An Assessment of Data Quality in Routine Health Information Systems in Oyo State, Nigeria.

ADEDAPO ADEJUMO

A mini-thesis submitted in partial fulfillment of the requirements for the degree of Master in Public Health at the School of Public Health,

University of the Western Cape

Supervisor: Ms. Verona Mathews

January 2017

KEYWORDS

Health Information

Health Management Information System

Data Quality

Routine Data

Completeness

Reliability

Use of Information

District Health Information System

Routine Data Quality Assessment

Nigeria

ABBREVIATIONS

DHIS - District Health Information System

DPT3 - Diphtheria-Tetanus-Pertussis Third Dose

DQA - Data Quality Assessment

FMOH - Federal Ministry of Health

HIS - Health Information System

HMIS - Health Management Information System

HMN - Health Metrics Network

LGA - Local Government Area

MOH - Ministry of Health

PHIN - Pacific Health Information Network

PMTCT - Prevention of Maternal to Child Transmission of HIV

RDQA - Routine Data Quality Assessment

WHO - World Health Organization

ACKNOWLEDGEMENTS

This work will not have been possible without the will of God and my gratitude goes to the almighty. My deepest admiration and thanks goes to my supervisor, Ms Verona Matthews for the first class supervision and guidance on this project as well as the high quality inputs into the completion of this thesis.

My sincere appreciation also goes to the school of public health administrators, Janine Kader and Corinne Carollisen for the provision of support during this challenging period of study at the University of Western Cape. Appreciation also goes to colleagues and teachers who have contributed in one way or the other to the completion of this degree.

Last but not the least is my lovely family who has been patient and very supportive during my studies. May God bless you immensely.

DECLARATION

I declare that “An Assessment of Data Quality in Routine Health Information Systems in Oyo State, Nigeria” is my own work, that it has not been submitted before for any degree or examination in any other University or College, and that all sources I have used or quoted have been indicated and acknowledged as complete references.

Adedapo Adejumo November 2016

Signed:...................................

ABSTRACT

Ensuring that routine health information systems provide good quality information for informed decision making and planning in health systems remain a major priority in several countries and health systems. The lack of use of health information or use of poor quality data in health care and systems results in inadequate assessments and evaluation of health care and result in weak and poorly functioning health systems. The Nigerian health system like in many developing countries has challenges with the building blocks of the health system with a weak Health Information System. Although the quality of data in the Nigerian routine health information system has been deemed poor in some reports and studies, there is little research based evidence of the current state of data quality in the country as well as factors that may influence data quality in routine health information systems.

This study explored the data quality of routine health information generated from health facilities in Oyo State, Nigeria, providing the state of data quality of the routine health information. This study was a cross sectional descriptive study taking a retrospective look at paper based and electronic data records in the National Health Management Information System in Nigeria. A mixed methodology approaches with quantitative to assess the quality of data within the health information system and qualitative methods to identify factors influencing the quality of health information at the health facilities in the district.

Assessment of the quality of information was done using a structured evaluation tool looking at completeness, accuracy and consistency of routine health statistics generated at these health facilities. A multistage sampling method was used in the quantitative component of the research. For the qualitative component of the research, purposive sampling was done to select respondents from each health facility

to describe the factors influencing data quality. The study found incomplete and inaccurate data in facility paper summaries as well as in the electronic databases storing aggregate information from the facility data.

Table of Contents

CHAPTER ONE - INTRODUCTION

Health Information Systems is one of the six building blocks of a Health System (World Health Organization, 2008), while the other 5 building blocks (health workforce; health services; health financing; governance and leadership; medical products, vaccines,) are vital to any health system the Health Information System building block provides vital information for effective decision making for the other building blocks (AbouZahr & Boerma, 2005). An information system refers to the "*structures and processes dedicated to the collection, storage, retrieval and use of information usually within the context of an organization*" (Callaos & Callaos, 2002, p. 2), thus a health information systems consists of the "*data collection, processing, archiving and use of the information required for the specific aim of improving health service efficiency and health systems*" (Lippeveld, 2001, p. 3).

In a health information system there are different types of information based on the frequency of information generation: routine and non-routine information. Routine health information provide information at regular intervals to meet predictable information needs (Hotchkiss, Aqil, Lippeveld, & Mukooyo, 2010) whereas non-routine health information like population censuses, demographic health surveys provide information on an ad hoc basis and over longer intervals usually to complement what is collected via routine health information (Lippeveld, Sauerborn, & Bodart, 2000). This research study focusses on the information collected routinely in a health information system which includes service delivery statistics from health service delivery units, health facilities and communities.

Even though decision making are largely based on surveys and ad hoc reviews, these methods are more expensive and provide information intermittently (Fernandes, Wagenaar, & Anselmi, 2014). In a health system, the measurement of success is determined by the health system's performance which depends on the generation and use of quality routine health data and information extracted from the health information systems (WHO, 2007). Unfortunately, though routine health information is important the poor quality of the routine health information impedes the effective use of information for decision making in health systems.

Despite the pivotal importance of good quality health data, it has been found that "*in practice, HMIS data have a number of limitations and quality problems, such as missing values, bias, and computation errors*"(World Health Organization, 2008, p. 11) .The health information system is bedevilled with "*serious limitations in the value of the health information that 'data-led' national information system has provided, particularly regarding its availability and usefulness for decision-making processes at local level*" (Health Reform Foundation of Nigeria, 2007, p. 304).

The National Health Management Information System (NHMIS) in Nigeria has over the years been noted to be weak specifically in terms of data completeness, reliability and use in supporting the health system (Anifalaje, 2009; Health Reform Foundation of Nigeria, 2007). In 2010, an assessment on the data quality of the routine health management information in one of the Nigerian states found poor data quality at health facility and district levels to consist of missing values, inconsistent data and poor usability (Makinde, 2012). Furthermore, poor data quality is experienced despite of routine health management information systems being part of international donor investments in health systems strengthening (Warren, Wyss, Shakarishvili, Atun, & de Savigny, 2013). Huge financial and non-financial investments have been made in the data collection and maintenance of the health information systems but the information is not used for decision making (Shaw, 2005).

The effects of poor data quality impact several aspects of the health system including planning, resource financing and management (Mavimbe, Braa, & Bjune, 2005). Thus, to ensure high quality data in routine health information systems the root causes of poor data quality and the factors that affect data quality has to be identified.

In Nigeria, the National Health Management Information System provides a framework on the process for data collection and collation, analysis and use of health data in the country (Appendix 1) (Federal Ministry of Health, 2006). In the framework, primary health care is at the core of the Nigerian health system and this is where health service statistics are reported from the health facilities on a monthly basis to the supervising districts. It is mandatory for each health facility to report a minimum set of health information data on a monthly basis to the respective supervising health districts. The minimum set of health information data is collected on a set of data collection tools at the health facilities. The data collection tools starts with daily registers collecting data on each individual patient daily which is then aggregated to monthly summary forms and eventually the data is captured on an electronic database, DHIS (District Health Information System Software). Thus, there are three (3) sources of health information integrated to form the National Health Information System: daily data collection tools, monthly health facility summary form, and finally the data capturing form on the electronic database DHIS.

Data quality is usually assessed through conducting data quality assessment exercises to health facilities by district and subnational officials. This assessment is typically executed by the District Monitoring and Evaluation Officer on a quarterly basis. For this data quality assessment exercise, a nationally approved tool is used to evaluate data quality at the health facility. However, these exercises are often donor driven and on an ad hoc basis depending on the availability of resources to conduct these exercises. In summary, good quality routine health information system is essential to the success for health information system and the overall health system. However, information within routine health information systems is often of poor quality and hindering the use of information.

Problem Statement

In Nigeria the use of routine information for decision making at all administrative levels is limited and this has been ascribed to the poor data quality of routine health information (Federal Ministry of Health, 2013). However, there is a paucity of research based evidence on the current state of data quality in Nigeria as well as the factors that may influence data quality in routine health information systems (Hahn, Wanjala, & Marx, 2013).

It is thus necessary to assess the quality of routine health information as well as to investigate the factors that affect data quality in Nigeria to generate good quality data in routine health information systems for the use of information for decision making and planning.

CHAPTER TWO - LITERATURE REVIEW

The Literature Review chapter focusses on the concept of data quality in routine health information systems. The thematic areas covered in this chapter are the definition of data quality, dimensions of data quality, the measurement of data quality, and finally the factors that influence data quality in routine health information systems.

Definitions of data quality

The Health Metrics Network (HMN) describes Health Information Systems as six building blocks: HIS resources, indicators, data sources, data management, information products and dissemination, and finally the use of information (WHO, 2007). In addition, the HMN states that the health information system collects data from the health sector and other relevant sectors, analyses the data and ensures their overall quality, relevance and timeliness, and converts data into information for health-related decision-making (WHO, 2016).

The Pacific Health Information Network (PHIN) adds to this definition by categorizing health information systems into 2 systems based on the frequency of data collection systems - routine and non-routine health information systems (Lippeveld et al., 2000) . The authors describe non routine health information systems as information systems that collect or gather data in a non-periodic fashion or an ad hoc basis including surveys, research studies and case studies (Lippeveld et al., 2000). Whereas, routine health information systems refers to *"a systematic and periodic method of collecting data in a defined time periodicity"* and is often part of a continuous system within countries or organizations (Lippeveld et al., 2000, p. 54). Both routine and non-routine health information systems together provide information to health systems. This research study focuses on routine health information systems.

Routine health information systems are considered important in this research study because it provides information to different levels of the health system, it is used for the planning of health system interventions and for the effective monitoring and evaluation of health systems (AbouZahr & Boerma, 2005). Planning of the health interventions cut across several health decisions including resource allocation, monitoring and evaluation of health program goals, micro and macro planning of health activities and public health systems research (AbouZahr & Boerma, 2005; Bowen, Erickson, Martens, & Crockett, 2009).

Data quality is a vital component of health information systems and the importance of the availability of usable routine health information is central to the use of the information for planning and decision making (World Health Organization, 2008). However, despite the recognized importance of the production and use of good quality data for effective health systems monitoring and evaluation, health information systems is ascribed a poor level of data quality (Gething et al., 2006; Ndabarora, Chipps, & Uys, 2013). Good quality data is said to be when the information available fits or meets the intended goals of its users (Chen, Hailey, Wang, & Yu, 2014). Unfortunately, health information systems in developing countries often fall into a vicious cycle of poor data quality and poor information use (Heywood & Rhode, 2001) .

Dimensions of data quality

In the past, data quality referred to the *"fitness for use"* (Tayi & Ballou, 1998; Wand & Wang, 1996; Wang & Strong, 1996). More recently, researchers propose that there are properties of data that determines data quality and there appears to be an agreement that data quality is a multidimensional concept (Chen et al., 2014; Wand & Wang, 1996; Wang & Strong, 1996; Zozus et al., 2014). However, though there is no agreement on the dimensions of data quality there are cross cutting dimensions identified by the literature: completeness, timeliness, consistency, accuracy , reliability and precision (Glèlè Ahanhanzo et al., 2014; Ndabarora et al., 2013; Pipino, Lee, & Wang, 2002; Wand & Wang, 1996). The figure below illustrates the interconnectedness of these different dimensions in data quality.

Figure 1 Data Quality Dimensions

Specifically, the dimensions completeness, accuracy, consistency, and timeliness were found the most commonest reviewed in the literature (Chen et al., 2014). Completeness is defined as "*a measure of the presence of expected data items in a given dataset or collection*" (Wand & Wang, 1996 pp 23). Accuracy is described as "*the closeness of data values to the truth or the veracity of the information received*" (Chen et al., 2014, pp2). Data consistency and accuracy are considered separate data quality dimensions but consistency can only be achieved if data is accurate and valid because the stability of data ensures consistency (Hahn et al., 2013).Timeliness is viewed as the "*extent to which a particular set of data is current in relation to a specified time*" (Vaziri, 2012, p. 6). Chahed et al (2013) used the completeness and accuracy dimensions to evaluate the immunization data with focus of reported values for DPT3 using daily and monthly PHC reporting forms. In this research study the dimensions that will be investigated are completeness and consistency.

Measuring data quality

The measurement of the quality of the data in routine health information systems have used various methodologies typically using one or more of dimensions of data quality mentioned above (Chen et al., 2014). However, the measurements of data quality appears to have focused on identifying poor quality data such as data inconsistencies, data accuracy errors and misrepresentations (Chen et al., 2014).

One particular measurement tool, the Routine Data Quality Assessment tool (RDQA) developed by Measure Evaluation (Measure Evaluation, 2008) has been used in many countries to assess data quality in routine health information systems (Abah, 2012; Edgard-marius, Alphonse, Leve, & Makoutode, 2014). The tool uses a two-pronged approach looking at data verification and system assessment to evaluate data quality and can be used either in its original form or adapted to meet specific needs.

The first of the two-pronged approach is the data verification part of the tool. The data verification recounts reported data values against source documents and uses the values to calculate a ratio comparing the values obtained. The second approach in the RDQA is the systems assessment focusing on a qualitative approach to assess the data management and reporting systems at data administrative levels. The assessment covers the training, indicator definitions, data requirements, data management and quality control measures in the data management process (Abah, 2012). The RDQA basically combines both quantitative and qualitative methods to assess data quality.

The RDQA tool has been used in various countries and for different purposes illustrating its usefulness in assessing data quality. For example, the tool was used in Nigeria, to evaluate the quality of HIV data to improve Grant Applications using both approaches; Data Verification and Systems Assessment (Abah, 2012). This study found poor quality data in ART clinics in Nigeria and attributed this to late submission of data from health facilities as well as a high turnover rate of health facility staff.

In another example, the tool was used in a cross country data quality assessment of Immunization data involving 27 countries (Ronveaux et al., 2005). Using the verification factor technique, the study found that only about 42% of districts evaluated fell between the desired range of .85 and 1.15 (Ronveaux et al., 2005). The verification factor was calculated by recounting DPT3 values at health facilities and comparing with values reported at the districts and national levels (Ronveaux et al., 2005).

Similarly, researchers in Tunisia examine the consistency of immunization data using verification factors (Chahed, Bellali, Alaya, Ali, & Mahmoudi, 2013). The consistency of the reporting system was determined by comparing reported DPT3 values with written documentation in health facilities and districts. Chahed et al found large discrepancies between the DPT3 values recorded in the facility registers, facility summary forms as well as district summaries. They found good data completion rates in their study although this was based on the availability of paper records. In another case, a process evaluation approach was adopted to assess immunization data (Mavimbe et al., 2005). In this process evaluation approach facility reports were crosschecked with tally sheets as well as district reports. Interviews were also conducted with health workers in the health facilities about data collection methods, quality of feedback on data quality as well as interactions with their respective district management teams. The study found poor consistency between values at health facilities and districts in Mozambique although data was complete (Mavimbe et al., 2005).

Using qualitative methods researchers in Mozambique tried to investigate the data quality of routine malaria data finding problems with the available malaria data in terms of completeness and accuracy (Chilundo, Sundby, & Aanestad, 2004). The results showed a discrepancy of 62% in the number of malaria cases reported at the district paper based records and electronic provincial records. The researchers attributed this to human errors in computation. In summary, the illustrations of the use of the RDQA tool described above provide evidence of the usefulness of the tool in investigating data quality in routine health information systems of vertical programs. The challenge is applying the tool to all the health programs in the health system.

Furthermore, several research studies in developing countries were identified conducting measurements of data quality in routine health information systems at facility and district levels (Chilundo et al., 2004; Ledikwe et al., 2014; Mphatswe et al., 2012; Ronveaux et al., 2005). Chilundo et al conducted a study on malaria data and found marked differences between laboratory registers of malaria tests and what was recorded in summary forms submitted to district and provincial levels (Chilundo et al., 2004). In this case, large discrepancies in values were found comparing the data at provincial and National levels in Mozambique.

Ledikwe et al (2014)conducted a qualitative study in Botswana using interviews based on the routine data quality audit tool. The interviewees were asked questions related to data management processes, monitoring and evaluation structures, indicator definitions and National information systems as related to the quality of health information in Botswana. The study found that there were generally good monitoring and evaluation systems in place with available human resources at the district and national levels to ensure the production of good quality health information in Botswana. There was however challenges reported with the electronic data systems in the country at the time of the study. These challenges included the presence of multiple systems lacking the ability to integrate the different health systems making the systems unreliable for most users in the system.

Looking at PMTCT data in KwaZulu Natal Province in South Africa, Mphatswe et al(2009) evaluated the quality of PMTCT data before and after a specific data quality intervention. The specific intervention involved the training of facility and district staff on the core principles of routine health information systems with focus on ensuring data quality. The study found improvement in both data completeness and accuracy after the specific data quality intervention in the health facilities.

In another study, the Global Fund on site Data Verification tool was used to rapidly assess the data quality in routine health information systems from health facilities to national levels in Mozambique (Gimbel et al., 2011). The study looked at three aspects; verification of the availability of monthly facility reports at the health facility and district health departments, evaluation of the reliability (concordance) of monthly statistics obtained from facility clinical registries, monthly facility reports, and the MOH electronic database as well as the examination of the validity of the HIS data by comparison with population-level surveys over time(Gimbel et al., 2011).

An example of a study focusing on facility level data is a qualitative study conducted in Tanzania (Wilms, Mbembela, Prytherch, Hellmold, & Kuelker, 2014). Wilms et al (2014) explored the implementation of National data collection tools and data quality at one district hospital in Tanzania. The researchers used various qualitative methods include interviews, direct observations as well as the retrospective inspection of reporting documents used in the health facility. Findings included pervasive inadequacies in the completeness and accuracy of health records in the secondary documents used for reporting outside the health facility. The study also found that the information available from the hospital was not used for any decision making process.

Factors determining data quality

The factors affecting data quality identified in routine health information systems are behavioral, infrastructural and systems based (Glèlè Ahanhanzo et al., 2014). The behavioral factors include health staff motivation, presence of incentives or disincentives; the infrastructural factors include availability of proper data collection tools and equipment, quantity and quality of human resources for health information systems and use of technology; and the systems factors include level of data demand and use, feedback mechanisms within health administrative levels, routine data quality checks and availability of robust routine health information system policies (Glèlè Ahanhanzo et al., 2014).

Furthermore, the human resources in routine health information systems remain a key factor in determining data quality because a strong health information system is built on well-functioning core building blocks driven by vibrant human resources for health (Health Metrics Network, 2005).

In Benin, researchers using the RDQA tools and Lot Sampling technique identified human resource levels, management and planning capacity as well as the state of infrastructure as some of the factors that influenced the quality of routine health information in Benin. Health Facilities with well trained staff and management capacity were found to have better quality health information (Glèlè Ahanhanzo et al., 2014) confirming their finding that human resources play a major role in determining the quality of data within a routine health information system and identified specifically health workers competence within the scope of their training as a factor.

The Benin study was a cross- sectional descriptive study that aimed to determine the factors affecting the quality of data in the routine health information system of Benin. Focus group discussions were used to collect information from health staff related to the collection of data in routine health information systems. The focus groups gave the following reasons as responsible for the poor data quality seen in the study ; large amount of data required in tools , format of data collection forms , demotivation of staff in routine heath information systems activities and poor capacity (Glèlè Ahanhanzo et al., 2014).

Using mixed methods in a cross sectional study Cheburet and Odiam looked at organizational factors that influence the quality of routine health information systems in one hospital in Kenya (Cheburet & Odhiambo-Otieno, 2016). This study obtained data via questionnaires administered to respondents in the health facility asking questions related to the process of data collection as well as data quality protocols available in the health facility. Data verification done in the study showed gaps in data completeness and consistency in the available data in the health facility. The study reported the presence of strong organization protocols for data quality as a major factor in determining the quality of data in routine health information systems.

Motivation or the lack of motivation has been described in literature as determinants of health staff performance and remains a major determinant of human resource performance. Since human resources play a major role in the functioning or routine health information systems, the general issues that affect staff performance and efficiency within health systems will have an influence on information systems. A literature review of performance based financing in health systems across several countries showed that financial incentives improved the quality and efficiency of health staff in health facilities (Witter et al., 2013).

Although financial incentives were not identified as a factor in most of the reviewed studies, it has been reported as a possible determinant of human resource performance in health systems and by extension health information systems (Ireland, Paul, & Dujardin, 2011; James et al., 2012; Witter et al., 2013). Reward for good quality data whether by direct inducement or incentives can contribute to the quality of data produced at health facilities, this is an extension of effects of performance based inducement programs for health workers making this a possible factor that influences data quality in routine health information systems (Miller, Musominali, Baganizi, & Paccione, 2014).

Regular data audits often referred to as data quality assessments were identified as contributors to the level of data quality in Health Information systems as well as investments in human resource development(Mutale et al., 2013). A study looking at the development of health information systems in five countries, Mutale et al (2013) identified engagement at the district and facility level to institutionalize routine data quality audits as a factor to improve the quality and subsequent use of data within routine health information systems. They posited that these regular audits accompanied by regular feedback on data quality will improve the quality of data in the health information system.

The capacity of the human resource available has also been identified in several studies as determinants of the quality of data in routine health information systems. The capacity of health staff to understand with specific training for health care workers on the importance of public health information, monthly data reviews and feedback, regular data audits (Mphatswe et al., 2012).

The review of literature has shown the importance of data quality in routine health information systems among HIS researchers. The review also showed the multiple methods used in assessing this property of data based on the multiple dimensions of data quality. A variety of results have also emerged from previous studies on the quality of data in health information systems. This variety is largely borne out of the multiplicity of the ways in which the quality of data can be evaluated as well as the different factors that can influence the quality of routine health information based on systems and environments The importance of data quality shown in the review underscores the purpose of this study to add to the body of knowledge on routine health information systems.

CHAPTER THREE - METHODOLOGY

The chapter begins with identifying the aim and objectives of this research study and the research approach used. It continues by describing the research strategy and research method and concludes with a description of the ethical considerations in this research study.

Aims and Objectives

The aim of the study is to assess the data quality in routine information systems in Oyo State Nigeria and to identify the factors that influence data quality in order to improve the use of routine health information for decision making.

The objectives of the study were as follows:

1. To assess the completeness and consistency of data collection tools in the routine health information system
2. To assess the accuracy of routine health information
3. To identify and describe the factors that influence data quality in routine health information systems

Study Design

A mixed methodology approach using quantitative methods to assess the quality of data within the health information system and qualitative methods to identify factors influencing the quality of health information at the health facilities in the district was employed. The quantitative section of the research study is a retrospective research study using paper based and electronic data records in the National Health Management Information System in Nigeria between the months of April and June 2015. The qualitative section consists of interviews with key informants from the health facilities.

Research Area

Nigeria is divided into 37 Administrative Areas called States. These States are further divided into Administrative Areas called Local Government Areas. The Local Government Areas is the equivalent of a health district. The Local Government Areas are divided into is Wards where the health facilities are located.

Oyo State is made up of 33 Local Government Areas and 347 Wards. Each Local Government Area has a health department that oversees health matters in the health district. Oyo State covers an approximate area of 28,454 square kilometers in the south west of Nigeria and has an estimated 5.5 million inhabitants spread across the 33 in the State (Federal Ministry of Health, 2012). There are 1234 health facilities in the state which are both publicly and privately owned (Federal Ministry of Health, 2012).

Population and Sampling

Study Population

There are 1234 health facilities in Oyo State distributed among the 33 Local Government Areas. The health facilities in Oyo State are either publicly or privately owned and all facilities are mandated to submit monthly records of health activities to the district using the National Minimum Dataset (Appendix 1).

Sampling

Multistage sampling method was used in this research study. The multistage stage sampling method utilizes a methodology taking selection of samples from increasing smaller samples of the population (Barreiro & Albandoz, 2001; Whittemore, 1997).The sampling involved 4 stages: Stage 1 LGA Selection, Stage 2 Health Facility Selection, Stage 3 Data Element Selection, and Stage 4 Selection of key informants. The Microsoft Excel Random Function was used for to conduct the selection for each stage. The Microsoft Excel Random Function creates an evenly distributed random real number greater than or equal to 0 and less than 1 and these series of randomly generated numbers can be assigned to cells in excel.

Stage 1 Local government area selection

The two LGAs were randomly selected from the 33 LGAs in the State using the Microsoft Excel Random Function. The list of LGA's was sorted alphabetically in Microsoft Excel and the Microsoft Excel Random Function generated 33 random numbers in a column adjacent to the list of LGA's. The random numbers created a new list and was sorted from smallest to largest where the first two LGA's were then selected from this list.

Stage 2 Health Facility Selections

The 2nd stage involved the random selection of 12 facilities from each of the two selected LGAs making a total of 24 health facilities in the study sample. The list of facilities in each LGA was sorted alphabetically in Microsoft Excel. The Microsoft Excel Random Function was then used to generate random numbers in a column adjacent to the list of Facilities. The same process as previously stated where followed and the 1st 12 health facilities in each LGA were then selected.

Stage 3 Data Element Selection

The data element selection is based on a list of data elements that contains 234 data elements covering the following PHC activities in health facilities: antenatal care, immunization, maternal and child health, facility attendance, HIV/AIDS, TB, Malaria, Family Planning and disease reporting. A sample of the data elements were selected due to the large number of the data elements. The sample size for the National Minimum Dataset was calculated using the Rao soft calculator with 95% Confidence Interval and 5% Margin of error which provided a sample size of 147 data elements (Appendix 3). The current National Minimum Dataset (Appendix 2) has 36 sub sections categorizing the different services offered in the health facility. One data element was randomly selected section by section using the Microsoft Excel Randomization function. This process was repeated until a sample of 147 data elements was reached.

Stage 4 Key Informant Selections

For this stage, random selection was not used but rather purposive sampling. Purposive sampling was used to select approximately 4 – 10 health workers in each health facility. Staff members were purposively selected from the health facility based on their health information systems roles and responsibilities in the health facility (Teddlie & Yu, 2007). The following inclusion criteria were used to select the participants.

- They must have worked for more than 2 years in that particular health facility.
- They have to work routinely with the health data records in the health facility
- They have to work directly in the compilation of the Facility Summary Sheet

Pretest of Instruments

The data collection tools were pretested in one health facility not participating in the research study by the researcher. The tools pretested included the quantitative data collection tool as well as the interview guide for the qualitative data. This pretest was done to ensure that the data collection tools collect information that will help towards the achievement of the goals and objectives of the study (Collins, 2003). In the pretest of the instruments, consent was obtained from the head of the health facility; data records were extracted from the health facility daily registers, monthly summary forms to fill the data collection tools. In addition, one interview was conducted with the facility head using the semi structured interview guide. A few modifications were made to the interviewer guide with more instructions on avoiding leading questions or statements included in the guide. The data from the pretest was not used in the final analysis of this study.

Analysis was done on the pretest data to ascertain the reliability of instruments. Completeness and accuracy ratios were calculated successfully for the pretest data. Although no changes were made to the data collection instruments after the pretest, guiding notes on how to ask questions were included. These notes emphasized the need to avoid leading phrases that may influence responses. Phrases ending in words like “as you know” or “as expected” were some of the phrases to be avoided in the interviews.

Data Collection

The researcher and two (2) research assistants collected data in the months of July and August 2015. The assistants were recruited from a local University and were completing their undergraduate programs in the Faculty of Basic Medical Sciences. The research assistants had no prior relationship to the participants in the research study. In preparation for data collection, the research assistants were trained on how to collect the data using the data collection tool. The research assistants visited all the selected health facilities and administered the research instruments. The research assistants extracted the data from the three sources of data for the quantitative data analysis. The researcher visited all the sampled health facilities and conducted the key informant interviews.

The accuracy of the data collected by the research assistants were checked by the research through a randomly cross checking of the data collected from two health facilities. All reported values were recounted in the different sources of data and a high congruence between what was collected and what was available in the health facilities were found.

Quantitative Data Collection Methods

In the case of the quantitative data, two data collation sheets were used to collect the data on accuracy and completeness. The first data collation sheet developed to assess completeness and consistency (Appendix 3) was used to collect data on the completeness of the sampled 147 data elements in 12 health facilities over a retrospective three-month period (April – June 2015). This data was collected from three data sources (Daily Register, Facility monthly Summary and the DHIS). The three data sources were compared to determine the completeness percentage of each source.

The second data collation sheet assess the accuracy of the health information data values of the sampled 147 data elements in 12 health facilities over a retrospective three-month period. The data was extracted from the following sources: Daily Registers at the health facility; Monthly Summary submitted to the District; Electronic database (DHIS). In this case, the actual values extracted from the three sources were captured on a data collation sheet for each health facility for a retrospective period of three months (Appendix 3). The data verification process is to see the consistency of values between the facility source documents and reported information to the districts and the DHIS. This process is conducted through a calculation of a verification factor using differences between the values in the different data sources for each data element and each facility (Ronveaux et al., 2005; World Health Organization, 2006).

Qualitative Data Collection Methods

For the qualitative data collection, a semi structured interview guide (Appendix 4) was used to interview key informants. A semi structured interview guide was used because it allows for both a structured yet liberal approach to obtaining responses from individuals (DiCicco-Bloom & Crabtree, 2006). The interview guide covered the following key areas: data collection process and practices, facility workload, work incentives, human resource development and supervision. The interviews were conducted face-to-face by the researcher and recorded using an audio recording device.

Data Analysis

In the case of the quantitative data, the data was analysed for the following three factors: completeness, consistency and accuracy. Completeness was calculated using the Data Collation Tool 1 and analysis was done using the following processes:

1. Counting the number of values present (for the sampled data elements)
2. A value of 1 was attributed to a data element whose value was present in the summary form or the DHIS.

3. The cumulative values were based on the number of expected values for the facility by using the Daily Register at the health facility used to capture daily records in the health facility as the benchmark for the expected values in both the Monthly Summary Form as well as the DHIS.
4. A value captured in the Daily Register is expected to be capture in both the Summary Form and the DHIS. Percentages were calculated using the sets of seen and expected values.
5. Collation and cleaning of the data was done in Microsoft Excel and statistical analysis was done on JASP software (Version 0.7.5.6). The cleaning was done by eyeballing the values and identifying numerical errors (Hellerstein, 2008). The errors seen included decimal points in some data values as well as alpha numeric characters. These errors accounted for 1.5% of the total values and were dropped from the dataset.

The accuracy of the data was determined using the consistency of the data values across all sources as a proxy measurement. It is expected that the values recorded from the daily registers represent the true representation of the events that took place in the health facility and is also the value that is transmitted to the monthly summaries and the DHIS.

1. The consistency in the values at these different data aggregation and transmission levels was evaluated using the verification factor methodology (Measure Evaluation, 2008; Ronveaux et al., 2005). Values in the source documents (Daily Registers) as well as in the Summary forms and the DHIS were obtained and recorded in the Data Collection Tool (Appendix 3). Consistency between the values for a particular health facility and period was measured using the calculation below.

$$\frac{Value\ in\ Source\ Document}{Value\ in\ Montly\ Summary\ Form\ \ or\ DHIS}$$

2. For each data element and month, the value in the source document (Daily Register) was divided by the values in the monthly summary and the DHIS. This generated 2 verification factors for each data element and each month.
3. An average verification factor was calculated from the two factors obtained for each data element and recorded. An average verification factor was calculated for each health facility using the average factors for each data element and for all the months reviewed. Collation and calculation of ratios were done on Microsoft Excel sheets and the final figures were put into the JASP software for statistical analysis.

For the qualitative data, semi-structured interviews were conducted with 13 individuals. Although we had planned to interview 1 staff per health facility making a total of 24 interviews, eleven (11) people were excluded from the interviews because they had only worked in the health facility for a period less than 3 months. The researcher is satisfied that saturation of data was reached with the thirteen interviews.

During the interviews, each selected health worker was interviewed individually after obtaining signed informed consent forms. All interviews were recorded with a voice recorder. During the interviews, the researcher looked out for themes and patterns and some of these were probed further during the interview. This was done to harvest rich descriptions about the data quality practices and issues in the health facility as perceived by health facility staff (DiCicco-Bloom & Crabtree, 2006). Each voice recording was then transcribed verbatim into transcripts for thematic analysis. Thematic analysis was utilized to look for similar phrases and ideas in conversations to form themes representing the major streams of thought of the interviewees (Thomas & Harden, 2008). The responses were analyzed by categorizing common patterns and themes on the factors that affect data quality. The identified themes were later consolidated to identify factors affecting the quality of data in the health information system.

Rigor

Validity is described as the extent to which an instrument measures what it has been designed to measure in a research study (Kimberly & Almut, 2008). The validity of research using mixed methods as we have done in this study requires the establishment of the quality of both quantitative and qualitative methods and tools (Onwuegbuzie, Anthony Burke, 2006). The data collection tool used in calculating the verification factors in this study has been used and validated in many studies (Chen et al., 2014; Ronveaux et al., 2005). The validity of the data collection tools was ensured by pretesting the tools at one health facility. The pretesting tools ensured that potential errors were detected and corrected thus ensuring usability and appropriateness of the tools (Kimberly & Almut, 2008).

The use of open-ended questions during interviews ensures satisfactory data saturation (Fusch & Ness, 2015). Data saturation is described in literature as a position in qualitative studies where the researcher can no longer obtain more information on a particular topic from the sample size (Russell & Gregory, 2003). The varying background of the informants interviewed (nurse, midwife, doctor, health information officer) in the study ensured triangulation based on data sources as similar responses were received despite the differences in the background of respondents (Shenton, 2004).

Scrutiny of the research data was also done by the study supervisor lending credibility (Russell & Gregory, 2003; Shenton, 2004) to the qualitative data obtained in the study.

Ethical Statement

Ethical approval from the University of the Western Cape Research and Ethics Committee as well as the Oyo State Ministry of Health was obtained (Appendix 7 and 8). Health data in Nigeria and by extension Oyo State is usually fiercely guarded and access to records usually require administrative permission either at the health facility level or district. Thus, the necessary administrative approvals from the State and district health authorities duly informing them of the nature of the study and the possible outcomes were obtained. These approvals were sought with formal requests to the local authorities and they responded with approval letters.

A clear and informative participant information sheet was explained and given to all interviewed staff in the health facility. This was accompanied by a consent form that was signed by the respondents once they agreed to participate in the study. Research and data collection was only conducted when respondents and facility staff fully understood and accepted the terms of the research by signing the consent form. In order to ensure that there was full understanding, participants were asked questions after the explanation process and where there was confusion or doubt, clarifications were provided.

The participants were also informed that were free to withdraw from the interview at any time without any repercussions. The participants were assured that their identities as well as the facilities will be kept anonymous to protect the identity of participants. Anonymity was assured using identifiers instead of actual names.

The published result of the study is expected to have some implications for people working in the health information sector of Oyo State. For example, while it will be beneficial to assess the quality of routine health information in the State, specifically where the weaknesses and gaps can be identified and solutions proposed), it can inadvertently hurt some individuals working either in the health facilities, districts or the State. In addition, the management staff may take poor data quality reports as a reflection of poor worker attitude and take adverse actions. In an attempt to address possible negative outcomes, the identities of health facilities involved in the research have not been included in the final report and the overall objective and purpose of the research which is to improve the health information system has been emphasized to the Ministry of Health in the State.

CHAPTER FOUR - RESULTS

Introduction

The results chapter is divided into two sections: the results from the quantitative analysis describing the data completeness, data accuracy and consistency, data quality by facility ownership and data element disaggregation; the results from the qualitative data analysis describing the health record responsibility, data collection process and management, data quality practices and challenges. The chapter concludes with a summary of the quantitative and qualitative findings.

Data Completeness

An average completeness of 89.42 % was found in the Monthly Summary Form, 65.24% in the DHIS computer software and an overall average of 77.33%. Completeness values were worse in the DHIS with just 54.17% of sampled health facilities with completeness values above 70. It will appear that the process of data transfer between paper tools and the electronic DHIS lead to some data loss. The following tables outline these details.

Table 1 Summary Form Data Completeness values for health facilities

Completeness Range	Percentage Facilities (N=24)
<50%	8 (2)
50-70%	13 (3)
70-100%	79(19)

Table 1 above represents the data completeness of the summary forms in the twenty-four facilities determined by a completeness range. 79 % of health facilities had completeness values between 70 and 100%, 12.5 % had between 50 - 70% while 8 % of the facilities had completeness values less than 50%.

Table 2 represents the data completeness of the summary forms in the twenty-four facilities determined by a completeness range. 54 % of health facilities had completeness values between 70 and 100%, 25 % had between 50 - 70% while 20 % of the facilities had completeness values less than 50%.

Table 2: DHIS Data Completeness values for health facilities

Completeness Range	Percentage(N)
<50%	20.83(5)
50-70%	25.00(6)
70-100%	54.17(13)

Table 3 below shows the average (Summary and DHIS) completeness values. 66 % of health facilities had completeness values between 70 and 100%, 18 % had between 50 - 70% while 15 % of the facilities had completeness values less than 50%.

Table 3: Average Data Completeness values for health facilities

Completeness Range	Number of Facilities	Percentage(N)
<50%	3	14.58(3)
50-70%	5	18.75(5)
70-100%	16	66.67(16)

Table 4 below shows the descriptive statistics for completeness values across the 2 data sources for data element values. The mean was lowest for the DHIS (65.24) indicating a high loss in data transfer.

Table 4: Descriptive Statistics for Completeness values

	Summary	DHIS	Average
Mean	89.42	65.24	77.33
Std. Deviation	18.36	23.66	18.37
Minimum	43.33	7.140	32.69
Maximum	100.0	90.48	95.24

Accuracy and Consistency

To calculate the verification factor, the value of the data element for each month in the source document (Daily Register) was divided by the values in the monthly summary and the DHIS. The calculation generated two verification factors for each data element per month. An average verification factor was calculated from the two verification factors obtained for each data element and recorded. In addition, an average verification factor was calculated for each health facility using the average factors for each data element based on all the months reviewed.

Accurate and consistent data is expected to fall between a score of .85 and 1.15 (Ronveaux et al., 2005). The study found an overall average figure of 1.16 as the verification factor for all sampled health facilities. The verification factors differed significantly between Summary Sheets and DHIS as well as in Facility types.

Table 5: Distribution of Verification Factor Values for Sampled Health Facilities by Data Source

Verification Factor	Summary Form	Summary Form Percentage	DHIS	DHIS Percentage
<0.85	0	0.00	1	4.17
0.85 - 1.15	19	79.17	13	54.17
>1.15	5	20.83	10	41.67

Table 5 above shows that verification factors was mostly in the abnormal range for DHIS values with only 54.17% of sampled facilities having average verification factors within expected range (0.85 and 1.15). Verification factors were worse in the evaluation of DHIS values, 1.26 (N=24) and in the privately owned health facilities (1.28). The Summary Forms had a higher percentage of facilities with normal verification factors (79 %).

The majority of the verification errors was due to "under reporting – values lower than the source values" with only one facility with VF less than 0.85 which will qualify as "over reporting" (English, 2014; Mutale et al., 2013). English et al (2014) found under reporting in neonatal deaths and fresh still births in facility summaries and over reporting of live births in Kenya.

Table 6: Descriptive Statistics for Verification Factors

	Average Verification Factor	Summary Form Verification Factor	DHIS Verification Factor
Mean	1.158	1.140	1.258
Median	1.055	1.000	1.130
Mode	1.030	1.000	1.050
Std. Deviation	0.2544	0.3259	0.3930

In Table 6, the lowest mean values for the verification factor were found in the Summary Forms (1.140). The mean value in the Summary Forms was also within the expected range for the verification factor. A slightly higher mean in the DHIS shows that the quality of the data reduced after the transmission to the DHIS software.

Verification Factor Value Distribution among Types of Health Facilities

The average values for verification factors were obtained for the sampled health facilities. Using the mean as a measurement stick can often mask the distribution of values among the health facilities specifically outliers. In order to unmask this value and show the true distribution, scatter plots have been used to analyze the deviation of the different facilities from the expected verification values. In this research study, 29.17% of the facilities had verification factors outside the expected range of 0.85 and 1.15. 70.83% of the health facilities had an average verification factor falling between 0.85 and 1.15. Two health facilities had values that fell extremely off the normal distribution, this result is mostly likely due very poor data management in these facilities or the figures provided by the health facility are fabricated.

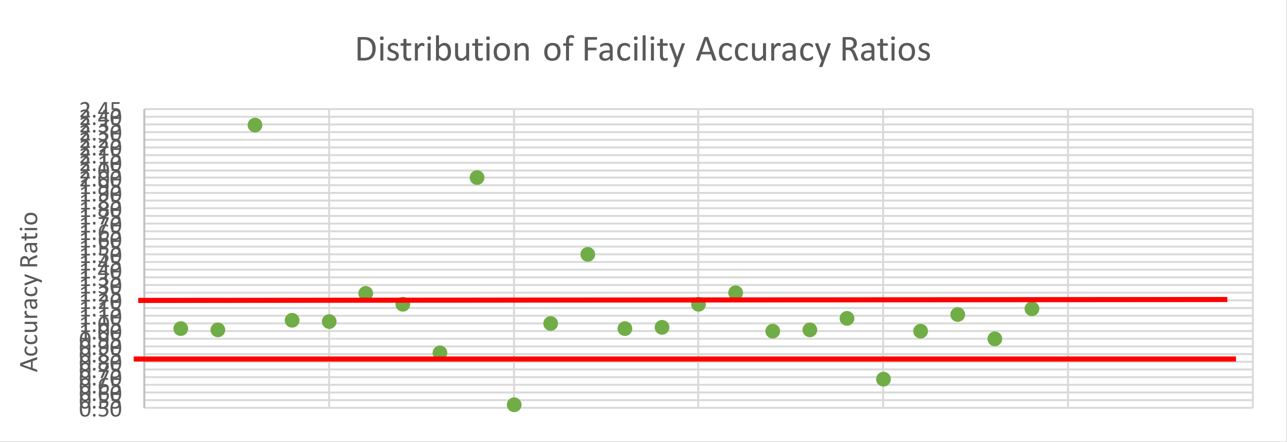


Figure 2: Distribution of facility accuracy ratios

The scatterplot illuminates that the majority of the health facilities fall in the expected band of verification factor (0.85 – 1.15).

Analysis of Facility Ownership and Data Element Dissaggregation Types (Public and Private Health Facility)

There are two (2) groups of facilities based on the ownership in the State. Private health facilities are owned and operated by private individuals or organisations while public health facilities are owned and operated by the Government.

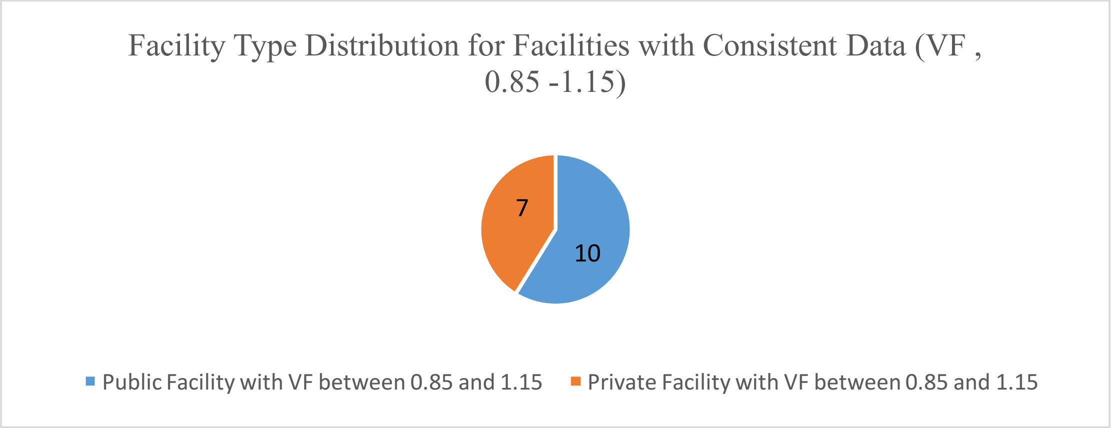


Figure 3: Distribution of Verification Factor values by Facility Type

Figure 4 above illustrates that more public health facilities had verification factors within the expected range compared to private health facilities.

Table 7: Completeness by Facility Type

	Completeness less than	Completeness greater than 75%	Total
Public	2	13	15
Private	4	5	9
Total	6	18	24

Table 7 above shows that the public health facilities had a higher percentage of facilities with average completeness values above 75%. However, it was found not to be statistically significant with a p-value of 0.1501.

Table 8: Verification factor by Facility Type

	Verification Factor within normal	Verification Factor outside normal	Total
Public	11	3	14
Private	4	6	10
Total	15	9	24

Table 8 above shows that public health facilities had a higher percentage of facilities with verification factors within normal ranges. However, it was found not to be statistically significant with a p-value 0.0918 (2 tailed Fisher's exact test was used to calculate the p-value). Although there was a slight propensity for public health facilities to have better completeness rates and verification factors, there was no significant statistical association between facility type and completeness values or verification factors.

Data Element Analysis

The analysis was done from the sampled data elements to investigate the quality of the data using verification factors. Two groups of data elements were analyzed: one group of data elements without disaggregation and a second group of data elements with disaggregation. Verification factor used in the analysis was generated from the facility based data (Registers vs Monthly Summary Form)

Table 9: Verification Factor for Data Element Groups

	VF Within Normal Values	VF Outside Normal Values	Total
Data Element without disaggregation	39	9	48
Data Element with disaggregation	38	61	99
Total	77	70	147

The two-tailed P value is less than 0.0001 making the association between disaggregation and verification to be statistically significant. The analysis shows that data elements without disaggregation were more likely to produce values with normal verification factors that data elements with disaggregation.

Qualitative Data Analysis

A total of thirteen (13) face to face semi-structured interviews were conducted. The thematic analysis found the following recurring themes the health record responsibility, data collection process and management, data quality practices and challenges.

Health Record Responsibility

The analysis of the data highlighted two main issues for the health record responsibility. Firstly, though the policy of the government is to have one health record professional in all health facilities this was not the case in the sampled health facilities: "*No – we do not have a health record officer in the health facility, myself and my assistant handle all data records in the health facility*" (Key Informant 1) Secondly, the bulk of the data management processes rests with the clinical staff: "*The nursing staff are responsible for collating the data when they have the time"* (Key Informant 2). In summary, there is a significant gap in the availability of required staff to manage routine health information systems in the state.

Data Collection Process and Management

Data collection in the health facility often starts with encounter with the patient when health staff record encounter details in paper notes.
When a patient comes to the health center, we register the patient first in our daily attendance register and after the patient is seen we then register the patient into the appropriate daily register. (Key Informant 1)

This data in then used to populate relevant daily registers usually at the end of the day by the health staff.
"At the end of the month we add up the data from the daily registers and put the figures on the NHMIS monthly form given to us by the LGA" (Respondent, Data Clerk)

Different daily registers are populated with details of clinic visits based on the type of service rendered such as outpatient attendance, immunization, antenatal care and family planning.

At the end of the month, a health facility staff who is usually dedicated to do this , brings the registers together and records summaries in a summary form.

"At the end of the month we add up the data from the daily registers and put the figures on the NHMIS monthly form given to us by the LGA" (Respondent, Data Clerk)

Box 1: Description of the data collection process

> *"We have health facility registers in the health facility. When a patient comes to the health center, we register the patient first in our daily attendance register and after the patient is seen we then register the patient into the appropriate daily register. We see mostly antenatal and immunization patients so we use those registers a lot and often"* (Key Informant 1)

At the end of the month, a staff (or a team of staff) of the health facility aggregates the data from all the daily registers to populate monthly summary forms. These records are collated on a daily basis as health facility staff interacts with patients and clients.

> *"At the end of the month we add up the data from the daily registers and put the figures on the NHMIS monthly form given to us by the LGA"* (Respondent, Data Clerk).

The monthly summary forms are completed in triplicate copies, a copy is retained in the health facility, a copy is sent to the LGA officer for capture into the DHIS.

> *"Usually, we use notebooks to collect patient data during consultation and then after the end of the day's work in the evening, I fill the relevant daily register provided by the government"* (Key Informant 3).

Although based on guidelines from the Federal Ministry of Health, this process expected to be done during or after patient encounters is sometimes forgotten or completed on the next day.

> *"I do not always remember to complete the daily register at the end of the day, we can be busy with other things or just tired"* (Key Informant 4).

In summary, the stipulated guidelines for the management of data are available in the health facilities.

Data Quality Practices

The main finding for data quality practices is that there were no specific processes dedicated to ensuring the quality of the data although a lot of time was spent in collecting and collating data into registers and summary forms.

> *"We don't have any specific things we do to ensure data quality, maybe entering data on time? Not very sure we do any specific thing".* (Key Informant 1)

In addition, the Monthly Summary Form has to be signed by a superior officer at the health facility verifying the data collated thus a data quality mechanism. However, this verification is not usually done and forms only signed to allow timely submission to the LGA.

Furthermore, the staff expects the quality to the verified and ensured by the next level officers at the district during routine data quality assurance visits to the health facility by the District and other sub national level officers. The Data Quality Assessments are conducted periodically by staff from the LGA and State (Abah, 2012). However, this was not done regularly and when done, only a small fraction of data elements are verified at the health facilities.

Challenges

The challenges facing health staff in the health facility in data collection and management included a severe deficiency in the availability of relevant human resource for health data management, clinical work overload for most health staff, use of complex and bulky forms, poor feedback mechanisms, delay in completing data records, lack of use of generated information at health facilities and inadequate training on health information systems.

Firstly, a lack of adequate human resource to manage health data in the health facilities was a common feature in many of the health facilities, health staff already overburdened with clinical duties are expected to also collect and manage the facility data – this duty was often relegated to the last hours of the day when health workers are already fatigued: *"If we have too many patients or on immunization days we may forget to enter all the patients in the daily registers or only do that after some days when we may have forgotten some of the details" (Key Informant 8).*

In addition to the scarcity of human resources is the complexity of the forms and registers expected to be completed by the health staff to satisfy the reporting requirements of the routine health information system. “*The forms are too big, we have many other things we do in the health facility, the government should employ more staff”* (Key Informant 9).

Many of the respondents had never had any formal training in the use of the data reporting tools, this poses a serious challenge in ensuring good quality data, when individuals do not understand data definitions it is difficult for them to understand or ensure that the correct data is collected in the health facility: “*I will wish that I attend more trainings on data management but there is no sponsorship or opportunities*” (Key Informant 11).

The delays in the completion of daily records or monthly records leave room for missing data records and incorrect entries. Feedback on data submitted to the district and captured on the DHIS was found to be very poor as most respondents had either never been given feedback on values captured on the DHIS by District Officers.

> “*We do attend meetings monthly to submit our data to the LGA but we don’t get any feedback on the data submitted to the LGA”* (Key Informant 1).

The findings of the study found incomplete data in facility summaries as well as in the DHIS. The completeness of the data was found to be worse in the DHIS records with more missing values compared to the expected values in the daily facility registers. The analysis of the data values transmitted between different data aggregation levels showed marked differences in the values reported in the daily facility registers, facility monthly summary sheets and the DHIS. The study also investigated any association of the quality of data and whilst no statistical association was observed in the type of facility ownership while the number of disaggregations in the data elements had a statistically significant impact on the quality of the data.

Qualitative analysis found a number of factors that affect the quality of routine health information including the availability of appropriate human resources, data management practices and process as well as specific capacity building for health facility staff. Health facility staff working on collection and collation of routine health information in the facilities reported that they were often severely overworked since they had other clinical duties to attend to. They also did not have regular training programs for development of specific health information skills. There was also a lack of adequate data management processes in place to ensure the production of good quality data at the health facilities on a routine basis.

CHAPTER FIVE – DISCUSSION

The chapter outlines the main findings of this study, explaining the implications with existing literature and setting a background for the provision of recommendations. The objective of the study was to investigate the quality of data within the routine health information system of the study area as well as the factors affecting data quality. In investigating the quality of the data, the study focused on two (2) dimensions of the data quality, completeness and accuracy using quantitative methods. Additional analysis was done on the effect of certain properties of information systems that may influence the quality of data. These properties include the ownership of health facilities as well as the complexity of data collection forms. These are discussed in this chapter. The chapter also discusses the factors affecting the quality of data as analysed in the qualitative data obtained from the study.

Data Completeness

Data completeness is one of the most important dimensions of data quality and was measured in terms of the percentage of expected data values present at the different data aggregation methods: the facility health data monthly summary form and the DHIS. The study found the data completeness for the Monthly Summary form at 89.3% whilst in the DHIS 65.2% with an overall average completeness of 77.3%. The completeness values in this study were found to be generally lower compared to similar countries. A study in Ghana found completeness to be estimated at 99.1 % for routine maternal health data in aggregate forms and 100% in the DHIS (Amoakoh-Coleman et al., 2015). However, lower completeness rates were found in Tanzania with a value of 64.2% for paper records (Simba & Mwangu, 2009). In South Africa, where the DHIS was developed, the completeness of HIV data in the DHIS were found to be 50.3% (Mate, Bennett, Mphatswe, Barker, & Rollins, 2009).

In Nigeria, the DHIS is the final repository for routine data generated from health facilities and is the source of information used for planning and decision making by the majority of the health managers in the country. This research study found that the completeness for publicly owned health facilities (86.66%) in the LGA were higher than in privately owned health facilities (55.55%). In the LGA, the private sector provides a significant portion of health care in developing countries and will contribute significantly to the data available in routine health information systems (Berman & Rose, 1996). Private health facilities in Nigeria usually feel a sense of non-duty to the government in terms of routine data submission to administrative levels. This phenomena is not limited to Nigeria only as other countries have also reported difficulty in the integration of public and private health information systems (Matshidze & Hanmer, 2007; Streveler & Sherlock, 2004). Furthermore, the challenges identified for private health clinics are; the absence of appropriate policies and framework , availability of human resources and a lack of engagement of the private sector by governments in routine health information systems.(Matshidze & Hanmer, 2007).

Our finding of lower completeness rates for higher levels of data aggregation and transfer shows that data values are lost in this process. This can easily be reversed through a more vigilant process of validating data aggregated from one medium to another (Cheburet & Odhiambo-Otieno, 2016).

Data Accuracy and Consistency

In Nigeria, paper records are used to collect and store individual level patient data that is also recorded on daily facility registers which are then used to collate health facility monthly summary forms. It is expected that some concordance exists between the data in the health facility registers, the monthly summary forms and the DHIS. However, the study found a large discordance between the values in the paper health facility registers and the monthly summary forms as well as the DHIS (14% and 25% respectively). This is similar to the findings (Amoakoh-Coleman et al., 2015; Simba & Mwangu, 2009; Wright & Odama, 2012) where inconsistencies were also found between values in source documents (patient notes or registers) and transmitted aggregated data either in paper summary forms or electronic databases. Amoakoh et al found data transfer errors between 14.6% and 35.6 % looking at total number of women with hemoglobin checked at 36 weeks' gestation data element in the Ghana DHMIS. In Tanzania, Simba and Mwangu found a variance of 36% to 92% in reported health records in a district.

Furthermore, we found this discordance to be much worse in the DHIS In this study with a weighted average verification factor of 1.25 (Expected is between 0.85 and 1.15). This is different from the finding of Amoakor et al (Amoakoh-Coleman et al., 2015) looking at the data in the DHIS within the Ghana Health Service where almost no discordance was found in the reported aggregated data and the values in the DHIS. The concordance between reported aggregated data and DHIS values was also found by Mate et al (2009) in South Africa looking at PMTCT data. Our finding may be due to the maturity of the DHIS system in the different environments.

This finding from our study shows both inconsistent data as well as over reporting of most data values in the DHIS for most data elements. Over reporting refers to a phenomenon whereby values higher than the correct values are recorded and reported. Over reporting was also reported in a number of studies and is described as a phenomenon in which values higher than values in source documents are recorded and transmitted. This can be due to the desire by health workers to receive a higher budget to manage certain conditions (Murray et al., 2003) or genuine data errors in the reporting process.

One of the main findings of this study was the absence of routine data quality checks in the health information system. The lack of a robust system to ensure routine data quality checks will likely account for the poor quality of data seen in the DHIS. The poor quality of data in the electronic system is itself not because of deficiencies in the software but likely because proper data quality methods and checks have not been put in place in the health information system.

Health Facility type and data quality

The study found no statistically significant association between the type of health facility and the quality of data in terms of completeness and consistency (p 0.15). This finding which may be due to the small sample size shows a weak indication that privately owned health facility increased chances of having poor data completeness and consistency of data values across aggregation levels raises a big concern about the contribution of privately owned health facilities in Nigeria to routine health information systems. Private health providers are often excluded from routine health information systems although they provide a significant portion of health care delivery in many countries. In Nigeria, some States have adopted legislations and laws to compel private health facilities to report routine health data to the government, some of these private health facilities in turn submit data to fulfil the obligation rather than see themselves as part of a unified health system. Ahanhanzo et al (2014) in the study in Benin did identify that private health facilities were more prone to data quality challenges because they were mostly excluded from the Routine health information system in the country (Glèlè Ahanhanzo et al., 2014)

Complexity of data reporting tools and quality

Data is often collected in routine information system by various dimensions for example gender and age. Age disaggregation can be simple with only 2 age groups or as complex as more than 5 age groups. The presence of multiple data element disaggregation in a form often complicates reporting forms and creates a burden on data collectors or collators with an impact on the quality. Our study found a significant association between data disaggregation and problems with data quality ($p < 0.0001$). This relationship between the complexity of reporting forms and the quality of the data was also demonstrated by Hahn et al. (201) looking at routine information systems in Kenya. They found that the time-consuming process of completing the data collection forms had an impact on the quality of the data in two (2) hospitals in Kenya. Apart from the association from the quantitative data analysis, our qualitative data also showed that this was a factor for the performance of health workers in the collection and management of routine health information in health facilities.

Factors affecting Data Quality

One of the objectives of this study was to investigate the factors that affected the quality of routine health facility data. The factors refer to conditions or phenomena that had a direct or indirect impact on the quality of data. We used qualitative methods to collect the data by conducting interviews with key informants at health facilities. The qualitative research revealed insights on the factors that influence the quality of data for example a number of practices and analysis from the key informant interviews. The factors identified by the findings were mainly on human resources, infrastructure and processes involved in the data collection and management of the routine health information.

Human Resources

A pervasive finding in this study is the absence of dedicated health staff for routine health information. A major finding of the study was that a lot of the health workers that worked in the collection and management of routine health data in the health facilities did this in addition to many other clinical duties. The clinical duties often take precedence as they were mostly clinical workers in this research study nurses or midwives. The clinical health workers doubled as health information officers only increased pressure on the performance of these health workers, this doubling or often times referred to as task shifting (WHO, 2007) adopted as a result of shortages of health information officers will have an adverse effect on the quality of data generated in the health facility.

The situation is compounded with the widely reported shortage of skilled health workers in many sub Saharan countries (Kinfu, Dal Poz, Mercer, & Evans, 2009; WHO, 2013). Clinical officers are more inclined to pay more attention and time to actual clinical duties leaving health information management as a secondary duty (Hahn et al., 2013). This makes it difficult for the workers working on the routine health information system to pay the required attention on the management of routine health data.

The importance of good quality human resource at the health facility level is further emphasized by the study carried out by Mphatswe et al., 2012 in South Africa where a specific intervention of training health workers on routine health information systems significantly improved the quality of routine health information systems. We found a severe lack of training on the management of routine health data among the health workers who worked on routine health data. This study found that the capacity of staff available to carry out data management activities in the health facilities was low. Although these facility officers had attended trainings on data management and quality, the frequency of those trainings were low with an average of once in two (2) years. The high turnover of staff at the health facilities compounded this challenging environment in the health system.

Feedback on data collected and submitted to the district office seems to be a major problem as evident by the wide gaps in the completeness of data and the discordance between data in the State. Regular feedback from district information officers who do the data capturing on the DHIS to the facility staff will reveal the discrepancies in the data available in the DHIS, Summary forms and health facility registers. The absence of feedback increases the lack of information use at the health facility and district levels. In addition, the knowledge of data quality and its importance among the health facility staff was quite low.

Data Collection processes

The study found that the way data collection process is done within health facilities affects the quality of the data. Although, an electronic system for routine health information system is in place in Nigeria (DHIS 2), a large part of data management is paper based. Daily patient encounters are recorded on government approved registers. Responsible staff is expected to complete these daily registers which are then aggregated into monthly summary forms at the end of the month. This process is a largely human resource driven one and the success depends on the way it is done. As previously mentioned, the lack of capacity of qualified human resources at most of the health facilities has a major impact in data collection process.

The burden of reporting on the health workers would have an impact on the overall quality of the data. Nicol, Dudley, & Bradshaw (2016) supports this in their review of PMTCT data in South Africa stating that “*weaknesses in data-collection processes and the multiplicity of data collection tools kept by multiple care givers at the facility level, which creates opportunities for underreporting and data compromise affect the quality of data in routine health information systems*” (Nicol et al., 2016, p. 3).

Data use

There is evidence that improved use of routine health data improves the quality of the data as more attention is paid to the usability of the data (Braa, Heywood, & Sahay, 2012). Poor data quality of available routine health information has also been given as one of the major reasons why there is poor use of the information available in routine health information systems (Wagenaar et al., 2015). This intricate relationship between the use of data and data quality underpins the importance of this factor. Our study found very little evidence that the generated routine health information system was been used. This lack of use was demonstrated by a lack of feedback to the health facilities from district officers who are the primary recipients of the aggregated data from the health facilities. If generated information within routine health information systems is not used for decision making, then the entire essence of the health information system is lost. Lack of use creates a vicious cycle of lack of data demand and production of good quality information.

CHAPTER SIX – CONCLUSION AND RECOMMENDATIONS

This study revealed poor data quality in routine health information systems of Oyo State Nigeria with gaps in completeness and accuracy of health information. Incomplete data was found in aggregated data reported from daily patient encounters in monthly summary forms as well as the electronic database (DHIS). The study also found inaccuracies in the reported values with discrepancies between facility register recounts, summary form values and the values in the DHIS. These inaccuracies were progressively worse as the data was aggregated and captured (Summary forms and DHIS). Incomplete data and data inaccuracies were found to be present in both public and private health facilities. These problems were however slightly more pronounced in the private facilities. Quality issues were also worse for data elements that had multiple disaggregations making their collection and reporting more complex.

Factors identified as affecting the quality of routine health information systems were human resources, data management processes and infrastructure. Routine health information systems remain a vital source of information for health management decision making and use. However, the data quality gaps revealed can easily be mitigated by recognizing the factors that will affect or influence the quality of data.

Recommendations

For routine health information systems to truly attain the goal of providing good quality data for decision making, the business of ensuring good quality data must be paid attention to. Based on the findings of this study, the following recommendations are suggested;

- There should be the development of appropriate human resources for routine health information systems. Specifically, dedicated trained staff should be available in health facilities to manage routine health information systems. In addition, there also must be a system to ensure that there is continuous training for these staff on a routine basis.

- A data quality system should be in place that ensures that there is routine data quality checks and assessments across all the repositories of data within the health information system. External checks should be conducted by the LGA and State staff. Internal processes for checks at the health facility should be instituted within the facility. For the facility staff, verification of collated summaries must be validated and signed off by senior staff and summaries also discussed before submission to the district or higher levels. This will ensure that a facility based health information system is in place to provide a base for the good functioning of the overall health information system.

- There should be a forum where data is presented at the district level involving health facilities. The use of information at the health facility will shift the focus of the health facility staff on satisfying the demand of reporting to the next level to the importance of having good quality data for local decision making.

- Data sheets should be printed, presented and discussed on a regular routine basis at both facility and district levels.
- The process of designing data collection tools for routine health information systems should be a carefully considered with due considerations given for burden on health staff and clarity of data elements in terms of definitions.
- Good quality data produced by health facilities and districts should be recognized in a non-financial incentives programme. Thus, creative non-monetary mechanisms should be put in place as a means to enhancing the performance of human resources working on producing data for the health information system.

REFERENCES

Abah, R. C. (2012). The Importance of Data Quality Assurance in Improving Grant Implementation : An Example from Nigeria . *Developing Country Studies*, *2*(7), 67–73.

AbouZahr, C., & Boerma, T. (2005). Health information systems: The foundations of public health. *Bulletin of the World Health Organization*, *83*(8), 578–583.

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. *BMC Research Notes*, *8*(1), 114.

Anifalaje, A. (2009). Decentralisation and Health Systems Performance in Developing Countries: Impact of "Decision Space" on Primary Health Care Delivery in Nigeria. *International Journal of Healthcare Delivery Intitiatives*, *1*(1), 25–47.

Barreiro, P. L., & Albandoz, J. P. (2001). Population and sample . Sampling techniques. *Management Mathematics for European Schools*, 6.

Berman, P., & Rose, L. (1996). The role of private providers in maternal and child health and family planning services in developing countries. *Health Policy and Planning*, *11*(2), 142–155.

Bowen, S., Erickson, T., Martens, P. J., & Crockett, S. (2009). More than "using research": The real challenges in promoting evidence-informed decision-making. *Healthcare Policy*, *4*(3), 87–102.

Braa, J., Heywood, A., & Sahay, S. (2012). Improving quality and use of data through data-use workshops: Zanzibar, United Republic of Tanzania. *Bulletin of the World Health Organization*, *90*(5), 379–384.

Caballero, I., Verbo, E., Calero, C., & Piattini, M. (2007). A Data Quality Measurement Information Model Based on Iso/Iec 15939. *12th International Conference on Information Quality (ICIQ)*, *15939*(Mim), 393–408.

Callaos, N., & Callaos, B. (2002). Towards a Systemic Notion of Information: Practical Consequences. *Informing Science*, *5*(1).

Chahed, M. K., Bellali, H., Alaya, N. Ben, Ali, M., & Mahmoudi, B. (2013). Auditing the quality of immunization data in Tunisia. *Asian Pacific Journal of Tropical Disease*, *3*(1), 65–70.

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.

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), 5170–207.

Chilundo, B., Sundby, J., & Aanestad, M. (2004). Analysing the quality of routine malaria data in Mozambique. *Malaria Journal*, *3*, 3.

Collins, D. (2003). Pretesting survey instruments: An overview of cognitive methods. *Quality of Life Research*, *12*(3), 229–238.

Cong, G., Fan, W., Geerts, F., Jia, X., Ma, S., Zhang, Z., Wang, S., Cong, G., Fan, W., Geerts, F., Jia, X., & Ma, S. (2007). Improving data quality: consistency and accuracy. *Proceedings of the 33rd International Conference on Very Large Data Bases*, *Vienna*, *Au*, 315–326.

DiCicco-Bloom, B., & Crabtree, B. F. (2006). The qualitative research interview. *Medical Education*, *40*, 314–321.

Edgard-marius, O. ;, Alphonse, K. ;, Leve, A. ;, & Makoutode, M. (2014). Data quality assessment in the routine health information system : an application of the Lot Quality Assurance Sampling in Benin. *Health Policy and Planning*, (2010), 1–7.

English, M. (2014). Assessing the ability of health information systems in hospitals to support evidence-informed decisions in Kenya. *Global Health Action*, *7*, 24859.

Fernandes, Q., Wagenaar, B., & Anselmi, L. (2014). Effects of health-system strengthening on under-5, infant, and neonatal mortality: 11-year provincial-level time-series analyses in Mozambique. *The Lancet Global Health*, *2*(8), 468–477.

Fusch, P. I., & Ness, L. R. (2015). Are we there yet? Data saturation in qualitative research. *The Qualitative Report*, *20*(9), 1408–1416.

Gething, P. W., Noor, A. M., Gikandi, P. W., Ogara, E. A. A., Hay, S. I., Nixon, M. S., Snow, R. W., & Atkinson, P. M. (2006). Improving imperfect data from health management information systems in Africa using space-time geostatistics. *PLoS Medicine*, *3*(6), 0825–0831.

Gimbel, S., Micek, M., Lambdin, B., Lara, J., Karagianis, M., Cuembelo, F., Gloyd, S. S., Pfeiffer, J., & Sherr, K. (2011). An assessment of routine primary care health information system data quality in Sofala Province, Mozambique. *Population Health Metrics*, *9*(1), 12.

Glèlè Ahanhanzo, Y., Ouedraogo, L. T., Kpozèhouen, A., Coppieters, Y., Makoutodé, M., & Wilmet-Dramaix, M. (2014). Factors associated with data quality in the routine health information system of Benin. *Archives of Public Health*, *72*(1), 25.

Hahn, D., Wanjala, P., & Marx, M. (2013). Where is information quality lost at clinical level? A mixed-method study on information systems and data quality in three urban Kenyan ANC clinics. *Global Health Action*, *6*(1), 1–10.

Health Metrics Network. (2005). *Issues in health information*.

Health Reform Foundation of Nigeria. (2007). Nigeria Health Review.

Hellerstein, J. M. (2008). Quantitative Data Cleaning for Large Databases. *United Nations Economic Commission for Europe*, 42.

Heywood, A., & Rhode, J. (2001). *Using Information for Action. A manual for Health Care Workers*.

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*(1), 188.

Ireland, M., Paul, E., & Dujardin, B. (2011). Can performance-based financing be used to reform health systems in developing countries? *Bulletin of the World Health Organization*, *89*(9), 695–698.

James, J., Damberg, C., Ryan, A., Agres, T., Schwartz, a, & Dentzer, S. (2012). Health Policy Brief. *Health Affairs*, *19*, 1–5.

Kimberly, C., & Almut, W. (2008). Validity and reliability of measurement instruments used in research. *American Journal of Health-System Pharmacy*, *65*.

Kinfu, Y., Dal Poz, M. R., Mercer, H., & Evans, D. B. (2009). The health worker shortage in Africa: Are enough physicians and nurses being trained? *Bulletin of the World Health Organization*, *87*(3), 225–230.

Ledikwe, J. H., Grignon, J., Lebelonyane, R., Ludick, S., Matshediso, E., Sento, B. W., Sharma, A., & Semo, B. (2014). Improving the quality of health information: a qualitative assessment of data management and reporting systems in Botswana. *Health Research Policy and Systems / BioMed Central*, *12*(1), 7.

Lippeveld, T. (2001). Routine Health Information Systems : The Glue of a Unified Health System. *World Health Organization, Geneva*.

Lippeveld, T., Sauerborn, R., & Bodart, C. (2000). Design and implementation of health information systems Design and implementation of health information systems Edited by. *World Health Organization, Geneva*.

Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, A. J., Ly, A., Gronau, Q. F., Smira, M., Epskamp, S., Matzke, D., Wild, A., Rouder, J. N., Morey, R. D. & Wagenmakers, E.-J. (2015). JASP (Version 0.7)[Computer software].

Makinde, O. (2012). *Assessment of Routine Health Management Information System in Delta State. Nigeria*.

Mate, K. S., Bennett, B., Mphatswe, W., Barker, P., & Rollins, N. (2009). Challenges for routine health system data management in a large public programme to prevent mother-to-child HIV transmission in South Africa. *PLoS ONE*, *4*(5), e5483.

Matshidze, P., & Hanmer, L. (2007). Health Information Systems in the Private Health Sector. *South African Health Review*, 89–102.

Mavimbe, J. C., Braa, J., & Bjune, G. (2005). Assessing immunization data quality from routine reports in Mozambique. *BMC Public Health*, *5*, 108.

Measure Evaluation. (2008). Data Quality Audit Tool. Retrieved December 12, 2014, from http://www.cpc.unc.edu/measure/tools/monitoring-evaluation-systems/data-quality-assurance-tools/dqa-auditing-tool-implentation-guidelines.pdf

Miller, J. S., Musominali, S., Baganizi, M., & Paccione, G. a. (2014). A process evaluation of performance-based incentives for village health workers in Kisoro district, Uganda. *Human Resources for Health*, *12*(1), 19.

Mphatswe, W., Mate, K. S., Bennett, B., Ngidi, H., Reddy, J., Barker, P. M., & Rollins, N. (2012). Improving public health information: a data quality intervention in KwaZulu-Natal, South Africa. *Bulletin of the World Health Organization*, *90*(3), 176–82.

Murray, C. J. L., Shengelia, B., Gupta, N., Moussavi, S., Tandon, A., & Thieren, M. (2003). Validity of reported vaccination coverage in 45 countries. *Lancet*, *362*(9389), 1022–1027.

Mutale, W., Chintu, N., Amoroso, C., Awoonor-Williams, K., Phillips, J., Baynes, C., Michel, C., Taylor, A., & Sherr, K. (2013, January). Improving health information systems for decision making across five sub-Saharan African countries: Implementation strategies from the African Health Initiative. *BMC Health Services Research*. BioMed Central Ltd.

Ndabarora, E., Chipps, J. a., & Uys, L. (2013). Systematic review of health data quality management and best practices at community and district levels in LMIC. *Information Development*, *30*(2), 103–120.

Nicol, E., Dudley, L., & 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*, 60–70.

Onwuegbuzie, Anthony Burke, J. (2006). The Validity Issue in Mixed Research Anthony J . Onwuegbuzie R . Burke Johnson University of South Alabama. *Research in the Schools*, *13*(1), 48–63.

Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data quality assessment. *Communications of the ACM*, *45*(4), 211.

Ronveaux, O., Rickert, D., Hadler, S., Groom, H., Lloyd, J., Bchir, a, & Birmingham, M. (2005). The immunization data quality audit: verifying the quality and consistency of immunization monitoring systems. *Bulletin of the World Health Organization*, *83*(7), 503–510.

Russell, C. K., & Gregory, D. M. (2003). EBN users' guide: Evaluation of qualitative research studies. *Evidence Based Nursing*, *6*(April), 36–40.

Shaw, V. (2005). Health information system reform in South Africa: developing an essential data set. *Bulletin of the World Health Organization*, *83*(8), 632–636.

Shenton, A. K. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education for Information*, *22*, 63–75.

Simba, D. O., & Mwangu, M. a. (2009). Quality of a routine data collection system for health: Case of

Kinondoni district in the Dar es Salaam region, Tanzania. *SA Journal of Information Management*, *7*(2).

Steve, L., & Judy R, W. (2008). Accuracy vs. Validity, Consistency vs. Reliability, and Fairness vs. Absence of Bias: A Call for Quality. In *Annual Meeting of the American Association of Colleges of Teacher Education (AACTE)* (pp. 1–23).

Streveler, D. J., & Sherlock, S. M. (2004). Health Management Information Systems for Resource Allocation and Purchasing in Developing Countries, (September).

Tayi, G. K., & Ballou, D. P. (1998). Examining data quality. *Communications of the ACM*, *41*(2), 54–57.

Teddlie, C., & Yu, F. (2007). Mixed Methods Sampling: A Typology With Examples. *Journal of Mixed Methods Research*, *1*, 77–100.

Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology*, *8*(1), 45.

Vaziri, R. (2012). A Questionnaire-Based Data Quality Methodology. *International Journal of Database Management Systems*, *4*(2), 55–68.

Wagenaar, B. H., Gimbel, S., Hoek, R., Pfeiffer, J., Michel, C., Manuel, J. L., Cuembelo, F., Quembo, T., Afonso, P., Porthé, V., Gloyd, S., & Sherr, K. (2015). Effects of a health information system data quality intervention on concordance in Mozambique: time-series analyses from 2009-2012. *Population Health Metrics*, *13*, 9.

Wand, Y., & Wang, R. (1996). Anchoring Data Quality Dimensions in Ontological Foundations. *Communications of the ACM*, *39*(11).

Wang, R., & Strong, D. (1996). Beyond accuracy : What data quality means to data consumers. *Journal of Management Information Systems*, *12*(4), 5.

Warren, A. E., Wyss, K., Shakarishvili, G., Atun, R., & de Savigny, D. (2013). Global health initiative investments and health systems strengthening: a content analysis of global fund investments. *Globalization and Health*, *9*(1), 30.

Whittemore, A. S. (1997). Multistage Sampling Designs and Estimating Equations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *59*(3), 589–602.

WHO. (2007). Task shifting to tackle health worker shortages. Retrieved December 2, 2015, from http://www.who.int/healthsystems/task_shifting_booklet.pdf

WHO. (2013). Global Health Workforce Crisis. *Global Health Workforce Alliance*, (11).

WHO. (2016). World Health Statistics 2016: monitoring health for the SDGs, sustainable development goals. Retrieved April 5, 2015, from http://www.who.int/gho/publications/world_health_statistics/2016/en/

WHO (World Health Organization). (2007). *Everybody business:Strenghthening Health Systems to Improve Health Outcomes:WHO's framework for Action*.

Wilms, M. C., Mbembela, O., Prytherch, H., Hellmold, P., & Kuelker, R. (2014). An in-depth, exploratory assessment of the implementation of the National Health Information System at a district level hospital in Tanzania. *BMC Health Services Research*, *14*(1), 91.

Witter, S., Toonen, J., Meessen, B., Kagubare, J., Fritsche, G., & Vaughan, K. (2013). Performance-based financing as a health system reform: mapping the key dimensions for monitoring and evaluation. *BMC Health Serv Res*, *13*, 367.

Woodall, J. P. (1988). Epidemiological approaches to health planning, management and evaluation. *World Health Statistics Quarterly. Rapport Trimestriel de Statistiques Sanitaires Mondiales*, *41*(1), 2–10.

World Health Organization. (2006). *GUIDE TO THE HEALTH FACILITY*.

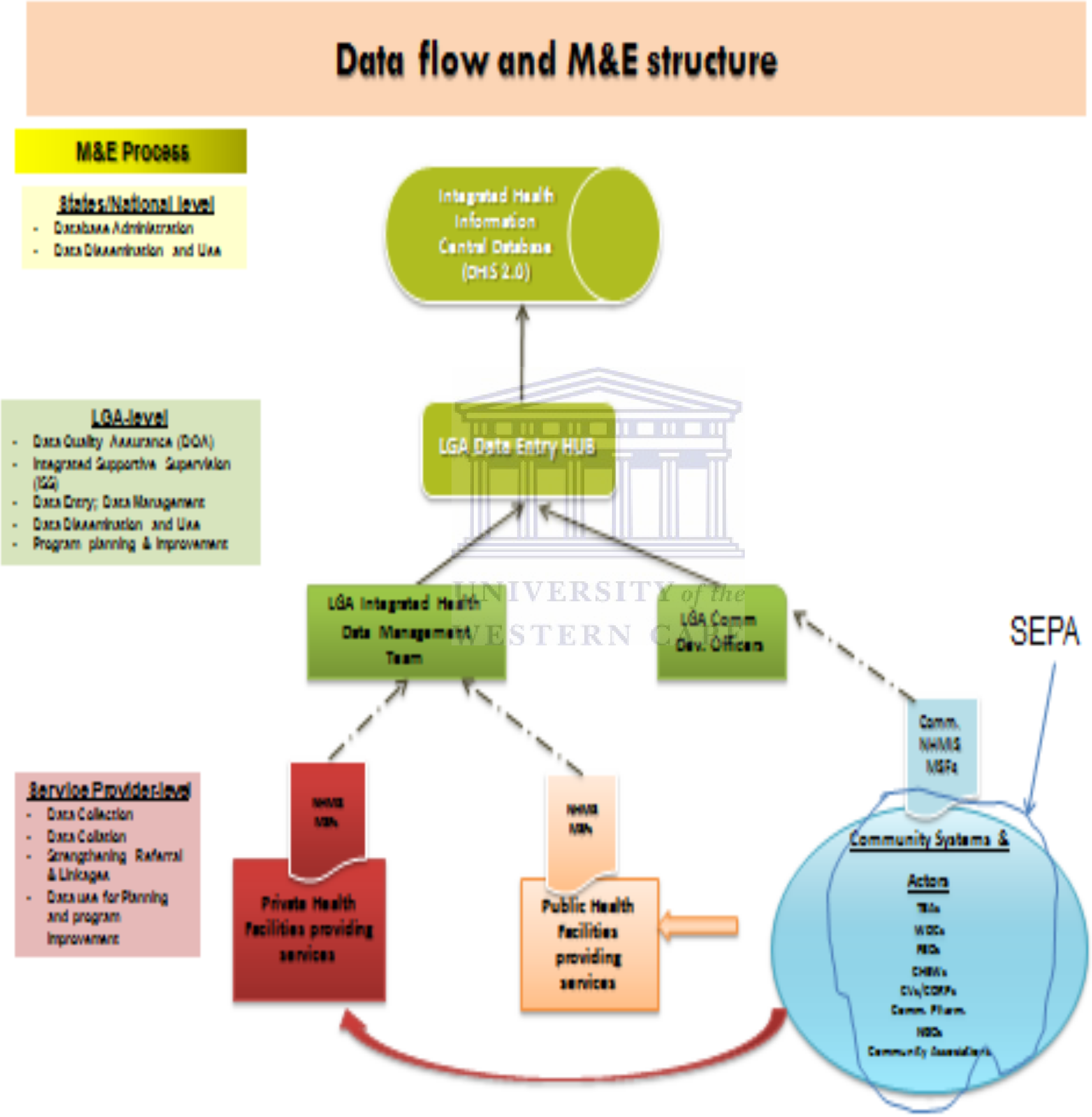
World Health Organization. (2008). Framework and Standards for Country Health Information Systems. *World Health*, *2nd Editio*(January), 72.

Wright, G., & Odama, A. (2012). Health Data Ownership and Data Quality: Clinics in the Nyandeni District, Eastern Cape, South Africa. *Engineering Management Research*, *1*(2), 146–152.

Zozus, M. N., Hammond, W. E., Green, B. B., Kahn, M. G., Richesson, R. L., Rusincovitch, S. a, Simon, G. E., & Smerek, M. M. (2014). Assessing Data Quality for Healthcare Systems Data Used in Clinical Research, (919), 1–26.

APPENDICES

Appendix 1



Appendix 2

NHMIS MONTHLY SUMMARY FORM FOR HEALTH FACILITIES

(Version 2013)

Identification

Health Facility:	Month:	
Political Ward:	Year:	
LGA:	Public: ☐	Private: ☐
State:	Beds:	
Facility code: ☐☐ ☐☐ ☐ ☐ ☐☐☐☐		

Health Facility Attendance

1	**Facility Attendance**	**Male**						**Female**						**Total**
		0 - 28d	29d - 11 m	12 - 59 m	5 - 9 Yrs	10 - 19 yrs	20 yrs+	0 - 28d	29d - 11 m	12 - 59 m	5 - 9 Yrs	10 - 19 Yrs	20 yrs+	

Maternal Health (Ante & Post natal Care)

Total

2	Antenatal attendance - total	
3	Antenatal first visit before 20 weeks	
4	Antenatal first visit 20 weeks or later	
5	Antenatal first visit – total	
6	Pregnant women that attended antenatal clinic for 4th visit during the month	

7	ANC syphilis test done	
8	ANC syphilis test positive	
9	ANC syphilis case treated	
10	Pregnant women who received malaria IPT1	
11	Pregnant women who received malaria IPT2	
12	Pregnant women who received LLIN	
13	Pregnant women who received IFAs (Iron and Folic Acid supplements)	
14	Postnatal attendance – total	
15	Postnatal clinic visits within 1 day of delivery	
16	Postnatal clinic visits within 3 days of delivery	
17	Postnatal clinic visits >=7 days of delivery	

Maternal Health (Labour and Delivery)

Total

18	Deliveries – total	
19	Deliveries - SVD (Spontaneous Vaginal Delivery)	
20	Deliveries - assisted	
21	Deliveries - caesarean section	
22	Deliveries - complications	
23	Deliveries – preterm	
24	Deliveries by HIV positive women	
25	Live birth by HIV positive women	
26	Deliveries amongst HIV positive women – Booked	

27	Deliveries amongst HIV positive women – Unbooked	
28	Deliveries monitored using a partograph	
29	Deliveries taken by a skilled birth attendant	

Tetanus Toxoid for Pregnant Women

Total

30	TT1	
31	TT2	
32	TT3	

NEW BORN HEALTH

Pregnancy Outcome - Live Births

		Male		Female		Total
33	**Live Births**	<2.5kg	>2.5kg	<2.5kg	>2.5kg	

Pregnancy Outcome - Still Births

Total

34	Still births	
35	Fresh still births (FSB)	
36	Abortions (Induced)	
37	Abortions (Total)	

Pregnancy Outcome - Complications **Male** **Female**

Total

38	Birth Asphyxia			
39	Neonatal sepsis			
40	Neonatal tetanus			
41	Neonatal jaundice			
42	Low birth weight babies placed in KMC			
43	Newborns with low birth weight discharged after KMC			

Immunization

		< 1 year		> 1 year		
	Antigen	**Fixed**	**Outreach**	**Fixed**	**Outreach**	**Total**

44	OPV 0 birth			☒	☒	
45	Hep. B 0 birth			☒	☒	
46	BCG			☒	☒	
47	OPV 1					
48	Hep. B 1					
49	Penta. 1					
50	DPT 1 (not when using Penta)					
51	PCV 1					
52	OPV 2					
53	Hep. B 2					
54	Penta. 2					
55	DPT 2 (not when using Penta)					
56	PCV 2					
57	OPV 3					
58	Penta. 3					
59	DPT 3 (not when using Penta)					
60	PCV 3					
61	Measles 1					
62	Fully Immunized < 1 year			☒	☒	
63	Yellow Fever					
64	Measles 2					
65	Conjugate A CSM					

Nutrition **Male** **Female**

Total

66	Children 0-59 months weighed – total			
67	Children 0-59 months weighing below the bottom line			
68	Children 0-6 months reporting being exclusively breast fed			
69	Children 6-11 months given Vitamin A			
70	Children 12-59 months given Vitamin A			
71	Children 12-59 months given deworming medication			
72	Children <5 years placed on treatment for severe acute malnutrition (OTP & SC)			
73	Children <5 years discharged (as healthy) from treatment for severe acute malnutrition (Recovered)			
74	Children admitted into CMAM Program			
75	Children defaulted from CMAM into intervention			

MALARIA PREVENTION (LLIN)

Total

76	Children under 5 years who received LLIN this month	

IMCI **Male** **Female**

Total

77	Diarrhoea new cases < 5 years			
78	Diarrhoea new cases < 5 years - given oral rehydration preparations (low osmolar ORS)			
79	Diarrhoea new cases < 5 years - given ORS and zinc			

	supplementation			
80	Pneumonia new cases < 5 years			
81	Pneumonia new cases < 5 years - given antibiotics (amoxyl PT)			
82	Measles new cases < 5 years			

Family Planning **Male** **Female**

Total

83	Clients counselled			
84	New family planning acceptors			
85	FP clients accessing HCT services			
86	Individual referred for FP services from HCT			
87	Individual referred for FP services from ART (ART Refill)			
88	Females aged 15 – 49 yrs using modern contraception			
89	Persons given oral pills			
90	Oral pill cycle (sachets) dispensed			
91	Injectables given			
92	IUCD inserted			
93	Implants inserted			
94	Sterilization			
95	Male Condoms distributed			
96	Female Condoms distributed			
97	Individual referred for FP services from PMTCT (HIV+ Pregnant Women)			

Referrals

Total

98	Referral in	
99	Referral out	
100	Malaria cases referred for further treatment	
101	Malaria cases referred for adverse drug reaction	
102	Women referred out for Pregnancy related complications	
103	Women seen and referred for Obstetric Fistula (VVF & RVF)	

Non-communicable diseases **Male** **Female**

Total

104	Coronary heart disease new cases			
105	Diabetes mellitus new cases			
106	Hypertension new cases			
107	Sickle cell disease new cases			
108	Road traffic accident new cases			
109	Home accident new cases			
110	Snake bites new cases			
111	Asthma new cases			
112	Athritis new cases			

Sexually transmitted infections

Total

113	STI treated new cases	

114	Male Urethritis new cases	

Laboratory

Total

115	ANC anaemia test done	
116	ANC anaemia test positive	
117	ANC proteinuria test done	
118	ANC proteinuria test positive	
119	HIV rapid antibody test done	
120	Sputum AFB - new diagnostic test done (incl. relapse)	
121	Sputum AFB - new diagnostic test done (incl. relapse) - tested positive	

Inpatient

Total

122	Functional beds	
123	Inpatient days – total	
124	Inpatient discharges – total	

Inpatient Admissions

125	**Total admissions**	**Male**						**Female**						**Total**
		0 - 28d	29d - 11 m	12 - 59 m	5 - 9 Yrs	10 - 19 Yrs	20 yrs+	0 - 28d	29d - 11 m	12 - 59 m	5 - 9 Yrs	10 - 19 Yrs	20 yrs+	

Pharmaceutical service

Total

126	Prescriptions issued	
127	Items dispensed	

Adverse Drug Reaction

128	Adverse drug reactions (ADRs) reported following immunization	
129	Adverse drug reactions (ADRs) reported following use of antiretrovirals	
130	Adverse drug reactions (ADRs) reported following use of antimalarials	

131	Antimalarials in the health facility with Mobile Authentification Service (scratch card)	With Scratch card	Without Scratch card

Mortality

		Male						**Female**						
132	**Total Deaths**	0 - 28d	29d - 11 m	12 - 59 m	5 - 9 Yrs	10 - 19 Yrs	20 yrs+	0 - 28d	29d - 11 m	12 - 59 m	5 - 9 Yrs	10 - 19 Yrs	20 yrs+	**Total**

Maternal Mortality

Total

133	Deaths of women related to pregnancy	

134	Causes of Maternal death	**Complications**									
		Antepartum haemorrhage	Pre/eclampsia	Post partum haemorrhage	sepsis	Obstructed labour	Abortion	Malaria	Anaemia	HIV	Unspecified

Neonatal Deaths

	Causes	**Complications**

135	of Neonatal death	Asphyxia	Sepsis	Prematurity	Neonatal Tetanus	Diarhoea	Congenital Malformation	Unspecified

HIV, TB, Malaria and Integrated Services

HIV Counselling & Testing			**Male**						**Female**					**Total**
		< 2 yrs	2-14 yrs	15 - 19 yrs	20 - 24 yrs	25 - 49 yrs	50 yrs+	< 2yrs	2-14 yrs	15 - 19 yrs	20 - 24 yrs	25 - 49 yrs	50 yrs+	
136	Individuals HIV counseled, tested and received results													
137	Individuals tested HIV positive													
138	Couples HIV counselled, tested &received results													
139	Couples HIV counselled, tested &received results that are sero-discordant													

HIV Care & Treatment		**Male**		**Female**		**Total**
		< 15 yrs	>= 15 yrs	< 15 yrs	>= 15 yrs	
140	HIV positive patients receiving cotrimoxazole prophylaxis					
141	ART patients receiving ARV refill					

SRH-HIV Integration **Male** **Female**

Total

142	HCT clients provided with SRH/HIV integrated services			

143	HCT clients referred for FP method			
144	HCT clients screened for STIs			
145	HCT clients treated for STIs			
146	Clients reached with SRH/HIV Integrated service			

TB/HIV **Male**

Female **Total**

147	Individuals clinically screened for TB			
148	Individuals clinically screened for TB score 1+ (TB suspects)			
149	Registered TB patients screened for HIV			
150	Individuals started on TB treatment - HIV -negative			
151	Individuals started on TB treatment - HIV Unknown			
152	HIV positive clients attending HIV care and treatment services and receiving TB treatment (count started on TB treatment)			
153	TB patients with HIV receiving ART			
154	Co-infected persons on CPT			

PMTCT - Mother

Total

155	ANC women with previously known HIV status (At ANC)	
156	Pregnant women who received HIV counseling, testing and received results at ANC	
157	Pregnant women who received HIV counseling, testing and received results at L&D	
158	Pregnant women who received HIV counseling, testing and received results at PNC	
159	Partners of HIV positive pregnant women tested HIV negative	
160	Partners of HIV positive pregnant women tested HIV positive	

161	Partners of HIV negative pregnant women tested HIV positive	
162	Partners of HIV negative pregnant women tested HIV negative	
163	HIV positive pregnant women assessed for ART eligibility by either clinical stage or CD4	
164	Pregnant HIV positive woman who received ARV prophylaxis for PMTCT(Triple)	
165	Pregnant HIV positive woman who received ARV prophylaxis for PMTCT(SdNVP in labour + (AZT +3TC))	
166	Pregnant HIV positive woman who received ARV prophylaxis for PMTCT(AZT)	
167	Pregnant HIV positive woman who received ARV prophylaxis for PMTCT(SdNVP in labour only)	
168	Pregnant HIV positive woman who received ARV prophylaxis for PMTCT	

PMTCT - Infant **Male**

Female Total

169	Infants born to HIV infected women started on cotrimoxazole prophylaxis within 2 months			
170	Infants born to HIV infected women started on cotrimoxazole prophylaxis 2 months & above			
171	Infants born to HIV-infected women who received an HIV test within two months of birth - (DNA-PCR)			
172	Infants born to HIV-infected women who received an HIV test after two months of birth - (DNA-PCR)			

173	Infants born to HIV-infected women who received an HIV test at 18months-(HIV Rapid test)			
174	Infant born to HIV-infected women who tested negative to HIV Rapid test at 18 months			
175	HIV exposed infants breast feeding and receiving ARV prophylaxis			

TB/LP **Male** **Female**

Total

		< 15yrs	≥ 15yrs	< 15yrs	≥ 15yrs	
176	TB cases (all forms) notified					
177	TB cases successfully treated among all forms (cured and completed)					
178	Registered TB patients tested for HIV					
179	DR-TB suspects tested for DR-TB					
180	Confirmed DR-TB patients enrolled for treatment					
181	Leprosy cases registered					
182	Buruli Ulcer patients notified					

Malaria Testing **MALE**

FEMALE **Total**

		< 5 years	**>= 5 years**	**< 5 years**	**>= 5 years**	
183	Persons with Fever					
184	Persons presenting with fever and tested by RDT					
185	Persons test positive for malaria by RDT					
186	Persons presenting with fever and tested by Microscopy (for malaria parasites)					

187	Persons tested positive for malaria by Microscopy					

Malaria in Pregnancy

Total

188	Pregnant women with clinically diagnosed Malaria	
189	Pregnant women with confirmed Malaria	

Malaria Cases **Male** **Female**

Total

		< 5 years	**>= 5 years**	**< 5 years**	**>= 5 years**	
190	Persons with clinically diagnosed Malaria					
191	Persons with confirmed uncomplicated Malaria					
192	Persons with severe Malaria					

Treatment with ACTs by type		**ACT 1**	**ACT 2**	**ACT 3**	**ACT 4**	**Total**
193	Children <5 years with confirmed uncomplicated Malaria receiving ACT			☒	☒	
194	Persons >=5 years with confirmed uncomplicated Malaria receiving ACT	☒				
195	Children < 5 years treated with ACT on the basis of clinical diagnosis only			☒	☒	
196	Persons >= 5 years treated with ACT on the basis clinical diagnosis only	☒				

Obstetric Fistula	**VVF**	**RVF**	**VVF & RVF**	**Total**

197	Women who reported leaking urine or feces	☒	☒	☒	
198	Women receiving surgery for fistula repair				
199	Women receiving a first repair				
200	Women receiving a second repair				
201	Women discharged after fistula surgery				
202	Women who had a closed and dry fistula at discharge				

Commodity Availability **Yes**

No

203	Stock out of any essential drug for 7 days consecutively in the last one month?		
204	Stock out of vaccine supplies in the past one month?		
205	Stock out of family planning commodities in the past one month?		
206	Stock out of Antiretroviral Drugs in the past one month?		
207	Stock out of ACTs for 7 days consecutively in the past one month?		
208	Stock out of RDTs for 7 days consecutively in the past one month?		
209	Stock out of SPs for 7 days consecutively in the past one month?		
210	Stock out of LLINs for 7 days consecutively in the past one month?		
211	Stock out of HIV test kits for 7 days consecutively in the past one month?		
212	Stock out of anti-TB drugs for 7 days consecutively in the past one month?		
213	Stock out of CPT for 7 days consecutively in the past one month?		
214	Stock out of INH for 7 days consecutively in the past one month?		

215	Stock out of Female Condoms in the past one month?		
216	Stock out of Implants in the past one month?		
217	Stock out of Emergency Contraception in the past one month?		
218	Stock out of Oxytocin in the past one month?		
219	Stock out of Misoprostol in the past one month?		
220	Stock out of Magnesium sulfate in the past one month?		
221	Stock out of Injectable antibiotics in the past one month?		
222	Stock out of Antenatal Corticosteroid (ANCS) in the past one month?		
223	Stock out of Chlorhexidine in the past one month?		
224	Stock out of Resuscitation Equipment in the past one month?		
225	Stock out of Amoxicillin DT in the past one month?		
226	Did you carry out planned preventive maintenance (PPM) of your equipment?		

Appendix 3 Completeness and Accuracy of Data Elements

	Data Source	Daily Register		Montly Summary Form		DHIS Database	
		Available (Yes/No/N/A(Not Applicable))	**Value**	**Available (Yes/No)**	**Value**	**Available (Yes/No)**	**Value**
1	**Dataelement1**						
2	**Dataelement2**						
3	**Dataelement3**						
4	**Dataelement4**						
5	**Dataelement5**						
6	**Dataelement6**						
7	**Dataelement7**						
8	**Dataelement8**						
9	**Dataelement9**						
10	**Dataelement10**						
11	**Dataelement11**						
12	**Dataelement12**						
13	**Dataelement13**						

14	**Dataelement14**						
15	**Dataelement15**						
16	**Dataelement16**						
17	**Dataelement17**						
18	**Dataelement18**						
19	**Dataelement19**						

Appendix 4

Health Facility Data Manager Interview Guide

1. **Facility Code**: ____________________________________

2. **Facility Type**: Primary ☐ Secondary ☐ Tertiary ☐

3. **Facility Location**: Urban ☐ Rural ☐

4. **Sector**: Public ☐ Private ☐

Questions

1. What is your specific role in this health facility?
2. What is your educational background?
3. Can you please describe the routine process of collating data in your health facility per day, per week and per month?
4. Are there any challenges you encounter in data management in your health facility? Please describe.
5. Do you work with anyone to ensure data quality in the health facility? Please describe.
6. Do you attend trainings on data management and specifically data quality? How often in a year?
7. What is your understanding of data quality?
8. Do you use any data quality mechanisms? Please describe.
9. How do you ensure data quality in your health facility?
10. What are the challenges you encounter ensuring data quality in your facility?

Appendix 5

CONSENT FORM

Title of Research Project: ***Assessment of Data Quality in Routine Health Information Systems in Oyo State, Nigeria.***

The study has been described to me in language that I understand. My questions about the study have been answered. I understand what my participation will involve and I agree to participate of my own choice and free will. I understand that my identity will not be disclosed to anyone. I understand that I may withdraw from the study at any time without giving a reason and without fear of negative consequences or loss of benefits.

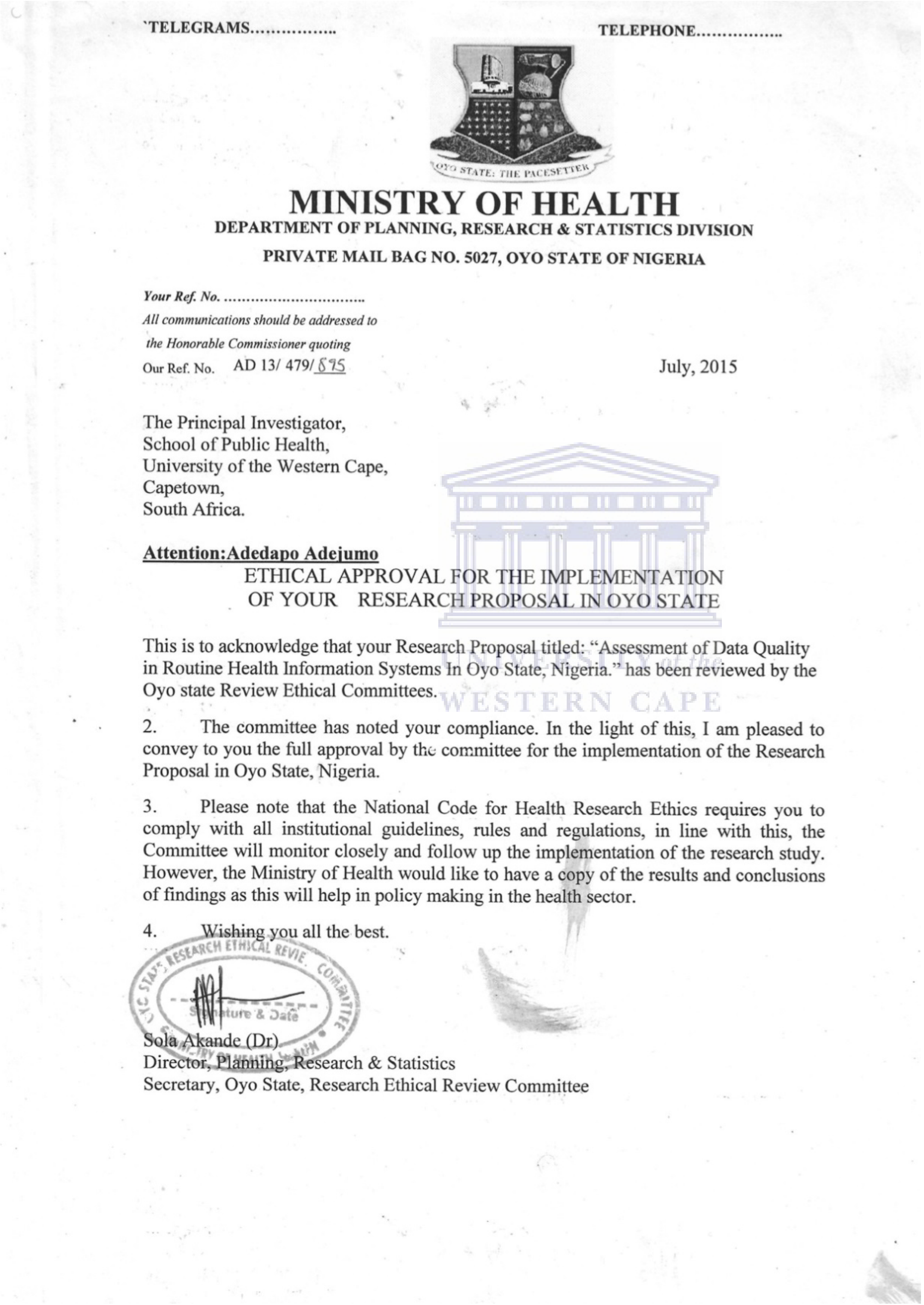
Participant's name………………………..

Participant's signature……………………………….

Date………………………

Appendix 6

Ethical Approval



`TELEGRAMS................ TELEPHONE................

MINISTRY OF HEALTH
DEPARTMENT OF PLANNING, RESEARCH & STATISTICS DIVISION
PRIVATE MAIL BAG NO. 5027, OYO STATE OF NIGERIA

Your Ref. No. *................................*
All communications should be addressed to
the Honorable Commissioner quoting
Our Ref. No. AD 13/ 479/875

July, 2015

The Principal Investigator,
School of Public Health,
University of the Western Cape,
Capetown,
South Africa.

Attention:Adedapo Adejumo

ETHICAL APPROVAL FOR THE IMPLEMENTATION OF YOUR RESEARCH PROPOSAL IN OYO STATE

This is to acknowledge that your Research Proposal titled: "Assessment of Data Quality in Routine Health Information Systems In Oyo State, Nigeria." has been reviewed by the Oyo state Review Ethical Committees.

2. The committee has noted your compliance. In the light of this, I am pleased to convey to you the full approval by the committee for the implementation of the Research Proposal in Oyo State, Nigeria.

3. Please note that the National Code for Health Research Ethics requires you to comply with all institutional guidelines, rules and regulations, in line with this, the Committee will monitor closely and follow up the implementation of the research study. However, the Ministry of Health would like to have a copy of the results and conclusions of findings as this will help in policy making in the health sector.

4. Wishing you all the best.

Sola Akande (Dr)
Director, Planning, Research & Statistics
Secretary, Oyo State, Research Ethical Review Committee