthening the functioning of the RHIS to ensure that good-quality data is available and utilized.

The cross-sectional study which was conducted by Rumisha et al. (2020) in Tanzania, found that the implemented RHIS is weak, and also identified that there are high data variations in the tool utilization and accuracy at facility and district levels. The findings showed that the rate of registers is 91%, while the utilization of the report forms is at 86.9%. Furthermore, the study attributed the inaccurate data in the RHIS to inappropriate recording in the source documents and poor adherence to the set procedures.

A study conducted by Maïga et al. (2019) assessed the data quality of the RHIS at the national level and subnational levels for a period from 2013 to 2017. The findings reveal that the RHIS has extreme data outliers, a lack of consistency for the reported data over time and between indicators, and challenges related to projected target populations. Furthermore, the researchers highlighted that the completeness of reporting for the 14 countries was generally high with a median of 95% of countries, and also the researchers identified that the facility data-based statistics analysis was not done regularly in the countries.

The study conducted by Chen et al. (2014) found that the prominent data quality dimensions in the public health system include completeness, accuracy, and timeliness of data. This is also validated in the study conducted in Nigeria, where researchers found that issues such as completeness, accuracy, and internal consistency of facility-based routine data are critical to measuring and/or impacting data quality (Bhattacharya et al., 2020). Further, the issue of data quality in the health system is identified as a challenge in the study conducted in the United State of America, where researchers found that there are several challenges on data unrepresentative, incomplete, and inaccurate from the health information systems (Ngugi et al., 2019).

Nicol et al. (2016) conducted a study in South Africa that included two health districts to assess the quality of routine data wherein the focus was on the prevention of the mother-to-child transmission of HIV (PMTCT) program. The study found that data completeness was relatively high at 91% at the facility level and 96% at the district level. The study reveals the considerable data quality amongst the source documents with an average accuracy between the register and routine monthly form at 51% and between the routine monthly form and RHIS at 84%. The study found that a major challenge in routine health data is accuracy which is attributed to the high margins of differences between the source documents and the health information system. The literature review study by Roomaney et al. (2017) also found that there is the high considerable challenge to the quality of data that is reported in the RHIS in South Africa.

A cross-sectional study by Teklegiorgis et al. (2016) in the Dire Dawa Administration health facilities in Ethiopia highlighted the data quality issues in the RHIS. The study findings revealed that the overall data quality was found to be 75.3%, and also confirmed the significant association between trained staff to fill format, decisions based on supervisor directives, and department heads seeking feedback and data quality. This is also validated by the study conducted by O'Hagan et al. (2017) in Malawi. The study found that the public health system is challenged with various issues of data which include the availability, completeness, and accuracy of data for programs as well as unsatisfactory level of comprehensiveness and reliability of the HMIS supervision, and inadequate staff training on HMIS at the facility level.

A study conducted by Moukénet et al. (2021) in the Massaguet district in Chad revealed data quality issues in the HMIS for both the district and health centre levels. The study findings show that health centre data completeness was high in the HMIS. Furthermore, the study showed that there was an association between workload and higher odds of inaccuracy in reporting data, and also showed an association between the presence of health information personnel and low inaccuracy in the reported data. The study by Shama et al. (2021) found that the level of good data quality in public health facilities was less than the expected national level. The study showed that the level of good quality data was 51.35% in public health facilities in the Harari Region in Ethiopia. Furthermore, the study showed that the lack of trained personnel able to fill the reporting format and feedback from higher-level was affecting data quality. Following a discussion on the state of quality of data collection using RHIS, the next section will discuss the factors that influence the data quality in the RHIS.

2.6 Factors Influencing Data Quality

The factors that influence data quality in the RHIS usually include technical capabilities, organizational structure and resources, and behavioural (Hoxha et al., 2020). An exploratory case study conducted in Kenya identified financial incentives given to health facilities as part of a project to improve maternal care services influenced data accuracy, timeliness, and completeness in the health facilities (Manya & Nielsen, 2016). The qualitative study conducted in two districts of Bangladesh revealed that a clear health strategy and framework for RHIS, regular training of staff, and incentives influence the performance and ultimately the data quality (Begum et al., 2020).

A literature review aimed at understanding system design barriers to data quality and use in LMICs and identifying any major research gaps revealed that the training of individuals plays an important part in strengthening the performance of the HIS and data quality (Kumar et al., 2018). Another study on India's data quality revealed that training and supportive supervision undertaken by higher levels of health managers were identified as playing important role in improving health data quality (Singh et al., 2016). Another study conducted in Ethiopia by Solomon et al. (2021) also highlighted a similar finding which affirms that supportive supervision and training are associated with data quality. The study found that only about 52.2% of healthcare workers were trained in health information management, and 62.5% had supervisory support visits (Solomon et al., 2021).

A study conducted in Tanzania by Mboera et al. (2021) assessed the data utilization and factors influencing the performance of HMIS. The study found inadequate analysis and poor data utilization practices, inadequate human resources, low supervision visits, and lack of standard operating procedures on data management that are significantly affecting the HMIS performance (Mboera et al., 2021). Poor performance of the HMIS has a direct impact on the data quality (Hlaing & Zin, 2020). Another cross-sectional study was conducted in Ethiopia by Getachew, Erkalo and Garedew (2022) which focused on determining the data quality and associated factors in the HMIS. The study showed that the overall level of data quality was 83% in Shashogo District health centres. In addition, the study found supportive supervision, checking data accuracy, filling registrations, and confidence level to be factors associated with data quality (Getachew, Erkalo, & Garedew, 2022).

Lemma et al. (2020) conducted a scoping review on the interventions for improving data quality and the use of routine health information system data in low- and middle-income countries. The study found that the interventions to improve data quality should include implementation of the improved technological solutions, capacity-building activities, and continuous data quality assessment and feedback systems within public health. The systematic review study conducted by Leon et al. (2020) also found that technical interventions have an impact on the improvement of data quality in particular the timeliness and accessibility of data.

The literature reviewed in this section highlighted various aspects that contribute to the data quality in the RHIS. However, this study will not be considering all factors due to the approach and objective adopted for the study. The literature will assist the researcher to formulate and collaborate the study findings with real-world experiences to recommend appropriate actions for the improvement of routine health data management.

2.7 Conclusion

The literature review showed the critical role of the routine health information system in the public health system. The literature review revealed how the DHIS software has evolved and been implemented in many countries in particular the LMICs to strengthen the quality of routine health data. Moreover, the review provided insight into the frameworks for the assessment of RHIS. In addition, the literature review highlighted the data quality in the RHIS, and finally, discussed the factors that are influencing the data quality. Although RHIS is taunted as an important management tool to assist with the improvement of health data and to address the challenge of good quality data, it is worth noting that obstacles and many factors are identified to have had a direct contribution to this phenomenon.

Chapter 3: Research Methodology

3.1 Introduction

This chapter aims at providing an overview of the research methodology followed in this study. It specifically focuses on the study research design, study population, sample methods, and sample size, characteristics of the study sites, health facility and data elements to be used as the unit of analysis. In addition, it explains the procedure for data collection, the technique used for data analysis, the definition of key data elements, and the study ethics considerations.

3.2 Research Design

Kothari (2004) describes research design as a conceptual structure within which research is conducted, and serves as the blueprint for the collection, measurement, and analysis of data. Research design facilitates the various research operations to ensure that the most valid, reliable, and credible research results and research objectives are achieved. Mouton (2012) states that research design focuses on the end product and the logic of the research particularly the activities of the research project. Research design is a plan or strategy of investigation that enable data collection which is relevant to answer both the research problem and questions. The retrospective-sectional study involves the investigation of a phenomenon, situation, problem, or issue that has happened in the past (Kumar, 2011). The retrospective study design enabled the researcher to easily and quickly collect data for the study and to measure the scale and performance of the implemented routine health information systems.

There are three types of research methodologies namely quantitative research, qualitative research, and mixed approach. The basic difference between these methodologies is that quantitative research relies on measurement, counting, and scales, whereas qualitative research uses words and sentences to describe the phenomenon, and mixed research uses both statistics and text, this is done to combine the advantages of quantitative and qualitative approach while as the same time avoiding disadvantages of these research approaches (Bless, Higson-Smith, & Sithole, 2013). The difference between these research approaches is not just that of quality, but rather that of the procedure that best suit the research project.

This study employs a quantitative approach using a retrospective descriptive study design to assess the quality of data within the WebDHIS software for the period from April 2017 to March 2020. The Eastern Cape Province implemented WebDHIS software in all public health facilities in April 2017. The WebDHIS software is one of the routine HISs that is used by public health facilities for capturing, analyzing, and storing routine health information. This research design was used because the outcome of interest already occurred and the data collection is done from records (Ranganathan & Aggarwal, 2018).

3.3 Study Setting

The research setting is the Eastern Cape Province, which is one of the nine provinces in the Republic of South Africa (RSA). The Province has two metropolitan municipalities namely Buffalo City Metropolitan and Nelson Mandela Metropolitan Municipalities and also has six District Municipalities which are further divided into thirty-one Sub-districts. The Eastern Province is home to approximately 6.7 million people and an estimated 89.3 percent rely on public health care services (Stats SA, 2020; ECDoH, 2020). The Eastern Cape Province covers an area of 168 966km2 and it constitutes a share of about 11.1 percent of the RSA population, and it is the second-largest province in surface area and third in terms of population size (ECDoH, 2020). The Eastern Cape Province Department of Health has a total of 860 public health facilities consisting of 727 Primary Health Care (PHC) Clinics, 41 Community Health Centres, and 92 hospitals as a platform for service delivery (ECDoH, 2020).

Figure 3.1: Districts of the Eastern Cape Province

3.4 Study population

The study population includes all public health facilities (850) in the Eastern Cape Province that are mandated to report health information every month as per National Indicator Dataset (NIDS) into WebDHIS software. In addition, 95 data elements comprising NIDS form part of the study population and represent the sampling frame. All public health facilities are expected to report data for each data element for all 12 months in a year. The NIDS data elements that will be included comprise only the Monthly Routine Core Health Facility, therefore it will exclude ART Quarterly; TB Quarterly; Monthly Routine Non-Facility Health Services; and Periodic campaigns, as well as central chronic medicine, dispensing and distribution programs which are captured in multiple software and not routinely reported every month.

Table 3.1 provide a summary of health facilities per district for the period under review, and Table 3.2 provide a summary of data elements as per NIDS categories.

Table 3.1: Public Health Facilities in the ECDoH distributed by district

Financial Year	Health District	Alfred Nzo	Amathole	Buffalo City	Chris Hani	Joe Gqabi	Nelson Mandela Bay	Oliver Tambo	Sarah Baartman	Total
2017/2018	Clinic	73	152	72	148	52	40	135	59	**731**
	Community Health Centre	2	5	5	7	0	9	10	3	**41**
	Hospital	7	14	5	16	11	7	13	15	**88**
	Province	**82**	**171**	**82**	**171**	**63**	**56**	**158**	**77**	**860**
2018/2019	Clinic	73	152	72	148	52	40	135	59	**731**
	Community Health Centre	2	5	5	7	0	9	10	3	**41**
	Hospital	7	14	5	16	11	7	13	15	**88**
	Province	**82**	**171**	**82**	**171**	**63**	**56**	**158**	**77**	**860**
2019/2020	Clinic	72	143	74	152	52	39	136	59	**727**
	Community Health Centre	2	5	5	7	0	9	10	3	**41**
	Hospital	7	14	5	16	11	7	12	15	**87**
	Province	**81**	**162**	**84**	**175**	**63**	**55**	**158**	**77**	**855**

Source: ECDoH (2017; 2018; 2019)

Table 3.2: National Indicator Dataset 2017

Routine Core Health Facility - Monthly	Data elements
Adolescent Health	2
ART Monthly	4
Child and nutrition	19
Communicable Diseases	2
EPI	11
Eye Care	3
HIV	20
Inpatient Management	77
Malaria	1
Management PHC	11
Maternal and neonatal	28
Mental Health	7
Non-communicable disease	5
Oral Health	4
Quality	8
Rehabilitation	4
STI	1
TB Monthly	10
Womens Health	12
Total	**229**

Source: NDoH (2017a)

Although the NIDS is established to be a standard list of data elements that are collected in the Department of Health, however, some health services are not rendered in all health facilities. Table 1.3 provides a summary of the dataset collection points per health facility category.

Table 3.3: National Indicator Dataset 2017 Collection Points

Routine Core Health Facility - Monthly	Clinic	Community Health Centre	Hospital
Adolescent Health		X	X
ART Monthly	X	X	X
Child and Nutrition	X	X	X
Communicable Diseases	X	X	X
EPI	X	X	X
Eye Care	X	X	X
HIV	X	X	X
Inpatient Management		X	X
Malaria	X	X	X
Management PHC	X	X	
Maternal and Neonatal	X	X	X
Mental Health	X	X	X
Non-communicable disease	X	X	X
Oral Health	X	X	X
Quality	X	X	X
Rehabilitation		X	X
STI	X	X	X
TB monthly	X	X	X
Women's Health	X	X	X
Total	**16**	**19**	**18**

Source: NDoH (2017a)

3.5 Sampling

A retrospective descriptive study design for the period from April 2017 to March 2020, was conducted using a multistage sampling approach to select public health facilities and data elements from the NIDS. Bless, Higson-Smith, and Sithole (2013) describe multistage-stage sampling as the sampling method that involves the selection of samples from a complete list of units of the population under investigation. The multistage sampling process involved 2 stages: the selection of public health facilities; and the selection of data elements. The Raosoft Sample size calculator will be used to calculate the sample size, which will include a 5% margin of error, 95% confidence level, and the population of 850 health facilities and 95 data elements. The sample size for the study is 265 for health facilities and 77 for data elements.

Stage 1: Public Health Facilities Selection

This stage involved the random selection of two hundred and sixty-five (265) facilities sampled from all the health districts. All health facilities were listed and sorted alphabetically in Microsoft Excel version 2013. The Microsoft Excel Random Function was then used to generate random numbers in a column next to the list of health facilities. The random numbers generated for health facilities were sorted from the smallest value to the largest. The first 265 smallest random numbers and facilities were then selected and included in the study.

Stage 2: Data Element Selection

This stage involved the random selection of seventy-seven data elements sampled from the NIDS list. All data elements were listed and sorted alphabetically in Microsoft Excel version 2013. The Microsoft Excel Random Function was then used to generate random numbers in a column next to the list of data elements. The random numbers generated for data elements were sorted from the smallest value to the largest. The first 77 smallest random numbers and data elements were then selected and included in the study.

3.6 Inclusion and exclusion criteria

a. Inclusion criteria

- All the selected health facilities that are government-owned and classified as clinics community health centres and/or hospitals.
- The NIDS 2017 data elements that were reported monthly by the public health facilities

b. Exclusion criteria

- Health Facilities that are not government owned
- Health facilities that were intermittent during the period under review
- Specialised health facilities such as TB Hospital, Psychiatric, etc.
- Mobile and satellite clinics, and other health facilities
- Data Elements that do not apply to all health facilities
- Data elements that were collected as part of the quarterly report, campaigns, and routine data from non-facility health services

3.7 Data Collection

This study used secondary datasets from WebDHIS software to measure the data quality metrics for routine data reported by the health facilities. The monthly routine health data is collected and reported by facilities and captured into WebDHIS software by the Information Cadre. This study utilized routine health data and data quality reports from WebDHIS software. Permission to extract data from WebDHIS software was given to the researcher by the Manager responsible for the Health Information Management Systems in the Eastern Cape Province. The WebDHIS software data quality reports are a built-in functionality; Data Validation Report (Appendix F), Outlier and Missing Report (Appendix G), and Marked for Follow-Up Report (Appendix H). These reports enable the system users to easily identify data quality issues as discovered in the captured data in the system. The WebDHIS software pivot table is one of the system analytic tools used for the manipulation or extraction of facility data (Appendix E).

Data Validation Report: This report shows data violations based on the pre-defined validation rules (HISP Team, 2016). The validation rules are expressed by conditions set between data elements. **Outlier and Missing Report**: This WHO Data quality tool was used to identify data gaps and outliers (HISP Team, 2016). **Marked for Follow-Up Report:** This report shows the list of data elements that data values marked for follow-up. The data values can be marked for follow-up by the data capture when the reported data value is not confirmed to be correct (HISP Team, 2016). **WebDHIS Routine Data Pivot Table:** This is an analytic tool that was used to summarise and arrange data according to dimensions such as data elements and indicators, periods, and organizational units (HISP Team, 2016).

Data were extracted from the Eastern Cape Department of Health WebDHIS software for all sampled health facilities and data elements for a period of three years, starting from April 2017 to March 2020. The extracted data were downloaded into a Microsoft Excel format. The researcher was responsible for the data collation which was done through Provincial Office, and this was done whilst adhering to Covid-19 protocols. The WebDHIS software data is aggregated numbers and it does not contain individual patient details, and the researcher adhered to the Protection of Personal Information Act and other relevant regulations within the Eastern Cape Department of Health.

3.8 Data analysis

The three dimensions which included completeness, consistency and internal cross-check were used to determine the quality of data according to WHO Data Quality review toolkit analysis techniques. All datasets with corresponding dates are captured into the WebDHIS software and the system data quality analysis tools that are utilized for data quality checks were used to measure the state of data quality. The data quality analysis tools data were all exported to Microsoft Excel version 2013 wherein the statistical analysis was performed by calculating frequencies and percentages. Simple descriptive statistics were used to analyze and display the study results. Data were summarized to obtain percentages of completeness, accuracy, and consistency scores and then graded according to the set criteria (Table 3.4). The graphs and tables were then created to display the results from the analysis.

The quality of data reported in the RHIS was assessed using the selected dimensions. The WHO routine data quality framework includes four dimensions that evaluate various aspects of the data quality reported inside and outside the health system (WHO, 2017). The first dimension involves the completeness and timeliness of health facility reporting and the completeness and timeliness of data. The second dimension focused on the internal consistency of reported data which includes the presence of outliers, consistency over time, consistency between indicators, and consistency of reported data and original records. The third dimension involves external consistency or cross-checking data with other data sources which includes comparing data from the RHIS with data from other sources such as survey results. The fourth dimension comprises the external comparison of population data which involves checking data consistency with population estimates (WHO, 2017). For this study, the researcher adopted the WHO framework for routine data quality as shown in Table 3.4.

Table 3.4: Assessing the quality of routine data reported by health facilities

Data Quality Metric	Analysis	Data Quality Review Guidance	Data Source
Dimension 1: Completeness			
Completeness of health facility reporting	The proportion of expected monthly reports submitted by health facilities	The completeness rate of reporting should be above 75%	WebDHIS software: • Facility Routine Data Report • Outlier and Missing Data Report
Completeness of data reported	The proportion of non-missing values for a given data element in expected monthly reports	The completeness rate of reported data should be 100%	
Dimension 2: Internal consistency of reported data			
Outliers	Number of extreme outliers (+3SD from the mean) of monthly values during the period under review	A data element value of above 3.5 on the modified Z-score is considered an Extreme Outlier.	WebDHIS software: • Outlier and Missing Data Report • Facility Routine Data Report
Consistency over time	The consistency over time was calculated as the cumulative value of the data element for the preceding years over the mean value of the data elements for the comparison year.	Assess the reported data element values by comparing the current year to the value predicted from the trend in the preceding years. It is expected that the reported values for the reference year be within a ratio of +/- 1.33 for the preceding years	
Dimension 3: Internal data cross-checking			
Consistency between related data elements	The relationship between two data elements at the facility level is assessed by comparing their correlation to the values reported.	This examines the extent to which two related data elements follow a predictable pattern.	WebDHIS software: • Data Validation Report • Data Marked for Follow-Up Report • Outlier and Missing Data Report
Data flagged for investigation	The data that is flagged for an investigation is still to be confirmed for accuracy	This examines the extent to which all data reported is confirmed.	
Completeness of data reported	The proportion of non-missing values for a given data element in expected monthly reports	The completeness rate of reported data should be 100%	

The data completeness was measured using both health facility reporting, and data element reporting. The health facility reporting was measured by comparing the number of health facilities reported data as compared to the number of eligible health facilities. Secondly, **data completeness** was measured to determine whether all data element values are reported for each month during the period under review. The data completeness is expected to be 100 percent for all health facilities and data elements.

The data consistency was measured by the data value in a series of values and whether or not it is extreme to the other values in the series. Data element value trends were compared using the predefined minimum and maximum values. To account for extreme outliers in the data elements, an exceptional tolerance level was applied by using a modified Z-score to measure outliers. The monthly data element value with an absolute value of modified Z-score of 3.5 and above standard deviation was identified as an extreme outlier and considered a potentially implausible data element value (WHO, 2017). The WHO deemed the data element values with a standard deviation that equals or surpasses the threshold of 3.5 as poor data quality (Bhattacharya et al., 2019; Agiraembabazi et al., 2021). This study accepts and adopts the WHO ascension of the 3.5 standard deviation threshold to identify extreme outliers . Secondly, **data consistency over time** was assessed to determine the plausibility of reported results for healthcare services in terms of the history of reporting of the data elements. Trends are evaluated to determine whether reported values are extreme to other values reported during the year or over several years (WHO, 2017).

The data accuracy was measured by cross-checking and quantifying the data capturing errors for each data element in the period under review. The data element with validation errors and/or data element values marked for follow-up are expected to be zero percent for all reported health facility data. In addition, data accuracy was assessed by determining whether all data element values are reported for each month during the period under the review.

3.8.1 Data Analysis Framework

This section describes and summarises the variables used in the study. The variables were used to measure and/or calculate the rating score for each data quality metric.

Table 3.5: Data Analysis Framework

Data Quality Metrics	Numerator	Denominator	Calculation of rate for Data Quality Metrics	Results Analysis, and Presentation
Health Facility (HF) Reporting	Count HF reported data (12 Months per Year), each HF should account for the number of DE times 12. Count the reported data per month from the HF. Only HF that report a value of 1 and above	100% of functional HF are expected to Report Routine Data (report DE 12 times per year)	% of HF Reported Data	Analysis and display were done per District and/or per HF type
Missing Data Element (DE) Values	Count missing data values in the DE. Each DE must be reported 12 times a year.	Count of Routine Data expected to be reported by HF per month (report DE 12 times per year)	% of Missing DE value % of Non-Missing DE Values. This is calculated by subtracting % Missing DE values from 100%	Analysis and display were done per District and/or per HF type
Outliers in the Reported Data	Count the DE with Extreme Outliers (12 Months per Year), each DE must be reported 12 times in a year. Count Outliers that are equal and above 3.5 in	Count of Routine DE values expected to be reported by HF per month (report DE 12 times per year)	% of DE Values with Outliers % of DE without Outliers. This is calculated by subtracting % of outliers from 100%	Analysis and display were done per District and/or per HF type
Data Consistency Over Time	Count the DE with values over an acceptable threshold of 33 percent and above. The average of the preceding years is compared to the reference year.	Count of DE values reported in the reference year	% of DE Values not consistent over time % of DE Values within acceptable consistency over time. This is calculated by subtracting % of DE Values not consistent over time from 100%	Analysis and display were done per District and/or per HF type
Validation Errors in the Reported Data	Count the identified validation errors in the reported routine data	Count of Routine DE values expected to be reported by HF per month (report DE 12 times per year)	% of DE Values with Validation error % of DE Values without Validation errors. This is calculated by subtracting % of DE Values with Validation error from 100%	Analysis and display were done per District and/or per HF type
DE Marked for follow-up in the Reported Data	Count DE values marked for follow-up in the reported data	Count of Routine DE values expected to be reported by HF per month (report DE 12 times per year)	% of DE Values Marked for Follow-up % of DE Values Not Marked for Follow-up. This is calculated by subtracting % of DE Values Marked for Follow-up from 100%	Analysis and display were done per District and/or per HF type

Health facility data quality was assessed using three dimensions completeness, consistency, and accuracy as described in the WHO recommended table. A weighted average of these dimensions was used to calculate a single weighted measure of data quality. Table 3.5 provide the summary of grading criteria for the assessment of routine health information system data.

Table 3.6: Criteria for Data Quality Status

Score	Criteria
90% and above	Excellent
75% to 89%	Good
74% and less	Poor

3.9 Validity and Reliability

Validity is defined as the extent to which an instrument measures what it is meant to measure (Heale & Twycross, 2015). The importance of ensuring the validity of the research instrument is to reduce errors in the measurement process. The WebDHIS software is used at the health facility level to capture and disseminate routine health information and measure and identify data quality issues and maintain data validity (NDoH, 2021). Capturing of data into WebDHIS software is done by information cadres and supervised by facility managers who manage the health facilities and assisted by program managers from the sub-district, district, and provincial offices who are part of the health system to ensure that reported data is trustworthy (NDoH, 2011; NDoH, 2012; NDoH, 2013). A study conducted by Kiberu et al. (2014) asserted the process of rollout of the WebDHIS involves training people such as record assistants, health information officers, and other health workers that are responsible for health data recording and reporting. Youssef et al. (2022) also affirmed the importance of training before implementing WebDHIS. In a study conducted in Lebanon, more than 80 training sessions were conducted throughout the country targeting health information officers, and focal persons who were working in all hospitals, laboratories, and medical centers (Youssef et al., 2022).

The WebDHIS software data collection process is guided through district health management information management system policy; standard operating procedures, as well as the compulsory and rigorous health information management training for information cadres, clinicians, and managers to ensure that accurate data is recorded and reported (NDoH, 2012; NDoH, 2013). Furthermore, facility data audits are also done to assess gaps; regular support visits are conducted by sub-district or district officials as well as supporting health partners to ensure data accuracy (NDoH, 2011). WebDHIS software has analytic tools which are built-in to allow system users to have data extraction and analysis capability (Farnham et al., 2020). The WebDHIS software analytic tools provide users with flexible and more efficient and accurate collection of data at all healthcare system levels with better quality control measures (Kanfe et al., 2021). For this study, WebDHIS software analytic tools will be used for data extraction that covers the actual area of investigation. The WebDHIS software analytic tools reduced measurement bias by extracting similar data from all the health facilities that are reporting routine health information. Selection bias was reduced by randomly selecting health facilities and data elements from the study population.

Reliability evaluates the stability of measures and the internal consistency of the measurement instrument as well as the interrater reliability of instrument scores (Kimberlin & Winterstein, 2008). The reliability of the data extraction tool will be ensured by using the same data extraction tool on health facility data. The WebDHIS data extraction tool is standardized and can produce similar and consistent data when used by any system user. Moreover, the researcher will engage Information Cadres from the Provincial Office to advise on the variables which can or cannot provide reliable measures. Data will be obtained from WebDHIS software which will then be converted into a Microsoft Excel version 2013 spreadsheet by the researcher and checked for duplicates and transcription errors.

3.10 Generalisability

Results and recommendations of the study are generalizable to all public health facilities in the Eastern Cape Province that utilizes the WebDHIS software and possibly to primary healthcare facilities in the country at large, however, the health facilities should have the same or similar data collection and analysis methods, as well as similar staffing levels and training.

3.11 Definition of key concepts

Data completeness refers to the extent to which all required data is available for a given task, representing the complete list of all eligible health facilities and data elements, and not just a fraction thereof (Liu et al., 2020).

Data consistency refers to the coherence of the data being evaluated (Adane et al., 2021). In this study, data consistency will be measured by the data value in a series of values and whether or not it is extreme to the other values in the series. Data element value trends will be compared using the predefined minimum and maximum values, and trends over time.

Data accuracy entails measuring the data against the set information standard and finding it to be correct and trustworthy (Measure Evaluation, 2019).

Routine Health Information refers to data that is generated and collected from healthcare facilities at regular intervals (Ahanhanzo et al., 2015).

3.12 Ethical considerations

Ethical approval from the University of Western Cape Higher Degrees and Humanities and Social Science Research Ethics Committee (HSSREC) was obtained (Appendix B). In addition, an information sheet (Appendix A) and permission letter (Appendix C and D) to conduct the research were received from the ECDoH Epidemiology and Research Office as well as from the Director of Health Information Management Systems to use the WebDHIS software data for research purposes. Minimal harm is expected as the study utilized secondary data reported in the WebDHIS software and all patient data is aggregated. Although the WebDHIS software provides the ability to aggregate information from different data elements and health programs to monitor the progress and milestones, and identify patterns and challenges, it also possess a threat to the privacy of health users and the population.

The confidentiality of the entities providing information used in this study was maintained throughout the research. To uphold confidentiality, pseudonyms were assigned to each health district and health facility. Moreover, the collected data were stored in a password-protected folder which will be kept for at least five years and be destroyed thereafter. Additionally, during the study, the researcher maintained and uphold the principles of the Protection of Personal Information Act (POPIA). The researcher maintained an ethical responsibility towards research data even though the research was a desktop in nature. Kajwang (2022) alluded that desktop studies use secondary data that has been acquired from existing resources. Moreover stated that desktop studies are preferred because they provide reliable information and have wide insights since the data is collected from well-known sources. The dignity and protection of the research stakeholders were upheld during the data extraction and analysis. Integrity was essential to this research, as well as the ethical principle of beneficence, which was applied in all stages of the research. Tripathy (2013) stated that the main concern for use of secondary data revolves around potential harm to participants and the issue of return for consent, and data anonymity must be used to prevent the identification of participants. The data from the WebDHIS software was anonymized as it does not contain the names and/or addresses of the patients whose health information is captured within the dataset. Standardized templates were adopted for data extraction to ensure there was uniformity and that data was collected appropriately. Pseudonyms were used to hide the identity of the health district included in the study. The study will benefit the health information cadres and the healthcare workers in general to enhance the management and provision of healthcare services. Furthermore, no specific health program, data elements, health facilities, and/or health districts were targeted.

3.13 Conclusion

This chapter discussed the research design, the study setting, sampling, and inclusion and exclusion criteria for the study unit of analysis. This was followed by a discussion on the data collection, data analysis, and validity and reliability of data. The chapter ended with generalisability, the definition of key concepts, and ethical considerations. The following chapter presents the study results and data analysis.

Chapter 4: Results and Data Analysis

4.1 Introduction

This chapter presents the results of the assessment of completeness, consistency, and accuracy of the WebDHIS software as used in the Eastern Cape Province. The routine data quality was measured based on the routine data quality framework as shown in Table 3.4. The results of the data quality metrics which were used to measure data completeness, internal consistency, and accuracy are analyzed and presented. The data quality results are presented by district, and disaggregation by the type of health facility. The chapter concludes with a summary of the findings or observations from the results.

4.2 Description of the Study Sample of Records

The dataset extracted from the WebDHIS software included the public health facilities that reported routine health data. The clinics were well-represented in the WebDHIS software with 85.66 percent of the 265 sampled health facilities, followed by hospitals with 8.68 percent, and CHCs with 5.66 percent (Table 4.1). In addition, the dataset included a total of 77 data elements which were expected to be reported by all health facilities as per the NIDS 2017 for the period under review (Table 4.2). Table 4.1 summarise health facilities that were part of the study, and Table 4.2 shows the data elements that were included in the study.

Table 4.1: List of Sampled Health Facilities

Health District	Clinic	CHC	Hospital	Total
District B	19	0	0	**19**
District T	48	2	4	**54**
District D	24	3	1	**28**
District F	52	2	4	**58**
District H	13	0	2	**15**
District P	13	4	1	**18**
District E	37	4	6	**47**
District L	21	0	5	**26**
Grand Total	**227**	**15**	**23**	**265**

Table 4.2: List of Sampled Data Elements

Data Elements Group	Total
ART Monthly	4
Child and nutrition	4
Communicable Diseases	1
EPI	11
Eye Care	1
HIV	16
Malaria	1
Maternal and neonatal	8
Mental Health	3
Non-communicable diseas	5
Oral Health	1
Quality	8
STI	1
TB Monthly	8
Womens Health	5
Grand Total	**77**

In this study, data from WebDHIS software was extracted from April 2017 to March 2020. The financial year period was considered which included 12 months (April to March). A total of 322 532 data element values were reported from all eight districts in the Province. Among the reported data values; there were 102 836 missing data, 4 767 were outliers, 1 280 were data validation errors, and 51 were data marked for follow-up (Table 4.3). The monthly data was extracted separately per health program, and/or district. The data sets were assembled and merged into one dataset for the province. The data were analyzed in the form of descriptive statistics such as frequencies, mean, minimum, or maximum values and presented in the form of tables and graphs. The next sections present results, starting with data completeness, followed by data consistency, and concluding with data accuracy.

Table 4.3: WebDHIS Software Data Extracted

Extracted Data	**2017-18**	**2018-19**	**2019-20**	**Total**
Health Facility Reports	3 156	3 165	3 168	**9 489**
Reported Data Element Values	106 057	105 753	110 722	**322 532**
Missing Data Values	35 326	34 953	32 557	**102 836**
Data Outliers	1 671	1 549	1 547	**4 767**
Data Validation Errors	456	494	330	**1 280**
Data Marked for Follow-up	25	19	7	**51**

During data clean-up, it was noted that the sum of reported data element values and missing data element values do not match the sum of the expected data element values. For this study, a total of 244 860 data element values per year was expected to be reported by all health facilities. The National Indicator Dataset framework clearly states the collection point for each data element, however, the allocation of data element groups in the WebDHIS software is done based on the healthcare services package that is rendered by the health facility (NDoH, 2017a). Also, it was observed that some data element values were blank spaces captured and thus not recorded in the missing report from the WebDHIS software.

4.3 Data Quality Results

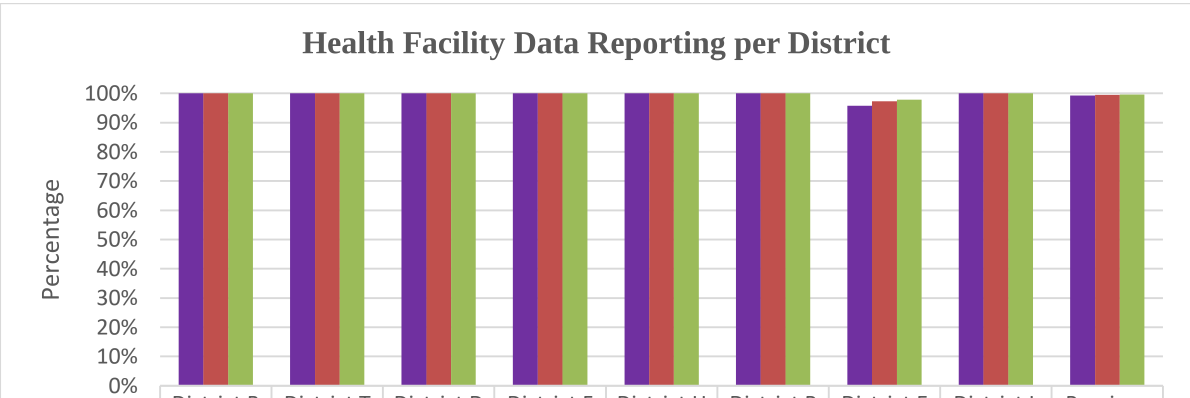
Data quality was assessed on three dimensions namely data completeness, data consistency, and data accuracy. Each dimension had two or more data quality metrics that were measured to assess the data quality. The results are presented below starting with data completeness, followed by data consistency, data accuracy, and finishing with determining the data quality.

4.3.1 Data Completeness

The health facilities are expected to submit reports on a monthly basis to the higher tier level of the health system. Unless stated otherwise, the functional health facilities should 100 percent report all health activities as per the NIDS. Data completeness was assessed using two data quality metrics namely health facility reporting and data element reporting. The completeness of health facility reporting was measured as the number of monthly data received divided by the expected number of reports in a year. Data element reporting was measured as the number of data element values reported divided by the expected number of data element values in a year. The overall rate of health facility reporting was 99.47 percent and data element reporting was 86.00 percent.

4.3.1.1 Health Facility Data Reporting

The public health facilities in the Eastern Cape Province are expected to submit the monthly routine health data to the District and ultimately to the Provincial Office and National Office. The health facility reporting rate for data completeness is defined as the total number of health facility reports received divided by the total number of health facilities expected to report monthly routine data and expressed as a percentage. The study results reveal that health facilities reported data 9 489 times out of 9 540 during the period under review. The health facility data reporting for all three financial years was very good with an average of 99.47 percent. The study results reveal that District E had one clinic that did not report any data in the first and third years, and in the second year, one clinic did not report data for a period of three months. Figure 4.1 shows the health facility data reporting rate per district.



	District B	District T	District D	District F	District H	District P	District E	District L	Province
2017-18	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	95,74%	100,00%	99,25%
2018-19	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	97,34%	100,00%	99,53%
2019-20	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	97,87%	100,00%	99,62%

Figure 4.1: Health Facility Data Reporting Per District

Seven districts managed to maintain 100 percent data reporting throughout the period under review (Figure 4.1). District E was the only district that did not achieve a reporting rate of 100 percent, which ultimately pulled down the provincial overall reporting rate. The overall health facility reporting gradually increased from 99.25 percent in the first year to 99.62 percent in the third year. The data reporting from health facilities showed a high rate of compliance and coverage of health data reporting in the Eastern Cape Department of Health.

The rate of data reporting was high across all health facility types with both CHCs and hospitals achieving 100 percent while the clinics achieved an average of 99.38 percent (Figure 4.2). The study showed that each year, there was at least one clinic that failed to report routine data.

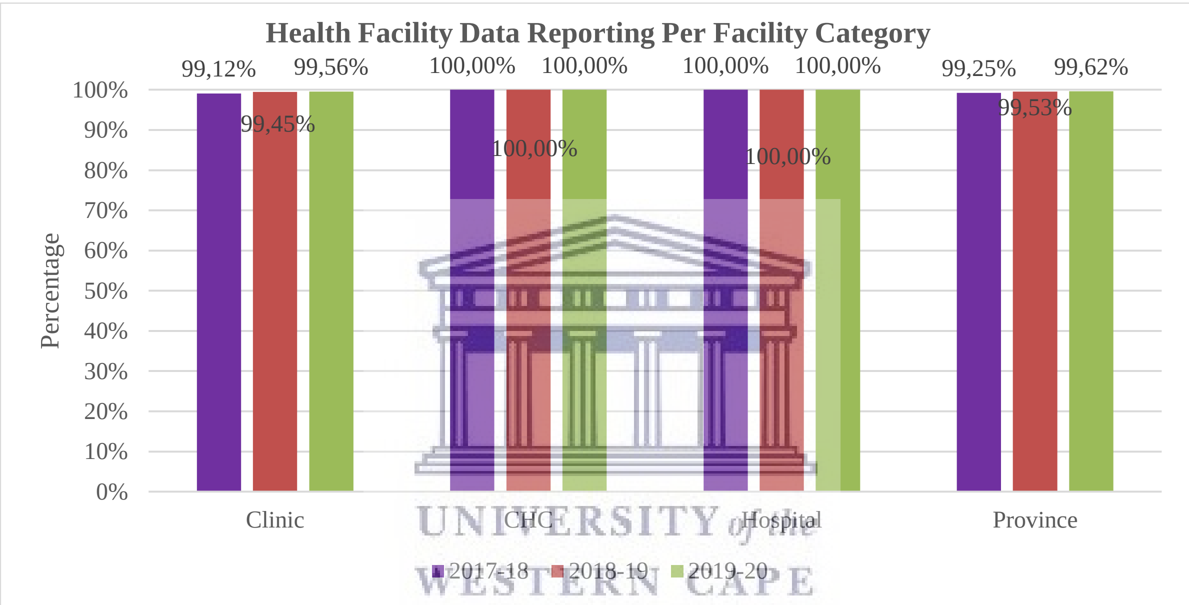
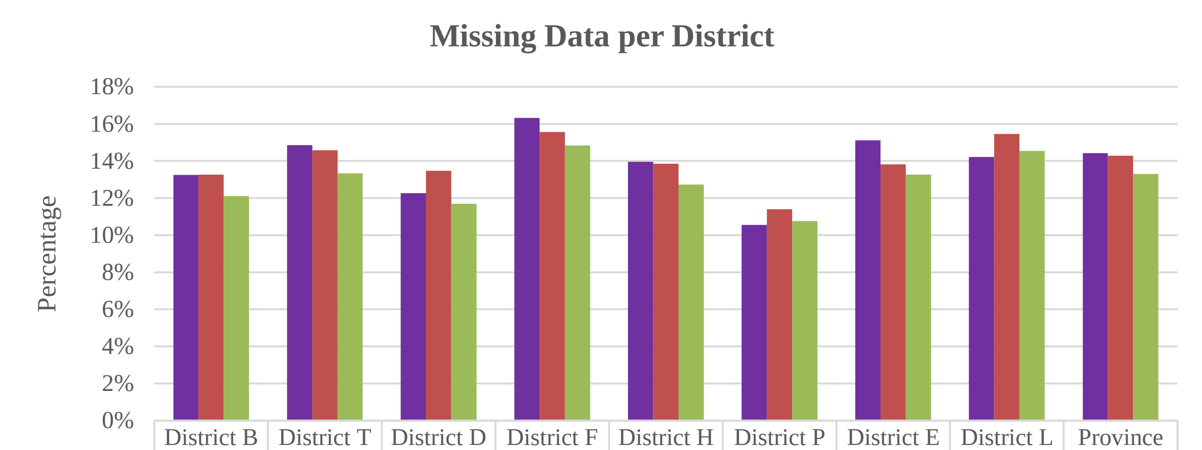


Figure 4.2: Health Facility Data Reporting Per Facility Type

4.3.1.2 Data Element Reporting

The missing data report counts the number of missing values for data elements from the health facilities in the WebDHIS. The missing data report was used to identify data gaps in the reports submitted by the health facilities. The complete reporting of data elements is those that do not have any missing values during the period under review. A total of 102 836 data element values were missing during the period under review. The missing data report showed that the data missing rate was relatively high with an average of 14 percent. Clinics had the highest missing data rate with an average of 14.20 percent and CHCs with the lowest rate at 11.98 percent.



	District B	District T	District D	District F	District H	District P	District E	District L	Province
2017-18	13,25%	14,85%	12,27%	16,33%	13,96%	10,55%	15,11%	14,21%	14,43%
2018-19	13,26%	14,58%	13,47%	15,56%	13,85%	11,39%	13,81%	15,46%	14,27%
2019-20	12,10%	13,34%	11,69%	14,83%	12,73%	10,76%	13,26%	14,54%	13,30%

Figure 4.3: Missing Data per District

The rate of missing data across all eight districts was relatively high with an average rate of 14.00 percent, however, a downward trend was observed with a reduction from 14.43 percent in 2017 to 13.30 percent in 2020. The missing data rate was the highest in District F with 15.57 percent of the expected monthly data, followed by 14.74 percent in District L, and District T at 14.25 percent (Figure 4.3). The districts that achieved the lowest data missing rate include District P with a rate of 10.90 percent and District D at 12.48 percent (Figure 4.3).

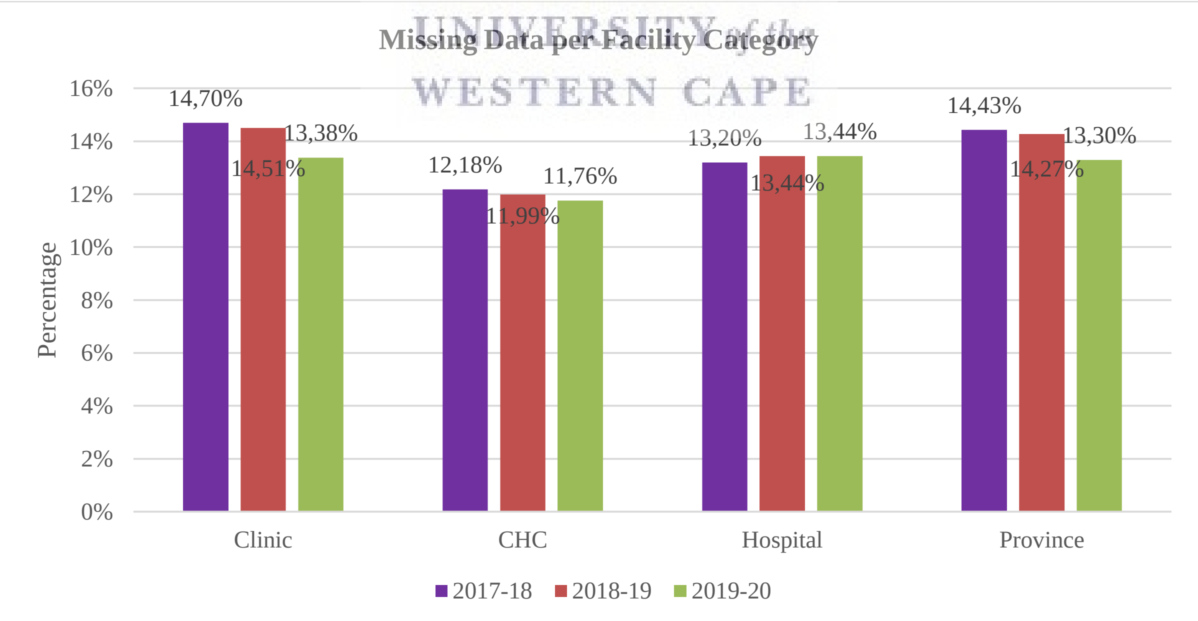


Figure 4.4: Missing Data per Facility Category

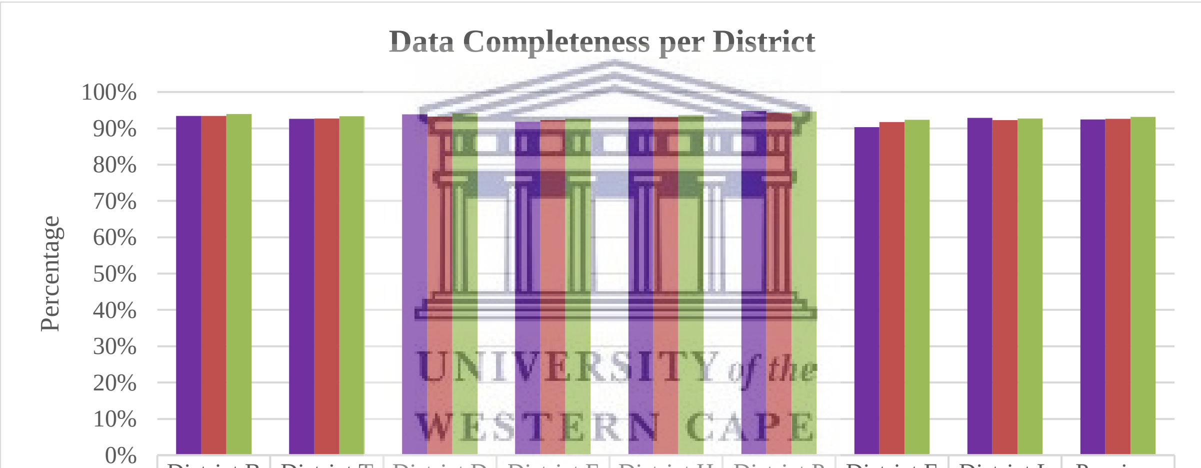
The health facility category that has the highest missing data rate was clinics at 14.20 percent of the expected monthly data from the health facilities, followed by the hospitals at 13.36 percent and the lowest rate of missing data was CHCs at 11.98 percent (Figure 4.4).

To calculate the rate for data completeness, the following formula was used, and Figure 4.5 shows the summary of data completeness per district

Health Facility (HF) Reporting = (HF Reported Data Values / Data Values Expected to be Reported by HF) * 100

Data Element (DE) Reporting = 100 – (Missing DE Value Rate)

Data Completeness = (HF Reporting + DE Reporting) / (2)



	District B	District T	District D	District F	District H	District P	District E	District L	Province
2017-18	93,38%	92,58%	93,87%	91,84%	93,02%	94,73%	90,32%	92,89%	92,41%
2018-19	93,37%	92,71%	93,26%	92,22%	93,07%	94,30%	91,76%	92,27%	92,63%
2019-20	93,95%	93,33%	94,15%	92,59%	93,63%	94,62%	92,30%	92,73%	93,16%

Figure 4.5: Data Completeness per District

The overall data completeness was very good for all districts which continued with an upward trend from 2017 up to 2020. Data completeness was the highest in District P with 94.55 percent of expected monthly health facility reports submitted, followed by 93.76 percent in District D, and District B at 93.57 percent (Figure 4.5). Although all eight districts achieved above 90 percent, there were two districts that recorded the lowest rate for data completeness which included District E at 91.46 percent, and District F at 92.21 percent (Figure 4.5).

4.3.2 Data Consistency

The health facility data is expected to be consistent or grow within the acceptable threshold or standards. Data consistency of the reported data was assessed using two data quality metrics namely outliers and data consistency over time. The outliers were deemed as data errors and quantified from the reported health facility data. Extreme outliers were defined as those data element values that had a modified Z-score of 3.5 and above standard deviation. The second data quality metric involved the quantification of the consistency of data element values that were reported over time. Data element values to be consistent are expected to change within the 33 percent as recommended by WHO. The overall rate of outlier and data consistency over time was 1.48 percent and 25.41 percent respectively. Table 4.1 shows the results of the outlier, followed by Figure 4.6 shows the results of the data consistency over time and concludes with Figure 4.7 which shows the overall results of data consistency.

4.3.2.1 Outliers

This data quality metric assesses whether a data value in a series of values is out of range in relation to the other values in the series (WHO, 2017). The study used extreme outliers to quantify data consistency. Extreme outliers are defined as the monthly values that vary with a modified Z-score of 3.5 and above standard deviations within the expected values of each data element (WHO, 2017). Table 4.4 presents the frequency of extreme outliers in the reported health facility data.

Table 4.4: Outlier per District

Health District	**2017-18**	**2018-19**	**2019-20**
District B	1,89%	1,74%	1,33%
District T	1,39%	1,31%	1,21%
District D	1,42%	1,37%	1,25%
District F	1,53%	1,39%	1,39%
District H	1,53%	1,45%	1,50%
District P	1,90%	1,64%	1,70%
District E	1,79%	1,74%	1,59%
District L	1,38%	1,22%	1,37%
Province	**1,58%**	**1,46%**	**1,40%**

Table 4.4 shows that the rate of the data element value that had an extreme outlier from the reported data was relatively low with an average rate of 1.48 percent across all eight districts. The downward trend in the rate of data outliers was observed in six districts namely District B,

District T, District D, District F, District H, and District E. The study result showed an upward trend of data outliers in District P and District L. The overall rate was on the decline which showed an improvement in the quality of data reported by health facilities.

4.3.2.2 Data Consistency Over Time

Consistency over time assessed the plausibility of reported results for health programs in terms of the history of reporting routine data element values (WHO, 2017). The WHO framework for routine data quality recommends that data consistency over time should be measured using four years of data, however, this study used three-year data due to study parameters and the lifespan of the NIDS 2017 (WHO, 2017; NDoH, 2017a). The financial year 2019-20 data is compared to the mean of the two preceding years. Figure 4.6 presents the frequency of data element values that were above the acceptable consistency over time in the reported health facility data.

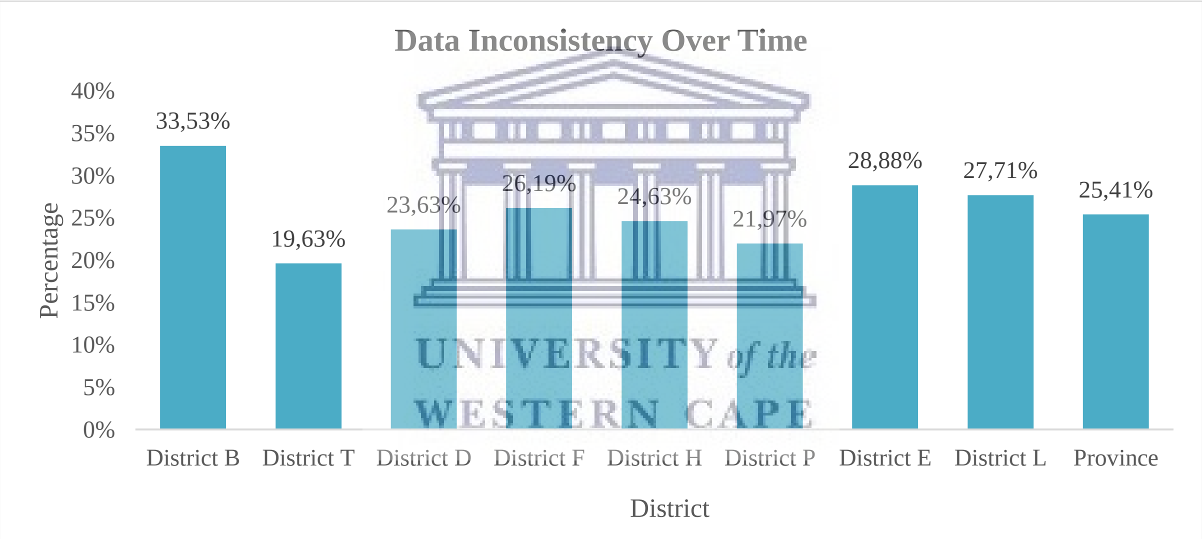


Figure 4.6: Data Consistency over Time per District

Figure 4.6, shows the result of data element value changes that were above the acceptable threshold of plus or minus 33 percent as recommended by WHO. The overall rate of data inconsistency over time for the reported health facility data was a bit higher with an average of 25.41 percent. The districts that had the lowest rate include District T at 19.63 percent, followed by District P at 21.97 percent, and District D at 23.63 percent. Two districts that recorded a high percentage that is above the province average which District B with 33.53 percent, and District E with 28.88 percent.

To calculate the rate for data consistency, the following formula was used, and Figure 4.7 shows the summary of data consistency per district.

Data Values without Outliers = 100 – (Data Values with Outliers)

Data Consistency over Time = 100 – (Data Values Inconsistency Over Time Rate)

Data Consistency = (Data Values without Outliers + Data Consistency over Time) / (2)

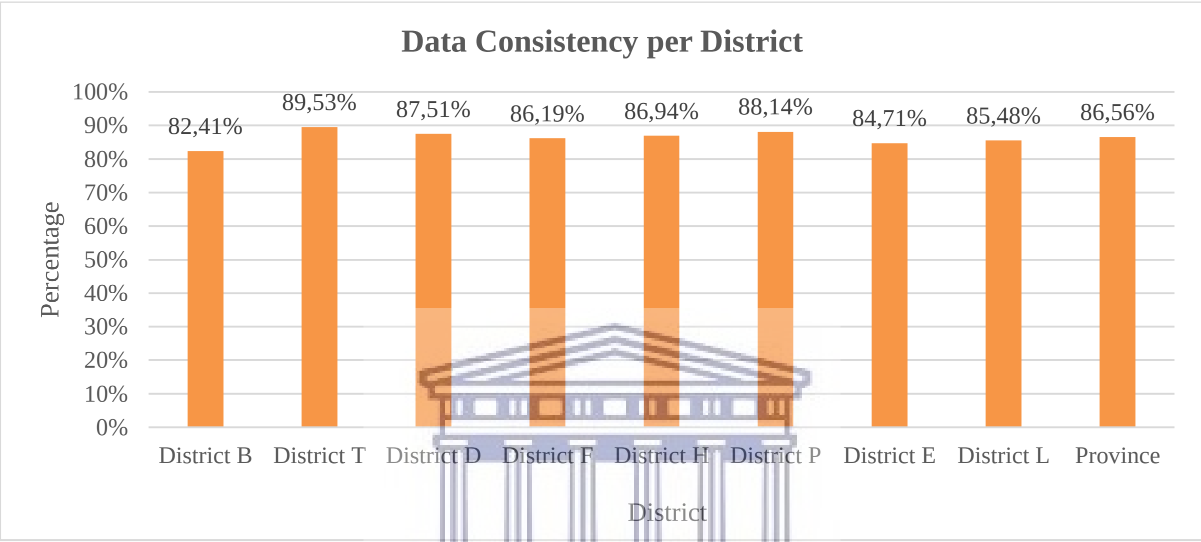


Figure 4.7: Data Consistency per District

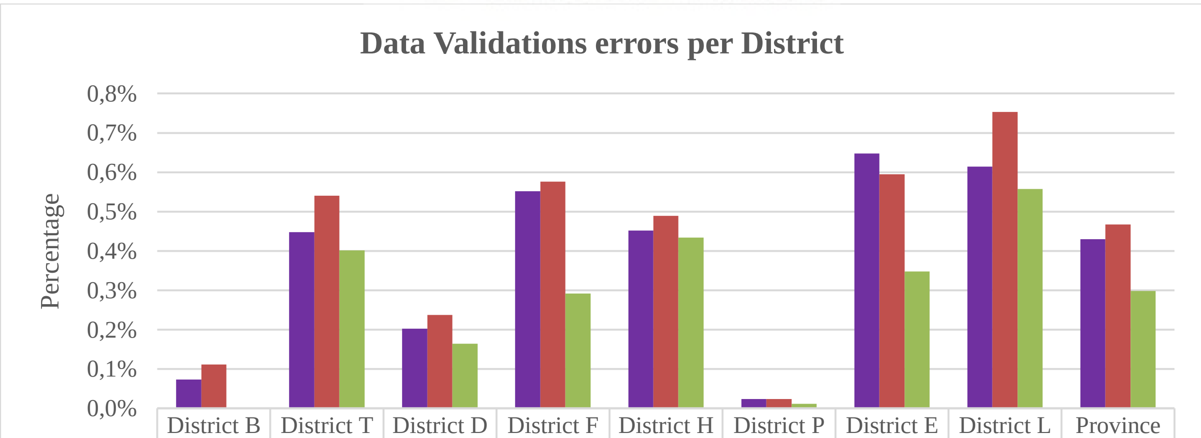
In Figure 4.7, the overall rate of internal data consistency for the reported data element values showed a good performance with an average of 86.56 percent. Data consistency was the highest in District T with 89.53 percent in the reported monthly health data, followed by 88.14 percent in District P, and District D at 87.51 percent (Figure 4.7). The district that had the lowest rate was District B with 82.41 percent, followed by District E at 84.71 percent and District L at 85.48 percent. The consistency of data reported in the WebDHIS software was good, however, the study revealed a high number of data element values that were above the acceptable ratio for appropriate data changes, and this phenomenon was observed in all eight districts.

4.3.3 Data Accuracy (Internal Data Cross-checking)

Internal data cross-checking dimension was used to assess the accuracy of data captured in the WebDHIS software for the period under review. The accuracy of the reported data was assessed using three data quality metrics which included data validation errors, data marked for follow-up, and missing data values. The data validation errors were defined as data element values that do not correspond with the other data element values as per the predefined data validation rules in the system. The data marked for follow-up was defined as the data element value that was flagged with or without the comments to be followed up for data correctness. The missing data values were defined as the number of missing values for data elements from the reported health facility data. Figure 4.8 show the results for data validation errors, followed by results of data marked for follow-up in Table 4.2, then Figure 4.9 shows results for missing data, and concludes with Figure 4.10 shows the overall results for data accuracy.

4.3.3.1 Data Validation Errors

The data validation measures the data violations based on the pre-defined validation rules which are expressed by the conditions set between data elements. The overall rate of data validation errors was lower in all districts with an average of less than 0.5 percent which is an excellent showing of good quality data. The district that had the highest validation errors was District E and District L with an average of 0.64 percent and 0.53 percent respectively. Three districts that reported the lowest number of validation errors include District P with an average of 0.02 percent, followed by District B at 0.06 percent, and District D at 0.20 percent (Figure 4.8). The hospitals were responsible for reporting monthly data that has the highest validation errors with an average of 3.46 percent for the period under review. Hospitals had the highest data validation errors in the year 2018/19, and the clinics as well as CHCs showed a slight downward decline.



	District B	District T	District D	District F	District H	District P	District E	District L	Province
2017-18	0,07%	0,45%	0,20%	0,55%	0,45%	0,02%	0,65%	0,61%	0,43%
2018-19	0,11%	0,54%	0,24%	0,58%	0,49%	0,02%	0,60%	0,75%	0,47%
2019-20	0,00%	0,40%	0,16%	0,29%	0,43%	0,01%	0,35%	0,56%	0,30%

Figure 4.8: Data Validation errors per District

4.3.3.2 Data Marked for Follow-up

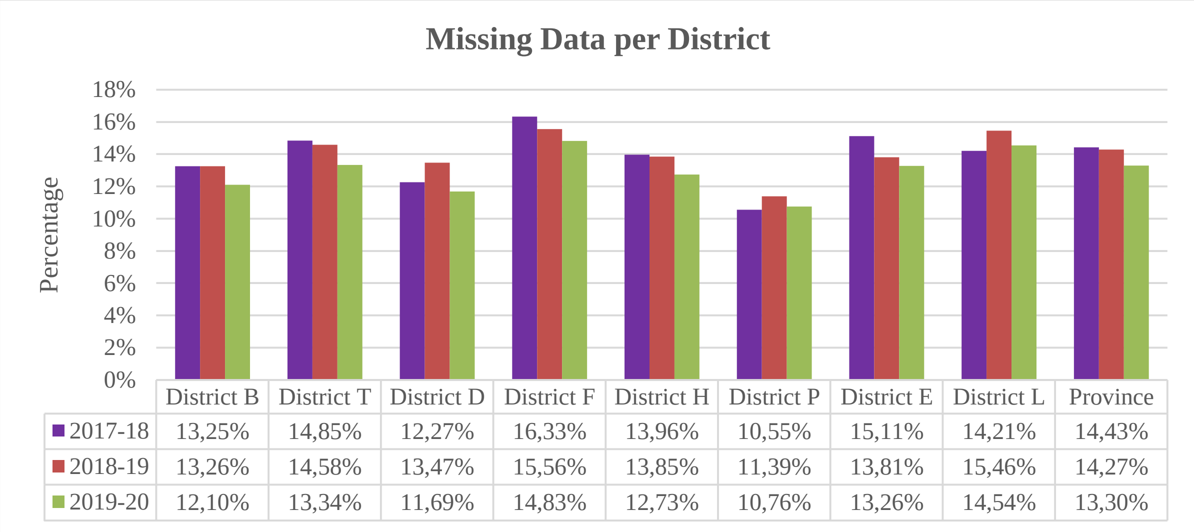
When assessing the extent of data accuracy it is expected that the reported data should not have any unresolved queries. Data marked for follow-up quantify the number of data element values that are flagged for correctness. The districts that recorded the highest number of data element values flagged include District E with an average of 0.04 percent, followed by District T, and District P with 0.02 percent respectively (Table 4.5). District H, District B, District D, and District L had the lowest data element values flagged for corrections with an average of 0.013 percent while District F had zero data element values flagged (Table 4.5).

Table 4.5: Data Marked for Follow-up Report

Health District	2017-18	2018-19	2019-20
District B	0,04%	0,00%	0,00%
District T	0,06%	0,01%	0,00%
District D	0,00%	0,01%	0,02%
District F	0,00%	0,00%	0,00%
District H	0,00%	0,02%	0,00%
District P	0,01%	0,04%	0,00%
District E	0,04%	0,07%	0,02%
District L	0,01%	0,00%	0,00%
Province	**0,02%**	**0,02%**	**0,01%**

4.3.3.3 Missing Data

The missing data report counts the number of missing values for data elements from the health facilities in the WebDHIS. The missing data report is used to identify data gaps in the reports submitted by the health facilities. The complete reporting of data elements is those that do not have any missing values during the period under review. The missing data report showed that the data missing rate was relatively high with an average of 14 percent. Clinics had the highest data missing rate with an average of 14.20 percent and CHCs with the lowest rate at 11.98 percent.



	District B	District T	District D	District F	District H	District P	District E	District L	Province
2017-18	13,25%	14,85%	12,27%	16,33%	13,96%	10,55%	15,11%	14,21%	14,43%
2018-19	13,26%	14,58%	13,47%	15,56%	13,85%	11,39%	13,81%	15,46%	14,27%
2019-20	12,10%	13,34%	11,69%	14,83%	12,73%	10,76%	13,26%	14,54%	13,30%

Figure 4.9: Missing Data per District

The rate of the missing data values across all eight districts was relatively high with an average rate of 14 percent, however, a downward trend was observed with a reduction from 14.43 percent in 2017 to 13.30 percent in 2020. The missing data rate was the highest in District F with 15.57 percent of the expected monthly data, followed by 14.74 percent in District L, and District T at 14.25 percent (Figure 4.9). The districts that achieved the lowest data missing rate include District P with a rate of 10.90 percent and District D at 12.48 percent (Figure 4.9).

To calculate the rate for data accuracy, the following formula was used, and Figure 4.10 shows the summary of data accuracy per district.

Data Values without Validation errors = 100 – (Data Values with Validation errors)

Data Values not Marked for Follow-up = 100 – (Data Values Marked for Follow-up)

Data Element (DE) Reporting = 100 – (Missing DE Value Rate)

Data Accuracy = (Data Values without Validation errors + Data not Marked for Follow-up + DE Reporting) / (3)

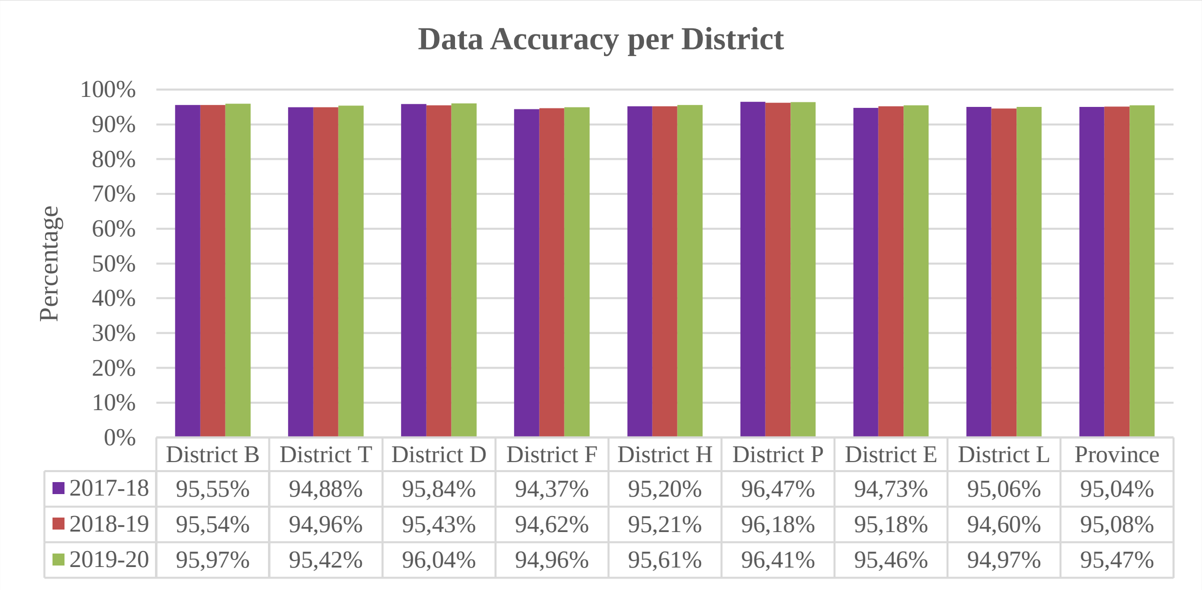


Figure 4.10: Data Accuracy per District

In Figure 4.10, the overall rate of data accuracy for the reported data elements values showed a very good performance with an average of 95.34 percent. Data accuracy was the highest in District P with 96.36 percent in the reported monthly health facility, followed by 95.69 percent in District B, and District D at 95.77 percent (Figure 4.10). The accuracy of data captured in the WebDHIS is high when assessed using the above-mentioned data metrics, however, a worrying concern with the data missing shows a high rate with an average of 14.00 percent in all eight districts (Figure 4.10).

4.3.4 Data Quality

In this study, the data quality was assessed based on three dimensions namely; completeness, consistency, and accuracy. Each dimension had two or more data quality metrics that were reviewed and quantified to assess the quality of data in the WebDHIS software. The calculation of scoring results for data quality dimensions was done as shown in Table 4.6.

Table 4.6: Calculation of Data Quality Dimensions

Data Quality Dimensions	Calculations
Data Completeness	Average Rate of HF Reporting, and DE Reporting
Data Consistency	Average Rate of DE values without Outliers, and DE values that are Consistency Over Time
Data Accuracy	Average Rate of DE without Validation Errors, DE values not Marked for Follow-up, and DE without Missing Values

To calculate the rate for data quality, the following formula was used, and Table 4.7 shows the summary of data accuracy per district.

Data Quality = (Data Completeness + Data Consistency + Data Accuracy) / (3)

The overall results for data quality in the WebDHIS software were at 91.5 percent for the period under review (Table 4.7). Two data quality dimensions were graded above 90 percent, which was rated as a very good performance. The data completeness was assessed using health facility data reporting and data elements reporting. The overall score on data completeness was 92.7 percent, which was rated as very good (Figure 4.5). Missing data values had about 14.00 percent which was the highest contributing factor to reducing the rate of data completeness (Figure 4.3). The data consistency was assessed using an outliers report, and the data consistency over time and the overall score rating was 86.6 percent (Figure 4.7). A high rate of 25.41 percent for data element values that were above the acceptable threshold for consistency over time contributed to the reduction of the data consistency rate (Figure 4.6). The data accuracy was assessed using data validation errors, data marked for follow-up, and missing data values. The overall score for data accuracy was very good with a rating of 95.2 percent (Figure 4.10). Table 4.7 presents a summary of data quality assessment results per district.

Table 4.7: Data Quality status of Health Facilities per Health District

Health District	Completeness	Consistency	Accuracy	Data Quality
District B	93,6%	82,4%	95,7%	90,6%
District T	92,9%	89,5%	95,1%	92,5%
District D	93,8%	87,5%	95,8%	92,3%
District F	92,2%	86,2%	94,7%	91,0%
District H	93,2%	86,9%	95,3%	91,8%
District P	94,6%	88,1%	96,4%	93,0%
District E	91,5%	84,7%	95,1%	90,4%
District L	92,6%	85,5%	94,9%	91,0%
Province	**92,7%**	**86,6%**	**95,2%**	**91,5%**

The study results revealed that the health facility data completeness was very good at 92.7 percent, however, data reported in the WebDHIS software had a concern with the high rate of 14.00 percent for missing data element values compared to the expected values in the information system as per NIDS requirements. Moreover, the missing data were also observed as a challenge, during the data clean-up a high number of data element values were blank but the missing data reported from the WebDHIS software did not include them. This unaccounted or misrepresented data can contribute more to missing data values which can result in decreasing data completeness and data accuracy.

The results of the assessment revealed good data quality with internal data consistency, however, the data element values that were not consistent over time were high at 25.41 percent in the reported data during the period under review. The study results showed that the data accuracy in the WebDHIS software was very good with fewer data errors in the reported health facility data. The study results showed a very good outcome of data quality at 91.5 percent, and all three data quality dimensions had an upward trend in the reported data in WebDHIS software.

4.4 Conclusion

This chapter presents the study results, starting with a description of sampled records, followed by the results of data completeness which included health facility data reporting, and data element reporting, then followed by data consistency which included outliers and data consistency over time, and data accuracy which included validation errors, data flagged for follow-up, and missing data. The chapter the results of the data quality. The following chapter discusses the study results and identifies the findings.

Chapter 5: Discussions

5.1 Introduction

This chapter explains the study findings in relation to the available evidence gathered in the literature review. The chapter has been arranged according to a sequence of study objectives. The chapter starts with data completeness and then moves on to discuss data consistency, then followed by data accuracy, and closes with data quality of the reported health data.

5.2 Data Completeness

Data completeness is one of the dimensions of data quality used and was assessed by comparing the number of health facilities that reported routine data with the number of the health facilities excepted to reported routine data, and also comparing the number of missing data elements values with the number of expected data element values in the WebDHIS software. In this study, the overall data completeness was found to be 92.7 percent, which is in a similar range to the countries that implement RHIS as the data source. The study results are similar to the findings of the study conducted in Guinea which found data completeness for disease surveillance to be at 98.5 percent in the WebDHIS software (Reynolds et al., 2022). A study in Senegal found that the completeness of facility reporting was 97.5 percent in the WebDHIS software for malaria data (Muhoza et al., 2022). A study conducted by Kiberu et al. (2014) affirmed that the implementation of WebDHIS software helps to improve completeness in reporting routine health data from the health facility to the national level.

To assess the data completeness, two data quality metrics were used. The study results showed that facility reporting was very good with an overall mean of 99.62 percent which was well above the minimum threshold of 75 percent as recommended by WHO (WHO, 2017a). The CHCs and hospitals maintained a 100 percent reporting performance, whereas clinics were slightly down at 99.38 percent. The study conducted in Bangladesh found that there was an increase in the completeness of data reporting over time and associated this performance with the strong commitment from the authorities, extensive support, and positive attitudes from the healthcare workers (Begum et al., 2020). Another study conducted in the Democratic Republic of Congo found data reporting from health facilities to be good which ranged between 83.3 percent and 93.2 percent (Malembaka et al., 2021).

Another data quality metric used was data element reporting which was measured by assessing the number of missing data element values from the reported facility data. The overall missing data scored a rating of 14.00 percent which was above the tolerance level of less than 10 percent as recommended by the WHO in the data quality assessment framework (WHO, 2017a). The study result revealed a better achievement when compared to a study conducted in Ghana which had a score of 27.6 percent of the reported monthly facility data (Nsiah et al., 2022). The high missing data contributed immensely to the reduction of the rate for data completeness reported in the WebDHIS by the health facilities. This occurrence is also observed in the study conducted in Nigeria in the state of Gombe which found facility-reported data in WebDHIS software to be incomplete by at least 40 percent of the events documented in the facility registers (Bhattacharya et al., 2020). The high number of missing data in the WebDHIS software revealed the data capturing challenge and/or inappropriate management of the database, wherein the health facilities are not properly assigned with the data element groups.

The findings on data completeness rate which is a higher scoring prove that health facilities are reporting health data and are compliant with NIDS requirements. However, missing data was relevantly high as compared to a recommendation of less than 10 percent by the WHO (WHO, 2017a). The high rate of missing data can be improved through a more proactive approach of regular data verification activities from the health facility to the district level.

5.3 Data Consistency

Data consistency was measured using two data quality metrics namely frequency of outliers and data consistency over time. An outlier analysis report was used to assess the reported data element that had extreme values. Secondly, the routine data was used to assess the data consistency over time to determine the extent of consistency of the reported data element values by comparing the reference year to the mean value of the preceding two years. In this study, the overall data consistency was found to be 86.6 percent, and the districts that had the lowest score of 84.7 percent and the highest achieved was 89.5 percent. The study results showed a good performance for data consistency which was a similar observation from the study conducted in Ethiopia which found data internal consistency to be good in the routine health information system (Adane et al., 2021).

Outlier is one of the most important metrics to measure data consistency in the WebDHIS software. In the study, extreme outliers were considered wherein a data element value recorded a standard deviation of more than three and a half from the mean value. The overall results for the outliers were low at 1.48 percent in the reported health facility data. When study results were disaggregated by health facility type, CHCs had more outliers at 1.81 percent, and hospitals had the least number at 1.36 percent. The low presence of outliers is similar to the study conducted in Senegal which found the frequency of extreme outliers at 2.3 percent in the reported malaria data in the WebDHIS software (Muhoza et al., 2022). Another study conducted in Uganda found that the presence of extreme outliers was rare in the reported health data with National and Sub-regions recording zero extreme outliers, and districts recording only 2 values with outliers in a single year (Agiraembabazi et al., 2021).

The second data quality metric used to measure data consistency was consistency over time, which evaluated the ratio of data element values for the financial year 2019/20 to the mean of the two preceding years. The study considered the ratio of data element values which were 1.33 and above as not consistent. The study results found data element values not consistent over time at 25.4 percent for all eight districts. The study results also revealed that hospitals had a higher rate of data not consistent over time at 29.69 percent and the CHCs had the lowest rate at 21.29 percent. This finding was in disagreement with the study conducted in Kenya which found that data consistency over time was good, and highlighted the important role of using WebDHIS software to monitor data quality (Maina et al., 2017). Another study conducted in Cape Coast Metropolis in Ghana found the level of data consistency over time to be good at 7 percent for the reported maternal and child health data (Lasim, Ansah, & Apaak, 2022).

The finding on data consistency was the lowest score rating among the three data quality dimensions used. Notable, the data inconsistency over time had a high scoring rate which contributed more to the reduction of the overall rate for data consistency. The study finding on data consistency highlights the ineffective use of the data quality check features in the WebDHIS software to monitor and correct the identified data errors. WebDHIS software has standardized data quality features to check data inconsistencies and outliers which makes it easier for information cadres and managers to identify problems and make necessary corrections or add meaningful comments. A study that was conducted in Kano State in Nigeria, suggested that data consistency can be improved through staff training on how to use data tools correctly (Akerele et al., 2020). To improve data consistency training of health workers and managers must be conducted, and the use of data quality features in the WebDHIS software.

5.4 Data Accuracy (Internal Data Cross-checking)

Internal data cross-checking was used to assess the accuracy of data. Three data quality metrics were included namely data validation errors, data element values marked for follow-up, and missing data element values. The study analyzed the frequency of the data errors from the reported health facility data in the WebDHIS software. The study results indicate a good score rating for data accuracy with a rating score of 95.2 percent for all eight districts. This scoring is similar to the finding by Nicol, Dudley, and Bradshaw (2016) from the study conducted in two districts in South Africa and showed a similar trend in data accuracy with 84 percent for the prevention of mother-to-child transmission of HIV program.

The contradictory results were shown in the study conducted in Ethiopia which found that only half of the health facilities reported accurate data in the health information management system (Endriyas et al., 2019). The challenge of data inaccuracy was also reaffirmed by the results study conducted in Tanzania, which found that about 26.4 percent of the reported data had discrepancies (Kabakama et al., 2016). The contrasting findings on the data accuracy in the RHIS affirm the view that organizational, behavioural, and technical factors had in the performance of the country RHIS (Lemma et al., 2016).

Three data quality metrics were used to measure data accuracy and the overall results showed a low number of data element values that had errors with an exception of missing data values which had a higher rate. The overall rating score for data validation errors, data marked for follow-up, and missing data was 0.40 percent, 0.02 percent, and 14.00 percent respectively. The low rate results on data validation errors and data marked follow-up showed a similar trend to the result of the study conducted in Ghana found data errors between WebDHIS and primary source data to be low ranging from 0.0 to 4.9 percent (Amoakoh-Coleman et al., 2015). In contrast, a study conducted by Nsiah et al. (2022) found a data discrepancy of 13.5 percent between the primary source and WebDHIS software. The study conducted in Botswana found a data discrepancy of over 10 percent in the reported data elements which affirmed the concern about high data errors in the WebDHIS (Tlale et al., 2019).

Another data quality metric assessed involved quantifying the frequency of missing data in the reported data from health facilities. The results score for missing data was a bit higher at 14.00 percent which was above the tolerance level of less than 10 percent as recommended by the WHO (WHO, 2017a). The study result showed a similar trend when compared to a study conducted in the Democratic Republic of Congo which found a decrease in data missing over time for health facilities, with health posts at 20 percent and 5 percent for health centres and hospitals respectively (Feng et al., 2021).

The high rate of missing data was also found to be a contributing factor to poor data quality in the study conducted in Kenya which revealed a data gap for reported malaria data across all health facilities (Githinji et al., 2017). This finding on data missing was also observed in the study conducted in Nigeria which found improvement in the data reporting, however, the incidence of data missing had a rating score of 10 percent and above of the reported monthly facility data (Shuaib et al., 2020). The missing was the only data quality metric that had the biggest influence on the data accuracy in the WebDHIS software.

The finding of a high rating score for data accuracy was similar to the study findings from the study conducted in Kenya which found data accuracy to be high in the reported data (Manya & Nielsen, 2016). The study finding affirms the credibility and trustworthiness of the reported data in the WebDHIS software. However, the missing data that had a score rated above the accepted tolerance level as recommended by WHO, showed a serious challenge in the data-capturing process and associated activities. Another issue that was observed was the discrepancy in the reported data when compared to the expected data from the health facilities as per NIDS (NDoH, 2017a). This observation was linked to system administration and management, wherein the health facilities are not properly assigned to the data element groups.

5.5 Data Quality

Quality of data is an important aspect of generating reliable and valid health information that enables authorities and healthcare workers to monitor performance and make appropriate decisions for continuous improvement (Blödt et al., 2018). This study evaluated the quality of routine health data in the WebDHIS software as reported by health facilities in the Eastern Cape Province. The study results revealed a very good rating of the reported data from the health facilities in the WebDHIS software. The overall result for data quality was rated 91.5 percent in the WebDHIS. A study conducted by Getachew, Erkalo and Garedew (2022) found data quality at 83 percent which affirmed a good rating for data in the HMIS. In contrast, to study results, a study conducted in Ethiopia by Shama et al. (2021) found good quality data to be at the lowest rate of 51.35 percent for reported data from public health facilities. Another study that showed a low level of good quality data was conducted by Teklegiorgis et al. (2016) found data quality to be only 75.3 percent in the HIMS as reported from the health facilities.

The data quality was measured using three dimensions namely data completeness, data consistency, and data accuracy. The overall results of data completeness, data consistency, and data accuracy level in the study were 92.7, 86.6, and 95.2 percent respectively. All data quality dimension results were rated above 75 percent which was regarded as good performance. The fewer data errors in the RHIS are attributed to interventions such as HIMS training of health workers to increase their knowledge and skills as well as supportive supervision and feedback (Shama et al., 2021; Kanfe et al., 2021). Although the results of the data quality had a rating score that was very good, the study revealed data quality metrics that had a bad outcome. The study results showed the reported data in the WebDHIS software had a high percentage of missing data which was above the tolerance level of less than 10 percent as recommended by WHO (WHO, 2017a). Also, the study revealed a high percentage of data element values that were not consistent over time, the WHO recommends a threshold of less than 33 percent for data changes over time (WHO, 2017a).

Good quality data not only helps to encourage data utilization but makes it easy for managers to take informed decisions pertaining to the planning and management of healthcare services (Mboera et al., 2021; Bhattacharya et al., 2020). This study results showed that the ECDoH had good quality data since the adoption of WebDHIS software which put the department in a favourable position to ensure that appropriate decisions are taken, and regular monitoring and evaluation of health activities are undertaken (Lutge et al., 2016; Maïga et al., 2019). The findings from the study (Hung et al., 2020; Jinabhai et al., 2021) concurred with the critical role of good data quality which found that RHIS data was not only used as a management support tool but was used to conduct program evaluation, monitoring and assessing services, and epidemiological research purposes. Another study conducted by Amouzou et al. (2021) also affirmed the importance of good data quality and found RHIS data to be useful to support planning processes and the performance assessment of the implemented health plans.

Also, good data quality helps to strengthen accountability and transparency as important principles of management. In South Africa, Auditor General is constitutionally mandated to strengthen the oversight, accountability, and governance in the public sector through auditing the performance and conduct of public institutions (RSA, 2004). Auditor General needs good-quality data from public institutions in order to ensure that the outcomes of the assessment are accurate and reliable (AGSA, 2022). A study conducted by Hilber et al. (2016) asserts that collected information is useful in strengthening accountability by providing evidence-grounded context and understandable details. Data quality enables appropriate management and ensures there is transparency and accountability in decision-making.

To conclude, during the study, the researcher observed a concern regarding the NDoH Data Dictionary database which is an official and public platform for the storing of NIDS and a list of public health facilities in South Africa. The database does not allow data extraction by Microsoft Excel or CSV format which are essential tools to sort and analysis of big data. Lastly, the study followed the WHO routine data quality framework as a guide, however, not all data quality dimensions were used to evaluate reported data. Also, private and non-government-owned health facilities were not included in this study.

5.6 Conclusion

In summary, this chapter discussed the study results and elaborated on the observed study findings in terms of data completeness, data consistency, data accuracy, and data quality. The study also highlights the observation about the NDoH Data dictionary website, and data quality assessment framework used as the guide for assessment. The discussion in this thesis acknowledges that different measurements have been utilised to determine data quality in various studies however this thesis compares the outcomes with that in mind. The following chapter presents the conclusions and makes recommendations regarding the data quality of routine health information systems.

Chapter 6: Conclusions and Recommendations

6.1 Introduction

This chapter provides an overview of the findings of the study as they were presented and reviewed in the previous chapter and also makes recommendations to attempt to find a solution to the set research problem. The general aim of the research was to assess the data quality in the routine health information system used in the Eastern Cape Province. The findings indicated there is a significantly high level of good-quality data in the routine health information system. This chapter starts with a summary of the thesis, followed the study findings, and closes with recommendations.

6.2 Summary of Chapters

The *first chapter* of the study focused on the background and rationale, research problem, research question, and objectives of the study. Moreover, the chapter discussed the significance of the study as well as the structure of the thesis. The *second chapter* focused on the literature review about health information systems, and the frameworks used to assess the data quality in the routine health information systems, and ends with a discussion on the factors that are influencing data quality in the routine health information systems. In *chapter three*, an overview of the research methodology was discussed which started with research design, study population, sampling method, and sample size, and the description of the study site. In addition, the chapter discussed the procedure for data collection, the technique used for data analysis, and the definition of key data elements and closes with ethical considerations.

Chapter four presented and summarised study results as per data quality dimensions that were aligned to answer the study questions. *Chapter five* discussed the study findings per data quality metrics. The study findings were discussed and critically analysed by comparing them with the existing literature. *Chapter 6* starts with the introduction, then makes a summary of chapters, and is followed by conclusion remarks on the findings regarding the research questions. This chapter closes with recommendations to maintain and/or improve data quality.

6.3 Findings

This study evaluated the data quality in the routine health information system as reported by the health facilities in Eastern Cape Province. The quality of the reported health facility data in the routine health information system was found to be very good in the Eastern Cape Province. Missing data was found to be high in the reported health data in the WebDHIS software. The high number of data element values were not captured in the system which was prevalent across all eight districts. The study also found inconsistent data over time in the reported data from health facilities. Over twenty-five percent of data element values changed above the acceptable tolerance level as recommended in the WHO data quality framework. Inconsistent data element values over time significantly contributed to reducing the rate of data consistency in the WebDHIS software. The challenge of missing data element values and inconsistent data over time were higher in the clinics than in hospitals and community health centres. Although the data quality issues were found in the health facility data they did not adversely affect the overall data quality, however, their presence poses a threat to the reliability of data in the WebDHIS software.

6.4 Recommendations

This study provided an overview of the data quality reported in the WebDHIS software from the health facilities in the Eastern Cape Province. The study results revealed the component of routine health information management that needs serious attention. Based on the findings of this study, the following recommendations are suggested:

a. There is a need for regular analysis of the WebDHIS software data quality reports to track progress and identify areas for continuous improvement in the reported data. The institutionalization of the data quality review in all tier levels of the health system.

b. Review data consistency trends for a longer period of 4 years and above in order to determine if the higher rate of data not consistent is accurate.

c. To strengthen the data capturing and reporting process of the routine health data in health facilities. Continuously conduct training and capacity building of information cadres and managers in all tier levels of the health system.

d. To strengthen the management of the routine health information system, especially the allocation of data element groups to health facilities for data collection. It is recommended that the system administrator should assign all data element groups to health facilities as per NIDS requirements.

e. The good quality data produced by health facilities and districts should be recognized in a non-financial incentives program such as awarding certificates to the best performers. The recognition event for handing certificates can be done on a semester or annual bases to encourage the performance of staff members working on producing and reporting health data.

Bibliography

Adane, A., Adege, T. M., Ahmed, M. M., Anteneh, H. A., Ayalew, E. S., Berhanu, D., Berhanu, N., Beyene, M. G., Bhattacharya, A., Bishaw, T., Cherinet, E., Dereje, M., Desta, T. H., Dibabe, A., Firew, H. S., Gebrehiwot, F., Gebreyohannes, E., Gella, Z., Girma, A., Janson, A. (2021). Routine health management information system data in Ethiopia: consistency, trends, and challenges. *Global Health Action*, 14(1). https://doi.org/10.1080/16549716.2020.1868961

Agiraembabazi, G., Ogwal, J., Tashobya, C., Kananura, R. M., Boerma, T., & Waiswa, P. (2021). Can routine health facility data be used to monitor subnational coverage of maternal, newborn and child health services in Uganda? *BioMed Central Health Services Research*, 21(Suppl 1), 1–10. https://doi.org/10.1186/s12913-021-06554-6

Ahanhanzo, Y. G., Ouendo, E. M., Kpozèhouen, A., Levêque, A., Makoutodé, M., & Dramaix-Wilmet, M. (2015). Data quality assessment in the routine health information system: An application of the Lot Quality Assurance Sampling in Benin. *Health Policy and Planning*, 30(7), 837–843. https://doi.org/10.1093/heapol/czu067

Akerele, A., Obansa, R., Adegoke, O. J., Haladu, S., & Adeoye, O. B. (2020). Determinants of improved data consistency across routine immunization data tools for health facilities in. *Pan African Medical Journal*, 36(148). https://doi.org/10.11604/pamj.2020.36.148.20498

Amoakoh-Coleman, M., Kayode, G. A., Brown-Davies, C., Agyepong, I. A., Grobbee, D. E., Klipstein-Grobusch, K., & Ansah, E. K. (2015). Completeness and accuracy of data transfer of routine maternal health services data in the greater Accra region. *BioMed Central Research Notes*, 8(1). https://doi.org/10.1186/s13104-015-1058-3

Amouzou, A., Faye, C., Wyss, K., & Boerma, T. (2021). Strengthening routine health information systems for analysis and data use: a tipping point. In *BioMed Central Health Services Research* (Vol. 21). https://doi.org/10.1186/s12913-021-06648-1

Andargoli, A. E., Scheepers, H., Rajendran, D., & Sohal, A. (2017). Health information systems evaluation frameworks: A systematic review. *International Journal of Medical Informatics*, 97, 195–209. https://doi.org/10.1016/j.ijmedinf.2016.10.008

Aqil, A., Lippeveld, T., & Hozumi, D. (2009). PRISM framework: A paradigm shift for designing, strengthening and evaluating routine health information systems. *Health Policy and Planning*, 24(3), 217–228. https://doi.org/10.1093/heapol/czp010

Auditor General South Africa (AGSA). 2021. 2020/21 The Citizen's Report. https://www.agsa.co.za/Portals/0/Reports/PFMA/202021/Citizen's%20Report%20PFMA%202020-21%20FINAL%20-%20no%20spine.pdf?ver=2022-02-22-161944-810. Accessed on the 25th October 2022.

Barro, S. G., Dicko, A., de Lame, P. A., & Staccini, P. (2020). Health Metrics Network (HMN-WHO), a tool to assess the quality of the information system in Burkina hospitals. *European Federation for Medical Informatics (EFMI) and IOS Press*, 270, 602–606. DOI: 10.3233/SHTI200231

Begum, T., Khan, S. M., Adamou, B., Ferdous, J., Parvez, M. M., Islam, M. S., Kumkum, F. A., Rahman, A., & Anwar, I. (2020). Perceptions and experiences with district health information system software to collect and utilize health data in Bangladesh: A qualitative exploratory study. *BioMed Central Health Services Research*, 20(1), 1–13. https://doi.org/10.1186/s12913-020-05322-2

Bhattacharya, A, A., Umar, N., Audu, A., Felix, H., Allen, E., Schellenberg, J. R. M., & Marchant, T. (2019). Quality of routine facility data for monitoring priority maternal and newborn indicators in DHIS2: A case study from Gombe State, Nigeria. *PLoS ONE*, 14(1), 1–21. https://doi.org/10.1371/journal.pone.0211265

Bhattacharya, A. A., Allen, E., Umar, N., Audu, A., Felix, H., Schellenberg, J., & Marchant, T. (2020). Improving the quality of routine maternal and newborn data captured in primary health facilities in Gombe State, Northeastern Nigeria: A before-and-after study. *British Medical Journal Open*, 10(12), 1–10. https://doi.org/10.1136/British Medical Journalopen-2020-038174

Bless, C., Higson-Smith, C., & Sithole, S.L. 2013. Fundamentals of Social Research Methods: An African Perspective 4th ed. Cape Town: Juta and Company Ltd

Blödt, S., Kaiser, M., Adam, Y., Adami, S., Schultze, M., Müller-nordhorn, J., Holmberg, C., & Adam, Y. (2018). Understanding the role of health information in patients' experiences: secondary analysis of qualitative narrative interviews with people diagnosed with cancer in Germany. *British Medical Journal Open*, 1–9. https://doi.org/10.1136/bmjopen-2017-019576

Braa, J. and Sahay, S. (2017) 'Chapter 42: The DHIS2 Open Source Software Platform: Evolution over time and space', *Global Health Informatics: Principles of eHealth and mHealth to improve quality of care*, (April), pp. 1–12. https://www.researchgate.net/publication/316619278_The_DHIS2_Open_Source_Software_Platform_Evolution_Over_Time_and_Space Accessed on the 02 May 2022

Cheburet, S. K., & Odhiambo-Otieno, G. W. (2016). Process factors influencing data quality of routine health management information system: Case of Uasin Gishu County referral Hospital, Kenya. *International Research Journal of Public and Environmental Health*, *3*(6), 132–139. https://doi.org/10.15739/irjpeh.16.017

Chen, H., Hailey, D., Wang, N., & Yu, P. (2014). A review of data quality assessment methods for public health information systems. *International Journal of Environmental Research and Public Health*, 11(5), 5170–5207. https://doi.org/10.3390/ijerph110505170

Choudrie, J., Islam, M. S., Wahid, F., Bass, J. M., & Priyatma, J. E. (2017). Applying ICT to Health Information Systems (HIS) in Low Resource Settings: Implementing DHIS2 as an Integrated Health Information Platform in Lao PDR. *IFIP Advances in Information and Communication Technology*, 504, 536–547. https://doi.org/10.1007/978-3-319-59111-7

Dehnavieh, R., Haghdoost, A. A., Khosravi, A., Hoseinabadi, F., Rahimi, H., Poursheikhali, A., Khajehpour, N., Khajeh, Z., Mirshekari, N., Hasani, M., Radmerikhi, S., Haghighi, H., Mehrolhassani, M. H., Kazemi, E., & Aghamohamadi, S. (2019). The District Health Information System (DHIS2): A literature review and meta-synthesis of its strengths and operational challenges based on the experiences of 11 countries. *Health Information Management Journal*, 48(2), 62–75. https://doi.org/10.1177/1833358318777713

DHIS2. 2022. About DHIS2. University of Oslo; HISP Centre. Available from: https://dhis2.org/about/. Accessed on the 30 July 2022

Eastern Cape Department of Health (ECDoH). 2017. Annual Performance Plan – 2017/18. Bisho: Eastern Cape Province Government. http://www.echealth.gov.za/index.php/document-library/annual-performance-plans?task=download.send&id=308&catid=7&m=0. Accessed 25 July 2022

Eastern Cape Department of Health (ECDoH). 2018. Annual Performance Plan – 2018/19. Bisho: Eastern Cape Province Government. http://www.echealth.gov.za/index.php/document-library/annual-performance-plans?task=download.send&id=648&catid=7&m=0. Accessed 25 July 2022

Eastern Cape Department of Health (ECDoH). 2019. Annual Performance Plan – 2019/20. Bisho: Eastern Cape Province Government. http://www.echealth.gov.za/index.php/document-library/annual-performance-plans?task=download.send&id=1019&catid=7&m=0. Accessed 25 July 2021

Eastern Cape Department of Health (ECDoH). 2020. Annual Performance Plan – 2020/21. Bisho: Eastern Cape Province Government. http://www.echealth.gov.za/index.php/document-library/annual-performance-plans/send/7-annual-performance-plans/1246-ecdoh-annual-performance-plan-2020-21. Accessed 27 July 2021

Ehsani-Moghaddam, B., Martin, K., & Queenan, J. A. (2021). Data quality in healthcare: A report of practical experience with the Canadian Primary Care Sentinel Surveillance Network data. *Health Information Management Journal*, 50(1–2), 88–92. https://doi.org/10.1177/1833358319887743

Endriyas, M., Alano, A., Mekonnen, E., Ayele, S., Kelaye, T., Shiferaw, M., Misganaw, T., Samuel, T., Hailemariam, T., & Hailu, S. (2019). Understanding performance data: health management information system data accuracy in Southern Nations Nationalities and People's Region, Ethiopia. *BioMed Central Health Services Research*, 7, 1–6. https://doi.org/10.1186/s12913-019-3991-7

Farnham, A., Utzinger, J., Kulinkina, A. v., & Winkler, M. S. (2020). Using district health information to monitor sustainable development. *In Bulletin of the World Health Organization* (Vol. 98, Issue 1, pp. 69–71). World Health Organization.

https://doi.org/10.2471/BLT.19.239970

Faujdar DS, Sahay S, Singh T, Jindal H, K. R. (2019). Public health information systems for primary health care in India: A situational analysis study. *Journal of Family Medicine and Primary Care*, 8(11), 3640–3646. https://doi.org/10.4103/jfmpc.jfmpc_808_19

Feng, S., Hategeka, C., & Grépin, K. A. (2021). Addressing missing values in routine health information system data: an evaluation of imputation methods using data from the Democratic Republic of the Congo during the COVID - 19 pandemic. *Population Health Metrics*, 1–14. https://doi.org/10.1186/s12963-021-00274-z

Garrib, A., Stoops, N., Mckenzie, A., Dlamini, L., Govender, T., Rohde, J., & Herbst, K. (2008). An evaluation of the District Health Information System in rural South Africa. *South African Medical Journal*, 98(7), 549–552.

Getachew, N., Erkalo, B., & Garedew, M. G. (2022). Data quality and associated factors in the health management information system at health centers in Shashogo district, Hadiya zone, southern Ethiopia, 2021. *BioMed Central Medical Informatics and Decision Making*, *22*(1). https://doi.org/10.1186/s12911-022-01898-3

Githinji, S., Oyando, R., Malinga, J., Ejersa, W., Soti, D., Rono, J., Snow, R. W., Buff, A. M., & Noor, A. M. (2017). Completeness of malaria indicator data reporting via the District Health Information Software 2 in Kenya, 2011-2015. *Malaria Journal*, 16(1), 1–11. https://doi.org/10.1186/s12936-017-1973-y

Hagel, C., Paton, C., Mbevi, G., & English, M. (2020). Data for tracking SDGs: Challenges in capturing neonatal data from hospitals in Kenya. *British Medical Journal Global Health*, 5(3), 1–10. https://doi.org/10.1136/bmjgh-2019-002108

Haug, A., Zachariassen, F. and van Liempd, D. (2011) 'The costs of poor data quality', *Journal of Industrial Engineering and Management*, 4(2), pp. 168–193. DOI: 10.3926/jiem.2011.v4n2.p168-193.

Heale, R. and Twycross, A. (2015) 'Validity and reliability in quantitative studies', *Evidence-Based Nursing*, 18(3), pp. 66–67. DOI: 10.1136/eb-2015-102129.

Hilber, A. M., Blake, C., Bohle, L. F., Bandali, S., Agbon, E., & Hulton, L. (2016). Strengthening accountability for improved maternal and newborn health: A mapping of studies in Sub-Saharan Africa. *International Journal of Gynecology and Obstetrics*, 135(3), 345–357. https://doi.org/10.1016/j.ijgo.2016.09.008

HISP Team, 2016. WebDHIS Foundation: Facilitator Guide. Pretoria: Pretoria HISP House. https://www.hisp.org/?page_id=4506 Accessed 15 June 2022

Hotchkiss, D. R., Aqil, A., Lippeveld, T., & Mukooyo, E. (2010). Evaluation of the Performance of Routine Information System Management (PRISM) framework: Evidence from Uganda. *BMC Health Services Research*, 10. https://doi.org/10.1186/1472-6963-10-188

Hoxha, K., Hung, Y. W., Irwin, B. R., & Grépin, K. A. (2020). Understanding the challenges associated with the use of data from routine health information systems in low- and middle-income countries: A systematic review. *Health Information Management Journal*. https://doi.org/10.1177/1833358320928729

Hung, Y. W., Hoxha, K., Irwin, B. R., Law, M. R., & Grépin, K. A. (2020). Using routine health information data for research in low- And middle-income countries: A systematic review. *BioMed Central Health Services Research*, 20(1), 1–15. https://doi.org/10.1186/s12913-020-05660-1

Jinabhai, C. C., Onwubu, S. C., Sibiya, M. N., & Thakur, S. (2021). Accelerating implementation of District Health Information Systems: Perspectives from healthcare workers from KwaZulu-Natal, South Africa. *SA Journal of Information Management*, 23(1), 1–8. https://doi.org/10.4102/sajim.v23i1.1435

Kabakama, S., Ngallaba, S., Musto, R., Montesanti, S., Konje, E., & Kishamawe, C. (2016). International Journal of Medical Informatics Assessment of four common underfive children illnesses Routine Health Management Information System data for decision making at Ilemela Municipal Council, Northwest Tanzania: A case series analysis. *International Journal of*

Medical Informatics, 93, 85–91. https://doi.org/10.1016/j.ijmedinf.2016.06.003

Kajwang, B. (2022). Effects of training and development practices on performance of insurance sector in Kenya. *International Journal of Research in Business and Social Science (2147-4478)*, *11*(4), 140–148. https://doi.org/10.20525/ijrbs.v11i4.1825

Kanfe, S. G., Debele, G. R., Berhanu, R. D., Ngusie, H. S., & Ahmed, M. H. (2021). Utilisation of district health information system and its associated factors among health professionals working at public health facilities of the southwest of Ethiopia: Cross-sectional survey. *BMJ Open*, 11(8), 1–7. https://doi.org/10.1136/bmjopen-2020-046578

Karuri, J., Waiganjo, P., Orwa, D., & Manya, A. (2014). DHIS2: The Tool to Improve Health Data Demand and Use in Kenya. *Journal of Health Informatics in Developing Countries*, 8(1), 38–60. https://doi.org/https://www.jhidc.org/index.php/jhidc/article/view/113

Kebede, M., Adeba, E., Adeba, E., & Chego, M. (2020). Evaluation of quality and use of health management information system in primary health care units of east Wollega zone, Oromia regional state, Ethiopia. *BioMed Central Medical Informatics and Decision Making*, 20(1), 1–12. https://doi.org/10.1186/s12911-020-01148-4

Kiberu, V. M., Matovu, J. K., Makumbi, F., Kyozira, C., Mukooyo, E., & Wanyenze, R. K. (2014). Strengthening district-based health reporting through the district health management information software system: The Ugandan experience. *BMC Medical Informatics and Decision Making*, 14(1), 1–9. https://doi.org/10.1186/1472-6947-14-40

Kimberlin, C. and Winterstein, A. (2008) ‘Validity and reliability of measurement instruments used in research’, *American Journal of Health-System Pharmacy*, 65(23), pp. 2276–2284. DOI: 10.2146/ajhp070364.

Kiwanuka, A., Kimaro, H. C., & Senyoni, W. 2015. Analysis of the acceptance process of district health information systems (DHIS) for vertical health programmes: a case study of TB, HIV/aids and malaria programmes in Tanzania. *Electronic Journal of Information Systems in Developing Countries* 70(1): 1–14.

Kothari, C.R. 2004. Research Methodology: Methods and Techniques. 2nd ed. New Delhi: New Age International (P) Ltd

Kumar, M., Gotz, D., Nutley, T., & Smith, J. B. (2018). Research gaps in routine health information system design barriers to data quality and use in low- and middle-income countries: A literature review. *International Journal of Health Planning and Management*, 33(1), e1–e9. https://doi.org/10.1002/hpm.2447

Kuyo, R. O., Muiruri, L., & Njuguna, S. (2018). Organizational Factors Influencing the Adoption of the District Health Information System 2 in Uasin Gishu County, Kenya. *International Journal of Medical Research & Health Sciences*, 7(10), 48–57.

Lasim, O. U., Ansah, E. W., & Apaak, D. (2022). Maternal and child health data quality in health care facilities at the Cape Coast. *BioMed Central Health Services Research*, 1–15. https://doi.org/10.1186/s12913-022-08449-6

Lee, J., Lynch, C. A., Hashiguchi, L. O., Snow, R. W., Herz, N. D., Webster, J., Parkhurst, J., & Erondu, N. A. (2021). Interventions to improve district-level routine health data in low-income and middle-income countries: A systematic review. *British Medical Journal Global Health*, 6(6). https://doi.org/10.1136/bmjgh-2020-004223

Lemma, S., Janson, A., Persson, L. Å., Wickremasinhe, D., & Källestål, C. (2020). Improving quality and use of routine health information system data in low- And middle-income countries: A scoping review. *PLoS ONE*, 15(10 October), 1–17. https://doi.org/10.1371/journal.pone.0239683

Leon N, Balakrishna Y, Hohlfeld A, Odendaal WA, Schmidt B-M, Zweigenthal V, Anstey Watkins J, D. K. (2020). Routine Health Information System (RHIS) improvements for strengthened health system management. *Cochrane Database of Systematic Reviews*. https://doi.org/10.1002/14651858.CD012012.pub2.www.cochranelibrary.com

Liu, C., Zowghi, D. and Talaei-Khoei, A. (2020) 'An empirical study of the antecedents of data completeness in electronic medical records', *International Journal of Information Management*, 50 (October 2018), pp. 155–170. DOI: 10.1016/j.ijinfomgt.2019.05.001.

Lutge, E., Moodley, N., Tefera, A., Sartorius, B., Hardcastle, T., & Clarke, D. (2016). A hospital based surveillance system to assess the burden of trauma in KwaZulu-Natal Province South Africa. *Injury International Journal of the Care of the Injured*, 47(1), 135–140. https://doi.org/10.1016/j.injury.2015.08.020

Maina, I., Wanjala, P., Soti, D., Kipruto, H., & Boerma, T. (2017). Using health-facility data to assess subnational coverage of maternal and child health indicators, Kenya. *Bull World Health Organ*, August, 683–694. https://doi.org/http://dx.doi.org/10.2471/BLT.17.194399

Maïga, A., Jiwani, S. S., Mutua, M. K., Porth, T. A., Taylor, C. M., Asiki, G., Melesse, D. Y., Day, C., Strong, K. L., Faye, C. M., Viswanathan, K., O'Neill, K. P., Amouzou, A., Pond, B. S., & Boerma, T. (2019). Generating statistics from health facility data: The state of routine health information systems in Eastern and Southern Africa. *British Medical Journal Global Health*, 4(5), 1–9. https://doi.org/10.1136/British Medical Journalgh-2019-001849

Malembaka, E. B., Altare, C., Bigirinama, R. N., Bisimwa, G., & Banywesize, R. (2021). The use of health facility data to assess the effects of armed conflicts on maternal and child health: experience from the Kivu, DR Congo. *BioMed Central Health Services Research*, 7, 1–11. DOI: 10.1186/s12913-021-06143-7

Manoj, S., Wijekoon, A., Dharmawardhana, M., Wijesuriya, D., Rodrigo, S., Hewapathirana, R., Siribaddana, P., Gunasekera, T., & Dissanayake, V. H. W. (2012). Implementation of District Health Information Software 2 (DHIS2) in Sri Lanka. *Sri Lanka Journal of Bio-Medical Informatics*, 3(4), 109. https://doi.org/10.4038/sljbmi.v3i4.5431

Manya, A., & Nielsen, P. (2016). Reporting Practices and Data Quality in Health Information Systems in Developing Countries: An Exploratory Case Study in Kenya. *Journal of Health Informatics in Developing Countries*, 10(1), 114–126.

Mboera, L. E. G., Rumisha, S. F., Mbata, D., Mremi, I. R., Lyimo, E. P., & Joachim, C. (2021). Data utilisation and factors influencing the performance of the health management information system in Tanzania. *BioMed Central Health Services Research*, 21(1), 4–11. https://doi.org/10.1186/s12913-021-06559-1

Measure Evaluation. (2011) ‘Performance of Routine Information Systems Management (PRISM) Tools’, *MEASURE Evaluation*. Available at: https://www.measureevaluation.org/resources/publications/ms-11-46-d/at_download/document.

Measure Evaluation. (2016) ‘Routine Health Information Systems’, *MEASURE Evaluation*, (September), p. 2. Available from: https://www.measureevaluation.org/our-work/routine-health-information-systems.

Moukénet, A., Cola, M. A. De, Ward, C., Beakgoubé, H., Baker, K., Donovan, L., Laoukolé, J., & Richardson, S. (2021). Health management information system (HMIS) data quality and associated factors in Massaguet district, Chad. *BMC Medical Informatics and Decision Making*, 7, 1–10. https://doi.org/10.1186/s12911-021-01684-7

Mouton, E.B.J. 2012. The Practice of Social Research: South African Edition. Cape Town: Oxford University Press Southern Africa (Pty) Ltd

Mphatswe, W., Mate, K. S., Bennett, B., Ngidi, H., Reddy, J., Barker, P. M., & Rollins, N. (2012). Amélioration des informations sur la santé publique: Une intervention sur la qualité des données dans le Kwazulu-Natal (afrique du sud). *Bulletin of the World Health Organization*, 90(3), 176–182. https://doi.org/10.2471/BLT.11.092759

Muhoza, P., Tine, R., Faye, A., Gaye, I., Zeger, S. L., Diaw, A., Gueye, A. B., Kante, A. M., Ruff, A., & Marx, M. A. (2022). A data quality assessment of the first four years of malaria reporting in the Senegal DHIS2, 2014–2017. *BioMed Central Health Services Research,* 22(1), 1–13. https://doi.org/10.1186/s12913-021-07364-6

National Department of Health (NDoH). 2011. District Health Management Information System Policy. Pretoria: Government Printer.

National Department of Health (NDoH). 2012. District Health Management Information System (DHMIS) Standard Operating Procedures: Facility Level. Pretoria: Government Printer.

National Department of Health (NDoH). 2013. District Health Management Information System (DHMIS) Standard Operating Procedures: District Level. Pretoria: Government Printer.

National Department of Health (NDoH). 2017. Annual Report 2016/17. Pretoria: National Department of Health

National Department of Health (NDoH). 2017a. National Indicator Dataset 2017: Data Elements. Pretoria: National Department of Health. https://dd.dhmis.org/dataelements.html?file=NIDS%20Integrated&source=nids&ver=a92f Accessed 21 May 2022.

National Department of Health (NDoH). 2021. Annual Report 2020/21. Pretoria: National Department of Health. https://www.health.gov.za/wp-content/uploads/2021/11/Annual-Report-2020-2021.pdf. Accessed on the 21st June 2022.

Nguefack-Tsague, G., Tamfon, B. B., Ngnie-Teta, I., Ngoufack, M. N., Keugoung, B., Bataliack, S. M., & Bilounga Ndongo, C. (2020). Factors associated with the performance of routine health information system in Yaoundé-Cameroon: a cross-sectional survey. *BioMed Central Medical Informatics and Decision Making*, 20(1), 1–8. https://doi.org/10.1186/s12911-020-01357-x

Ngugi, B. K., Harrington, B., Porcher, E. N., & Wamai, R. G. (2019). Data quality shortcomings with the US HIV/AIDS surveillance system. *Health Informatics Journal*, 25(2), 304–314. https://doi.org/10.1177/1460458217706183

Nicol, E., Dudley, L. and Bradshaw, D. (2016) 'Assessing the quality of routine data for the prevention of mother-to-child transmission of HIV: An analytical observational study in two health districts with high HIV prevalence in South Africa', *International Journal of Medical Informatics*, 95, pp. 60–70. DOI: 10.1016/j.ijmedinf.2016.09.006.

Nsiah, R., Takramah, W., Anum-doku, S., Avagu, R., & Nyarko, D. (2022). Assessment of neonatal mortalities and stillbirths data quality in Offinso North District of Ghana. *International Journal of Scientific Research and Management*, 10(01), 508–519.

https://doi.org/10.18535/ijsrm/v10i01.mp01

O'Hagan, R., Marx, M. A., Finnegan, K. E., Naphini, P., Ng'ambi, K., Laija, K., Wilson, E., Park, L., Wachepa, S., Smith, J., Gombwa, L., Misomali, A., Mleme, T., & Yosefe, S. (2017). National assessment of data quality and associated systems-level factors in Malawi. *Global Health Science and Practice*, 5(3), 367–381. https://doi.org/10.9745/GHSP-D-17-00177

Ranganathan, P., & Aggarwal, R. (2018). Study designs: Part 1-An overview and classification. *Perspectives in Clinical Research*, 9(4), 184–186. https://doi.org/10.4103/picr.PICR_124_18

Republic of South Africa (RSA). 2004. Public Audit Act. No 25 of 2004. Government Printer: Pretoria.

Reynolds E, Martel LD, Bah MO, Bah M, Bah MB, Boubacar B, Camara N, Camara YB, Corvil S, Diallo BI, Diallo IT, Diallo MK, Diallo MT, Diallo T, Guilavogui S, Hemingway-Foday JJ, Hann F, Kaba A, Kaba AK, Kande M, Lamarana DM, Middleton K, Sidibe N, Souare O, W. M. and M. P. (2022). Implementation of DHIS2 for Disease Surveillance in Guinea: 2015 – 2020. *Frontiers in Public Health*, 9(January), 9:761196. https://doi.org/10.3389/fpubh.2021.761196

Roomaney, R. A., Wyk, V. P., Awotiwon, O. F., Nicol, E., Joubert, D., Bradshaw, D., & Hanmer, L. A. (2017). Review Availability and quality of routine morbidity data: review of studies in South Africa. *Journal of the American Medical Informatics Association*, 24(June 2016), 194–206. https://doi.org/10.1093/jamia/ocw075

Rumisha, S. F., Lyimo, E. P., Mremi, I. R., Tungu, P. K., Mwingira, V. S., Mbata, D., Malekia, S. E., Joachim, C., & Mboera, L. E. G. (2020). Data quality of the routine health management information system at the primary healthcare facility and district levels in Tanzania. *BioMed Central Medical Informatics and Decision Making,* 20(1), 1–22. https://doi.org/10.1186/s12911-020-01366-w

Saigí-Rubió, F., Pereyra-Rodríguez, J. J., Torrent-Sellens, J., Eguia, H., Azzopardi-Muscat, N., & Novillo-Ortiz, D. (2021). Routine health information systems in the European context: A systematic review of systematic reviews. *International Journal of Environmental Research and*

Public Health, 18(9). https://doi.org/10.3390/ijerph18094622

Shama, A. T., Roba, H. S., Abaerei, A. A., Gebremeskel, T. G., & Baraki, N. (2021). Assessment of quality of routine health information system data and associated factors among departments in public health facilities of Harari region, Ethiopia. *BioMed Central Medical Informatics and Decision Making*, 1–12. https://doi.org/10.1186/s12911-021-01651-2

Shiferaw, A. M., Zegeye, D. T., Assefa, S., & Yenit, M. K. (2017). Routine health information system utilization and factors associated thereof among health workers at government health institutions in East Gojjam Zone, Northwest Ethiopia. *BioMed Central Medical Informatics and Decision Making*, 17(1), 1–9. https://doi.org/10.1186/s12911-017-0509-2

Shuaib, F., Garba, A. B., Meribole, E., Obasi, S., Sule, A., Nnadi, C., Waziri, N. E., Bolu, O., Nguku, P. M., Ghiselli, M., Adegoke, O. J., Jacenko, S., Mungure, E., Gidado, S., Wilson, I., Wiesen, E., Elmousaad, H., Bloland, P., Rosencrans, L., Vertefeuille, J. (2020). Implementing the routine immunisation data module and dashboard of DHIS2 in Nigeria, 2014-2019. *British Medical Journal Global Health*, 5(7), 2014–2019. https://doi.org/10.1136/bmjgh-2019-002203

Singh, M., Goel, M. K., Behera, B. K., & Punia, J. S. (2016). Health Management Information System Data Quality under NRHM in District Sonipat, Haryana. -. *International Journal of Health Sciences and Research (IJHSR)*, 6(9), 11–14.

Solomon, M., Addise, M., Tassew, B., Balcha, B., & Abebe, A. (2021). Data quality assessment and associated factors in the health management information system among health centers of Southern Ethiopia. *PLoS ONE*, 16(10 October). https://doi.org/10.1371/journal.pone.0255949

Statistics South Africa (Stats SA). 2020. Mid-Year Population Estimates 2020. Stats SA. http://www.statssa.gov.za/publications/P0302/P03022020.pdf. Accessed 09 July 2021

Tamfon, B. B., Bilounga Ndongo, C., Bataliack, S. M., Ngoufack, M. N., & Nguefack-Tsague, G. (2020). Routine health information system in the health facilities in Yaoundé–Cameroon: assessing the gaps for strengthening. *BioMed Central Medical Informatics and Decision Making*, 20(1), 1–11. https://doi.org/10.1186/s12911-020-01351-3

Teklegiorgis, K., Tadesse, K., Mirutse, G., & Terefe, W. (2016). Level of data quality from Health Management Information Systems in a resources limited setting and its associated factors, eastern Ethiopia. *South African Journal of Information Management*, 18(1), 1–8. https://doi.org/10.4102/sajim.v18i1.612

Tilahun, B., Teklu, A., Mancuso, A., Abebaw, Z., Dessie, K., & Zegeye, D. (2018). How can the use of data within the immunisation programme be increased in order to improve data quality and ensure greater accountability in the health system? A protocol for implementation science study. *Health Research Policy and Systems*, 16(1), 1–6. https://doi.org/10.1186/s12961-018-0312-2

Tlale, L. B., Morake, B., Lesetedi, O., Maribe, L., Masweu, M., Faye, C., & Asiki, G. (2019). Data quality self-assessment of child health and sexual reproductive health indicators in Botswana, 2016-2017. In *PLoS ONE* (Vol. 14, Issue 8). Public Library of Science. https://doi.org/10.1371/journal.pone.0220313

Tripathy, J. P. (2013). Letter to the Editor Secondary Data Analysis: Ethical Issues and Challenges. In *Iranian J Publ Health* (Vol. 42, Issue 12). http://ijph.tums.ac.ir

Tshabalala, A. and Taylor, M. (2016) 'Disaggregated data to improve child health outcomes', *African journal of primary health care & family medicine*, 8(1), pp. 1–7. DOI: 10.4102/phcfm.v8i1.1221.

Vasbotten, M., Klevstul, S., Tronerud, K., Solberg, O. (2015). The use of DHIS2 within Malawi, pp. 1–24. http://www.uio.no/studier/emner/matnat/ifi/INF5210/h15/project-reports/dhis2.pdf

Wagenaar, B. H., Sherr, K., Fernandes, Q., & Wagenaar, A. C. (2016). Using routine health information systems for well-designed health evaluations in low- and middle-income countries. *Health Policy and Planning*, 31(1), 129–135. https://doi.org/10.1093/heapol/czv029

World Health Organisation (WHO). 2007. Everybody business: strengthening health systems to improve health outcomes: WHO's framework for action. Geneva: World Health Organization

World Health Organisation (WHO). (2008). Framework and standards for country health information systems, 2nd ed. Geneva: World Health Organization. Available from: https:// apps.who.int /iris/handle/ 10665/43872

World Health Organisation (WHO). (2017). Data quality review: a toolkit for facility data quality assessment: Module 1: Framework and metrics. Geneva: World Health Organization. Available from: https://apps.who.int/iris/handle/10665/259224

World Health Organisation (WHO). (2017a). Data quality review: a toolkit for facility data quality assessment. Module 2: Desk review of data quality. Geneva: World Health Organization. Available from: https://apps.who.int/iris/handle/10665/259225

Wude, H., Woldie, M., Melese, D., Lolaso, T., & Balcha, B. (2020). Utilization of routine health information and associated factors among health workers in Hadiya Zone, Southern Ethiopia. PLoS ONE, 15(5), 1–11. https://doi.org/10.1371/journal.pone.0233092

Youssef, D., Yaghi, A., Jouny, A., Abou-Abbas, L., Chammaa, H., & Ghosn, N. (2022). Converting the existing disease surveillance from a paper-based to an electronic-based system using district health information system (DHIS-2) for real-time information: the Lebanese experience. *BioMed Central Health Services Research*, *22*(1), 395. https://doi.org/10.1186/s12913-022-07773-1

Appendix A: Information sheet

UNIVERSITY OF THE WESTERN CAPE

Private Bag X 17, Bellville 7535, South Africa

Tel: +27 21 959 2809 Fax: 27 21 959 2872

E-mail: soph-comm@uwc.ac.za

INFORMATION SHEET

Project Title: Evaluation of the data quality of routine health information system in the public health facilities in the Eastern Cape Province, South Africa.

What is this study about?

This is a research project being conducted by **Sibusiso Sifundo Thabethe** at the Eastern Cape Department of Health. We are inviting you to participate in this research project because you are a person responsible for health information management at the Eastern Cape Department of Health. The purpose of this research project is to determine the quality of the data collected and reported in the public health facilities as part of the routine health information system.

What will I be asked to do if I agree to participate?

You will be requested to authorise the use of routine health information from WebDHIS. After this data extraction tool will be requested and specific templates will be shared which is divided into 6 datasheets. No further information nor follow-up interview will be required of you after submitting the requested information.

Would my participation in this study be kept confidential?

The researcher undertakes to protect the participant's identity and the nature of their contribution. To ensure your anonymity, this study will make use of pseudonyms and will not contain information that may identify participants. To maintain data confidentiality, the extracted data will be kept in a password-protected spreadsheet and be stored in a USB and hard drive, which will be only accessible to the researcher. If the researcher writes a report or article about this research project, the participant's identity will be protected.

What are the risks of this research?

There may be some risks from participating in this research study. All human interactions and talking about self or others carry some amount of risks. We will nevertheless minimize such risks and act promptly to assist you if you experience any discomfort, psychological or otherwise during the process of your participation in this study. Where necessary, an appropriate referral will be made to a suitable professional for further assistance or intervention.

What are the benefits of this research?

The benefits to you include the change in behaviours and attitudes towards data quality to improve and/or maintain the processes of managing routine health information and information systems in the Eastern Cape Department of Health. Other provincial departments of health and health ministries who may come across the findings may also benefit similarly. The findings of this study will be used to make recommendations to improve the data quality of the routine health information systems in the public health facilities and strengthen the health information management in the health systems.

Do I have to be in this research and may I stop participating at any time?

Your participation in this research is completely voluntary. You may choose not to take part at all. If you decide to participate in this research, you may stop participating at any time. If you decide not to participate in this study or if you stop participating at any time, you will not be penalized or lose any benefits to which you otherwise qualify.

What if I have questions?

This research is being conducted by **Sibusiso Sifundo Thabethe and the School of Public health** at the University of the Western Cape. If you have any questions about the research study itself, please contact **Sibusiso Sifundo Thabethe** at 4002960@myuwc.ca.za or WhatsApp on +27738726977.

Should you have any questions regarding this study and your rights as a research participant or if you wish to report any problems you have experienced related to the study, please contact:

Prof U Lehmann
Head of Department: School of Public Health
University of the Western Cape
Private Bag X17
Bellville 7535
ulehmann@uwc.ac.za

Prof Anthea Rhoda
Dean: Faculty of Community and Health Sciences
University of the Western Cape
Private Bag X17
Bellville 7535
chs-deansoffice@uwc.ac.za

Appendix B: UWC Ethical Approval letter

UNIVERSITY *of the*
WESTERN CAPE

YEARS
of hope, action & knowledge

06 May 2022

Mr S Thabethe
School of Public Health
Faculty of Community and Health Sciences

HSSREC Reference Number:	HS22/3/2
Project Title:	Evaluation of the data quality of routine health information system in the public health facilities in the Eastern Cape Province, South Africa.
Approval Period:	**21 April 2022 – 21 April 2025**

I hereby certify that the Humanities and Social Science Research Ethics Committee of the University of the Western Cape approved the methodology, and amendments to the ethics of the above mentioned research project.

Any amendments, extension or other modifications to the protocol must be submitted to the Ethics Committee for approval.

Please remember to submit a progress report by 30 November each year for the duration of the project.

For permission to conduct research using student and/or staff data or to distribute research surveys/questionnaires please apply via:
https://sites.google.com/uwc.ac.za/permissionresearch/home

The permission letter must then be submitted to HSSREC for record keeping purposes.

The Committee must be informed of any serious adverse events and/or termination of the study.

Ms Patricia Josias
Research Ethics Committee Officer
University of the Western Cape

Director: Research Development
University of the Western Cape
Private Bag X 17
Bellville 7535
Republic of South Africa
Tel: +27 21 959 4111
Email: research-ethics@uwc.ac.za

NHREC Registration Number: HSSREC-130416-049

Appendix C: ECDoH Study Approval Letter

Province of the
EASTERN CAPE
HEALTH

Enquiries: Yvonne Gixela Tel no: 079 074 0859

Email: Yvonne.Gixela@echealth.gov.za / ygixela@gmail.com

Date: 10 May 2022

Evaluation of the data quality of routine health information system in the public health facilities in the Eastern Cape Province, South Africa. (EC_202205_003)

Dear Mr S. Thabethe

The department would like to inform you that your application for the above mentioned research topic has been approved based on the following conditions:

1. During your study, you will follow the submitted protocol with ethical approval and can only deviate from it after having a written approval from the Department of Health in writing.

2. You are advised to ensure, observe and respect the rights and culture of your research participants and maintain confidentiality of their identities and shall remove or not collect any information which can be used to link the participants.

3. The Department of Health expects you to provide a progress update on your study every 3 months (from date you received this letter) in writing.

4. At the end of your study, you will be expected to send a full written report with your findings and implementable recommendations to the Eastern Cape Health Research Committee secretariat. You may also be invited to the department to come and present your research findings with your implementable recommendations.

5. Your results on the Eastern Cape will not be presented anywhere unless you have shared them with the Department of Health as indicated above.

Your compliance in this regard will be highly appreciated.

SECRETARIAT: EASTERN CAPE HEALTH RESEARCH COMMITTE

2030 NDP

TOGETHER, MOVING THE HEALTH SYSTEM FORWARD

Appendix D: ECDoH HIMS Director Approval

Lesego Nemavhandu <lesegonem@gmail.com> | Sibusiso Thabethe; Roseline Nkobo

Re: Request to use DHIS data for Academic

Cc Roseline Nkobo

You replied to this message on 2022/09/10 08:52.

Action Items

Dear Mr Thabethe

Permission is hereby granted to use the aggregated DHIS data for the purposes of the research study only. Kindly abide by the provisions as per the Research Committee approval letter, and also make copy (s) of the findings to first your selected facilities and HIMS head office on completion of the study. I assume that you are not focusing on all facilities and all over 500 data elements that we collect?(Hence copies to selected facilities)

Wishing you all the best.

Best Regards
Ms Lesego Nemavhandu
Director: Health Information Management and Systems
ECDOH
lesego.nemavhandu@ecdoh.gov.za
040 608 1925

On Thu, May 19, 2022 at 10:05 AM Sibusiso Thabethe <Sibusiso.Thabethe@echealth.gov.za> wrote:

> Good morning
>
> Please note that I have registered with the University of Western Cape (UWC), and this year I'm required to do a mini-thesis.
>
> I have chosen EC Health as a Unit of Analysis, and I'm intending to use DHIS data.
>
> I would like to request your permission to use DHIS data for my study.
>
> 1. The study is a Desktop and Retrospective in nature
> 2. The study will focus on the NIDS 2017 data quality for the province
>
> I have attached the supporting documents, which are as follows;
>
> 1. EC Health Study Approval Letter
> 2. UWC ethics clearance
>
> I will appreciate your response
>
> Regards,

Appendix E: Monthly Routine Data

District Name:														
Financial Year														
Health Facility	Data Element Group	Data elements	Month 1	Month 2	Month 3	Month 4	Month 5	Month 6	Month 7	Month 8	Month 9	Month 10	Month 11	Month 12

Appendix F: Data Validation Report

District Name:								
Financial Year								
Health Facility	Period	Validation rule	Importance	Left side description	Value	Operator	Value	Right side description

Appendix G: Outliers and Missing Data Report

Province Name:																		
Financial Year																		
Health Facility	Data elements	Month 1	Month 2	Month 3	Month 4	Month 5	Month 6	Month 7	Month 8	Month 9	Month 10	Month 11	Month 12	Max Z score	Max modified Z score	Gap weight	Outlier weight	Total weight

Appendix H: Data Marked for Follow-Up Report

District Name:								
Financial Year								
Health Facility	Period	Validation rule	Importance	Left side description	Value	Operator	Value	Right side description