SCHOOL OF PUBLIC HEALTH
COLLEGE OF HEALTH SCIENCES
UNIVERSITY OF GHANA

ACCURACY AND COMPLETENESS OF HYPERTENSION DATA IN THE DHIMS-2
IN SELECTED HEALTH FACILITIES IN THE GREATER ACCRA REGION

BY
EMMANUEL TETTEH DOKU
(10636516)

THIS DISSERTATION IS SUBMITTED TO THE UNIVERSITY OF GHANA, LEGON
IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE AWARD OF
MASTER OF PUBLIC HEALTH DEGREE

JULY, 2018
DECLARATION

I, EMMANUEL TETTEH DOKU, hereby declare that with the exception of references to other people’s work which has been duly cited, this thesis was done entirely by me under due supervision. I further affirm that this thesis has never been presented either in whole or in part for any other degree in this University or elsewhere.

.......................................................... ..........................................................  
EMMANUEL TETTEH DOKU DATE  
(STUDENT)  

.......................................................... ..........................................................  
DR. ANTHONY DANSO-APPIAH DATE  
(SUPERVISOR)
ABSTRACT

Background: Hypertension is a major public health challenge in Ghana and with the prevalence increasing over the years, accurate estimation of the burden of the disease is required for effective prevention and control measures. The GHS in 2012, introduced a web-based database; DHIMS-2, to ensure quality health data however, aspects of this database; the maternal and neonatal data was reported to be good and reliable, yet the EPI data was poor, aside the lack of evidence based research on the morbidity data. The aim of this study therefore was to determine the quality of Outpatient Hypertension Data in the DHIMS-2 by assessing how accurate and complete hypertension data reported into the database is, and factors influencing hypertension data management in health facilities.

Methodology: Through a mixed method approach using secondary data from patient’s folders for the third quarter of 2016, two hospitals and two polyclinics were randomly selected out of highly ranked administrative units in the Greater Accra region. A structured Data Abstraction form and a semi-structured interview guide were used to extract information in relation to the study objectives. Hypertension data were assessed for accuracy (error rate) and completeness (% missing data). Primary source data from patient’s folders and consulting room registers were totalled, gathered and comparison was made with facility aggregate forms as well as the DHIMS-2. Twelve health staff were interviewed and through thematic analysis, factors influencing hypertension data quality were assessed.

Results: A total of 2268 of data from consulting room registers and patient folders were inspected. The mean completeness of hypertension data among the four health facilities was 60.69% (95%
Comparing new cases of hypertension in consulting room registers to patient folders, the mean percentage error was 72.9% (95% C.I=52.20% - 93.62%). The study found an overall of 21.5% (95% C.I= 4.49% – 38.48%) error rate in the transfer of hypertension data from consulting room registers to the monthly OPD morbidity form. There was 100% accuracy in transferring hypertension data from facility monthly aggregate form to the DHIMS-2, however, in comparing the DHIMS-2 data to new cases of hypertension captured in patient folders, the mean error detected was 79.2% (95% C.I= 49.02% - 109.42%). Factors contributing to incomplete and inaccurate hypertension data included, lack of understanding on indicators, less priority on hypertension data quality and lack of validation activities focusing on hypertension.

**Conclusion:** Over 50% of data captured into the consulting room registers were accounted for in the web-based database, but the accuracy of hypertension data in the DHIMS-2 did not reflect what existed in primary sources and this disparity was wide. Factors affecting hypertension data quality in hospitals and polyclinics were related to inadequate data verification, poor data management and inadequate training of staff on data management, therefore there is the need to regularly validate hypertension data and consistently train staff directly linked to managing hypertension data.
ACKNOWLEDGEMENT

I give thanks to the Almighty God for his love, care and protection throughout my stay on Campus.

I wish to express my heartfelt gratitude to my supervisor, Dr. Anthony Danso-Appiah for his priceless suggestions, guidance, counselling and constructive criticisms, without which this dissertation would not have become a reality.

I am so much grateful to all the health professionals I interacted with at the four health facilities for their immense assistance and contributions during the phases of my data collection. My deepest gratitude goes to my parents, Madam Emma Ayebea Otchere and Mr. Joseph Doku for your tremendous support and encouragement. God richly bless you abundantly and replenish all that you have lost in many folds. And to my siblings, Carl Ortis Doku, Harriet Narki Doku and Charles Tetteh Doku God bless you for the individual and collective roles you played towards making this thesis a reality. Special thanks also go to Emmanuel Aidoo (Statistics Department) for your statistical support. For all others whose names may not have been captured here but have contributed immensely towards the completion of this study, your contributions are very much appreciated. I, however wish to state that any shortcomings in this work remain solely my responsibility.
DEDICATION

This thesis is solely dedicated to my lovely learned wife who immeasurably provided for me in all aspects of life to carry out this research. She is the chief editor of this work, hence the quality of this work is greatly attributed to her.
# TABLE OF CONTENTS

DECLARATION ............................................................................................................................. i  
ABSTRACT ................................................................................................................................. ii  
ACKNOWLEDGEMENT .............................................................................................................. iv  
DEDICATION ................................................................................................................................ v  
TABLE OF CONTENTS ............................................................................................................... vi  
LIST OF TABLES ......................................................................................................................... ix  
LIST OF FIGURES ....................................................................................................................... x  
LIST OF ACRONYMS ................................................................................................................. xi  
CHAPTER ONE ............................................................................................................................. 1  
INTRODUCTION .......................................................................................................................... 1  
  1.1 Background to the Study ....................................................................................................... 1  
  1.2 Problem Statement ................................................................................................................ 7  
  1.3 Conceptual Framework ......................................................................................................... 9  
  1.3.1 Flow Chart of hypertension data capture ..................................................................... 10  
  1.4 Justification ......................................................................................................................... 11  
  1.5 Research Questions ............................................................................................................. 12  
  1.6 Objectives ............................................................................................................................ 13  
CHAPTER TWO .......................................................................................................................... 14  
LITERATURE REVIEW ............................................................................................................. 14  
  2.1 Introduction ......................................................................................................................... 14  
  2.2 Data Quality ........................................................................................................................ 14  
  2.3 Data Quality Dimensions .................................................................................................... 15  
  2.3.1 Completeness ................................................................................................................ 16  
  2.3.2 Accuracy ....................................................................................................................... 16  
  2.3.3 Accuracy of Hypertension data .................................................................................... 17  
  2.3.4 Timeliness ..................................................................................................................... 18  
  2.3.5 Consistency ................................................................................................................... 19  
  2.3.6 Validity ......................................................................................................................... 19  
  2.4 Measuring data quality ........................................................................................................ 19  
  2.5 Importance of Data Quality ................................................................................................. 20
APPENDICES .............................................................................................................................. 65
APPENDIX 1 ............................................................................................................................ 65
APPENDIX 2 ............................................................................................................................ 66
APPENDIX 3 ............................................................................................................................ 67
APPENDIX 4 ............................................................................................................................ 68
LIST OF TABLES

Table 1: Distribution of percentage completeness and error rates of hypertension data on sex and age groups in four health facilities in Greater Accra Region........................................................ 37

Table 2: Percentage completeness and percentage of missing hypertension data in four health facilities in the Greater Accra region during the third quarter of 2016 ...................................................... 38

Table 3: Percentage Error of hypertension data in four health facilities in the Greater Accra Region during the third quarter of 2016 .......................................................... 39
LIST OF FIGURES

Figure 1: A chart showing Determinants of Accuracy and Completeness of Hypertension data. 9
Figure 2: A chart showing the flow of hypertension data at the facility level. 10
Figure 3: Map showing the selected health facilities in the Greater Accra Region. 28
Figure 4: Distribution of true and deviated cases of hypertension among the health facilities. 42
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIDS</td>
<td>Acquired Immune-deficiency Syndrome</td>
</tr>
<tr>
<td>CHIM</td>
<td>Centre for Health Information Management</td>
</tr>
<tr>
<td>CRF</td>
<td>Case Report Form</td>
</tr>
<tr>
<td>CRR</td>
<td>Consulting Room Register</td>
</tr>
<tr>
<td>DHIS</td>
<td>District Health Information System</td>
</tr>
<tr>
<td>DHIMS</td>
<td>District Health Information Management System</td>
</tr>
<tr>
<td>DPT</td>
<td>Diphtheria, Pertussis and Tetanus</td>
</tr>
<tr>
<td>EDC</td>
<td>Electronic Data Collection</td>
</tr>
<tr>
<td>EPI</td>
<td>Expanded Program on Immunization</td>
</tr>
<tr>
<td>FMOH</td>
<td>Federal Ministry of Health</td>
</tr>
<tr>
<td>GHS</td>
<td>Ghana Health Service</td>
</tr>
<tr>
<td>GCDMP</td>
<td>Good Clinical Data Management Practise</td>
</tr>
<tr>
<td>HIS</td>
<td>Health Information System</td>
</tr>
<tr>
<td>HIV</td>
<td>Human Immune-deficiency Virus</td>
</tr>
<tr>
<td>HMIS</td>
<td>Health Management Information System</td>
</tr>
<tr>
<td>HMN</td>
<td>Health Management Network</td>
</tr>
<tr>
<td>HPT</td>
<td>Hypertension</td>
</tr>
<tr>
<td>HSE</td>
<td>Health Survey for England</td>
</tr>
<tr>
<td>ID</td>
<td>Identity</td>
</tr>
<tr>
<td>IT</td>
<td>Information Technology</td>
</tr>
<tr>
<td>M &amp; E</td>
<td>Monitoring and Evaluation</td>
</tr>
<tr>
<td>MOH</td>
<td>Ministry Of Health</td>
</tr>
<tr>
<td>NHS</td>
<td>National Health Service</td>
</tr>
<tr>
<td>NMCP</td>
<td>National Malaria Control Program</td>
</tr>
<tr>
<td>OPD</td>
<td>Out-Patient Department</td>
</tr>
<tr>
<td>PPME</td>
<td>Policy Planning Monitoring and Evaluation</td>
</tr>
<tr>
<td>PDC</td>
<td>Paper-based Data Collection</td>
</tr>
<tr>
<td>PHC</td>
<td>Primary Health Care</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>PHDS</td>
<td>Population Health Data Sets</td>
</tr>
<tr>
<td>SOP</td>
<td>Standard Operating Procedure</td>
</tr>
<tr>
<td>TB</td>
<td>Tuberculosis</td>
</tr>
<tr>
<td>THIN</td>
<td>The Health Information Network database</td>
</tr>
<tr>
<td>UK</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organization</td>
</tr>
</tbody>
</table>
CHAPTER ONE

INTRODUCTION

1.1 Background to the Study

Hypertension, (HPT) also known as high blood pressure is a major universal public-health challenge as a result of an increase in its incidence (Go et al., 2013). It is projected that the incidence of hypertension in the world is about one billion (Kottke, Stroebel, & Hoffman, 2003). The worldwide burden is higher in low middle income countries where about 1 in every 5 of the adult population is affected (Seedat, 2000), and this is even projected to increase to 3 out of every 4 people by 2025 (Opie & Seedat, 2005). Estimates now show that in Africa more than 40 percent of adults have hypertension (Addo, Smeeth, & Leon, 2007) and nearly 80 million adults with hypertension in sub-Saharan Africa in 2000 and projections built on epidemiological data submit that this will increase to 150 million by 2025 (Opie & Seedat, 2005). Furthermore, in the developing countries, the incidence of hypertension is growing and is anticipated to increase by more than 500 million by 2025 (Kearney et al., 2005; Fuentes et al., 2000).

Judicious management of the health sectors requires that estimations of the prevalence of hypertension are accurate to serve as a primary data for these health sectors (Kearney et al., 2005). It is in recent times that regular epidemiological data on hypertension have been reported from countries that are developing like Ghana. However, according to Hoang et al., 2007; Damasceno et al., 2009; Demaio et al., 2013, for the past 20 years, the data is still not adequate to evidently describe population based trends in the analyses of blood pressure.
There has been a substantial increase in the prevalence of hypertension in Ghana. The prevalence in both the urban and rural areas was reported to be 11.3% and 4.5% respectively in 1977 as posited by (Agyemang et al., 2012). In 2004, a study of the greater Accra area found an urban prevalence of hypertension to be 32.9% and a rural prevalence also to be 24.1% (Cappuccio et al., 2004). This drastic increase in the prevalence of HPT indicates a need for further review of the condition in Ghana. The scarcity of data on hypertension and its related risk factors in Ghana indicate that further investigation is merited.

Data accuracy is the rate at which data gathered is exact, and precisely denotes what it defines (Nektar Data Systems, 2016). It is described as “the closeness of data values to the truth or the veracity of the information received” (Chen et al., 2014). In other words, it can be defined as the extent to which data appropriately defines the "real world" item being defined (Dama et al., 2013). 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). Accuracy of a morbidity data is very critical in estimating the prevalence and incidence of the disease among a particular group of people. For instance if it was estimated in 2012 that the incidence of hypertension in a rural area in Kumasi was 35% (Hana et al., 2012), then it should mean that real data captured in health facilities on hypertension should reflect exactly same statistics. The lack of data on trends on the prevalence of hypertension in Ghana highlights the need for attaining more data using strict measures in order to help inform policy and practice on hypertension and cardiovascular diseases (Agyemang et al., 2012).

Data completeness is explained to be the rate at which data comprises the items or points that are needed to back the application it is intended to (Dekkers et al., 2014). Simply put, it tells if the
data has any loopholes looking at what was projected to be gathered, and what was really gathered. According to Legendre & Legendre, 2012; it is defined as “a measure of the presence of expected data items in a given dataset or collection”. Data that needs to be submitted must be checked to ensure that all estimated data is present to avoid issues with incomplete data. This is extremely challenging with the use of paper as it is susceptible to mistakes (Nektar Data Systems, 2016). However, data collection procedures which are paperless involves the use of smartphones and tablets which allows companies and the health sector to gather similar data that involves the use of paper, but data is documented on a secondary device other than paper (Hall, 2011).

The Ghana Health Service (GHS) implemented an innovation which is an online database known as District Health Information Management Systems – 2 (DHIMS-2). The purpose of this innovation was to enhance the DHIMS data quality by reducing processes used in acquiring data. Formerly, in managing the routine data collected by the health facilities, the DHIMS was used. With that, health services had to organize and forward their data to the districts. These districts also had to organize and forward the data to the regional level and finally to the national level before sending to the national data set. There have been tremendous improvements in data collection, reporting and analysis with the invention of the DHIMS-2 which is deployed nationwide and has reinforced the country’s health systems. This new method, however, still depends on paper-based data collection (PDC) but the length of processes involved in acquiring data has drastically reduced signifying possible development in the quality of the DHIMS-2 dataset likened to the DHIMS dataset that was formerly used. Despite all the benefits that can be derived from the DHIMS-2, there are some shortfalls that have been identified in its utilization. The issues have basically been grouped into three categories: organizational, technical and behavioral. The
current data management system also permits health centres to organise and send information straight to the DHIMS-2 dataset at all the three levels. This is a tremendous accomplishment comparable to the previous ways of managing data. However, according to (Kayode et al., 2014), in spite of all the benefits aligned with this upgraded way of managing data, the DHIMS-2 data quality has still not been duly expressed.

The quality of data with regards to routine health services delivery in the DHIMS-2 has been tested through several studies using accuracy and completeness as the yard stick. To measure the quality of a data file, its accuracy and completeness are estimated (Fong, 2001). Preferably, Nahm et al., (2008) posits that in data validation, there should be a comparison of the database to the source of the data that is, patient’s folder (primary data source) so that the error rate will not be underestimated. The quality of Outpatient Morbidity data in the DHIMS-2, which forms one of the pillars in determining the burden of diseases among geographical areas, however, has not been tested in detailed.

Health facilities in Ghana can be categorized into Public, Private, Quasi-government and Mission health facilities. For most of the privately owned facilities, the quasi-government and the Mission health institutions, especially those based in the urban parts of the country, health information data management is strictly based on Electronic Data Collection and later there is a transfer of data into the DHIMS-2 database for nationwide access. However, most of the Public Health Facilities resort to the Paper-Based Data Collection, and since Public Health Facilities relatively outnumber the Private facilities, very sizeable amount of the DHIMS-2 database emanates from the Public Health Institutions. In the Ghana Health Service, there is always a production of routine data at the facility
level and accumulated once-a-month at the facility and district levels and transferred to the regional and national levels (Amoakoh-Coleman et al., 2015).

The Paper-based Data Collection (PDC) and the Electronic Data Capturing (EDC) are ways of collecting data. These methods are liable to mistakes thus, it is very necessary to be cautious when assessing the quality of data which precedes analysis. Mistakes in clinical data can be identified by dual data entry, identifying outliers, relational encounters and pictorial verification (Fong, 2001). In developing countries like Ghana, routine clinical data are typically paper-based before they are uploaded to the database (Kayode et al., 2014). In Paper-based Data Collection, in most situations mistakes are likely to occur whiles transferring and processing data. This in turn results in data being inaccurate or incomplete. The amount of mistakes in data management are as a result of inadequate resources for organising data, unclear indicators, absence of training, lack of manpower, poor data structure, sub-standard operating systems etc. (Fong, 2001). Knowledge of the procedures employed in compiling data also aids in recognizing the areas where mistakes are likely to happen (Amoakoh-Coleman et al., 2015).

According to the Standard Operating Procedure (SOP) for Health Information, the consulting room register is the standardized GHS document that captures almost all morbidities at the OPD of health facilities. In capturing accurate data into the consulting room register, the user must be well trained and understand the various indicators in the register before the right data can be captured. Deviation from the SOP affects data quality and integrity. Practically, after a patient’s diagnosis is stated in the patient’s folder, either that same clinician or another clinician, in most cases a nurse, transfers the information into the consulting room register. In the case of hypertension; being a
chronic disease, if a patient is seen and diagnosed for the first time in his/her lifetime by a Physician, per the definition of the SOPs, that particular diagnosis must be captured as a new case of hypertension, however, if that patient has ever been diagnosed of hypertension and has initiated treatment to manage the condition, the diagnosis in this instance must be captured as an old case in the consulting room register. It is the new cases that are counted and reported on the Monthly Outpatient morbidity form (facility aggregate form) as hypertension whilst the old case is only captured as re-attendance on that same form. Furthermore, that hypertension case reported into the facility aggregate form is later transferred into the DHIMS-2 database and made available in real time at all the functional levels of Ghana Health Service. This data is what is being populated and reported periodically by all stakeholders during performance reviews. Also that data is mostly used by researchers and Epidemiologists in calculating for the incidence and prevalence rate of hypertension. However, assuming an old case of hypertension is wrongly captured as a new case, it will give a wrong picture of the burden of the disease condition.

Data extracted from the DHIMS-2 indicates that, hypertension cases has for the past five years, since the inception of the DHIMS-2 data base in 2012, been reducing substantially in all the Ten Regions of Ghana. Among all the Regions, the Greater Accra Region has recorded the highest cases of hypertension so far. As a matter of fact, hypertension still ranks among the top causes of Outpatient morbidity and mortality in most of the regions. For instance in the Greater Accra region alone, hypertension is the only non-communicable disease observed to have positioned itself comfortably among the first five topmost causes of Outpatient morbidity. Over the years this has raised many questions among health managers and stakeholders at various levels of the health sector as to whether these recorded hypertension cases are really new cases of the disease being
diagnosed, or perhaps they are the same old cases of the disease which are being captured wrongly and hence projecting such alarming rates of hypertension in the city.

Since data management is one of the key pillars in Epidemiology, it is therefore very necessary to determine if the hypertension cases recorded in the DHIMS-2 database is reliable and can be used for interventional programs, monitoring, evaluation and decision making. It is therefore in this light that this study was conducted to quantify hypertension data quality in the DHIMS-2 within selected health facilities in Accra.

1.2 Problem Statement

Data Quality has been found to be very beneficial in many organisations and the health sector is no exception. Empirical evidence suggests that several studies have been done in both developing and developed countries on the completeness and accuracy of neonatal data, Expanded Program on Immunization (EPI) and maternal health data. Review of literature revealed that the completeness and accuracy of the DHIMS-2 dataset for neonatal and maternal health was relatively good (Kayode et al., 2014), whereas that of EPI was poor (Informat et al., 2015). There is a dearth of research based evidence on the morbidity aspect of data in the DHIMS-2 in Ghana. In addition, the mixed results and errors that the DHIMS-2 has, since it still relies on Paper-based Data Collection, have made the accuracy and completeness questionable.

The lack of data on trends on the prevalence of hypertension in Ghana makes it necessary to attain more data using strict measures to help inform policy and management practices on hypertension and cardiovascular diseases (Agyemang et al., 2012). Also, most studies done in West Africa and Ghana to determine the incidence and burden of hypertension in a population has been community-
based studies and systematic reviews (Addo et al., 2006; Bosu, 2010). However, it is possible to use administrative data to accurately identify from a population sample those patients who have been diagnosed with hypertension. As long as administrative data are already routinely collected, their use is expected to be considerably less expensive equated with serial cross-sectional or cohort studies for surveillance of hypertension occurrence and outcomes over time in a large population (Tu et al., 2007). This study therefore sought to assess the accuracy and completeness of hypertension data in DHIMS-2 in selected health facilities within Greater Accra Region.
1.3 Conceptual Framework

National/Regional level
- MOH
- GHS
- PPME
- CHIM

Facility level Determinants
- Validation activities
- Use of data for decision making
- Supervision
- Data storage systems
- Data collection tools
- I.T. & Software’s

Personnel determinants
- Cadre of staff
- Years worked
- Understanding indicators
- Complexity of task
- Work load
- Motivation
- DHIMS competency

Outcome

Decision making points

INPUT

ACCURACY & COMPLETENESS

FEEDBACK SYSTEMS

Figure 1: A chart showing Determinants of Accuracy and Completeness of Hypertension data. Adapted from: Integrated Health Information Architecture - Power to the Users - Design, Development and Use. Braa & Sahay, 2012
1.3.1 Flow Chart of hypertension data capture.

**PATIENT IS DIAGNOSED OF HYPERTENSION**
Diagnosis is captured into Patients Folder at OPD

**CONSULTING ROOM REGISTER**
Diagnosis is captured and classified into new case or old case

**MONTHLY DATA COLLATION INTO FACILITY AGGREGATE FORM**
Only the New cases are captured as Hypertension in the form

**MONTHLY DATA ENTRY INTO THE DHIMS 2 AT THE FACILITY LEVEL OR DISTRICT LEVEL**
This captured data is always available at the regional and national level in real time

---

Data Entry, executed by H.I.O., B.A.

Data Collation, executed by H.I.O., B.A., Nurse

Data Capture, executed by M.O., PA, or Nurse

---

Figure 2: A chart showing the flow of hypertension data at the facility level.
Source: (Adapted from Kayode, 2014)
1.4 Justification

Globally, hypertension has become a general health problem and epidemiological data on hypertension in Ghana is required to inform policy and develop operative interventions. Over the past 30 years several studies have been done in both urban and rural settings to determine the burden of the disease in the country and there seem to be varying results. For instance in 2003 and 2004, studies were done in Accra and Kumasi respectively and revealed prevalence of 28.3% and 28.7% respectively (Amoah, 2003; Cappuccio et al., 2004).

In 2006, Addo, Amoah, and Koram also published a study that put the prevalence of hypertension around rural communities in Accra to be 25.4%. Again in 2008, a similar study on Civil Servants in Accra put the prevalence of hypertension at 30.3% according to Addo Smeeth, and Leon. A systematic review on 17 published studies (1970 – 2009) done on the prevalence of hypertension in Ghana estimated the prevalence to be 25% in Urban and 20% in Rural populations respectively (Bosu, 2010). Published reports and surveys conducted on hypertension in the year 1973 and 2009 also ranged the prevalence from 19.3% in rural to 54.6% in urban areas (Agyemang et al., 2012). In that same period, (Hana et al., 2012) estimated the prevalence of hypertension in a rural community in Kumasi to be 35%.

The varying statistics on the burden of hypertension even between robust studies done in recent times provides the need for further studies into the existing data available. The advent of the DHIMS-2 database notwithstanding its challenges is expected to provide reliable data on various disease conditions including hypertension to ascertain the risk and burden of the disease. Hence, the need for a study to quantify the accuracy and completeness of Hypertension data in the DHIMS-2. This study might therefore furnish health managers in Ghana, International Partners in Health and Researchers all over the world to have a fair idea about the
reliability of hypertension data in the DHIMS-2 and therefore appropriately estimate the burden of the disease with relative ease. It will also be a guide for similar studies to be conducted in other regions of the country and even within Africa. The results of this study will also guide various stakeholders e.g. GHS-PPME on where to empower human resource in data capturing activities within health facilities. Again just as TB and HIV/AIDS patients are being electronically tracked on the same DHIMS-2 platform to know exactly the burden of the disease in Ghana, same can be done for Hypertension.

Since the DHIMS-2 has a nationwide coverage and almost every geographical location and population base with their prevailing disease occurrence is captured into the system, it will be prudent and easy to compute and ascertain the incidence and prevalence of hypertension with DHIMS-2 as the source data. Hence, the need to test the accuracy and completeness of the hypertension data in the database.

1.5 Research Questions

MAIN QUESTION

- What is the quality of Outpatient Hypertension Data in DHIMS-2 in the Greater Accra Region?

SPECIFIC QUESTIONS

1. What is the accuracy of the Outpatient hypertension data in the selected facilities?
2. What is the completeness of the Outpatient Hypertension data in the selected facilities?
3. What factors contribute to incompleteness and inaccuracy of Hypertension data in the selected facilities?
1.6 Objectives

MAIN OBJECTIVE

• To assess the quality of Outpatient Hypertension Data in DHIMS-2 in selected health facilities in the Greater Accra Region.

SPECIFIC OBJECTIVES

1. To assess the accuracy of Outpatient Hypertension Data in the DHIMS-2.

2. To determine the completeness level of Outpatient Hypertension Data in the DHIMS-2.

3. To identify factors that contribute to inaccuracy and incompleteness of hypertension data in the selected health facilities.
CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

This chapter climaxes the concept of data quality in routine health information systems. The thematic areas covered in this chapter are the data quality, dimensions of data quality, the measurement of data quality, as well as factors that influence data quality in routine health information systems.

2.2 Data Quality

Quality in relation to data has numerous meanings but the widely and commonly used definition which is standardized is “fitness for use” or “potential use” (Devillers et al., 2007). Several researchers have studied issues related to quality, particularly for the quality of products, an instance is, quality is “the degree to which a set of inherent characteristics fulfill the requirements” or “conformance to requirements” (Cai & Zhu, 2015).

Data quality is generally looked at in three perspectives: organizational, architectural, and computational. These approaches must be looked at for an effective data quality task. The study of data quality came about as a result of a substantial growth in information technology. Studies on data quality commenced in the developed countries around 1990s where several researchers suggested varied ways of defining data quality (Sadiq, 2003).

Several studies sort to defined data quality as “fitness for use” (Fisher & Kingma, 2001; Devillers et al., 2007) and suggested that the quality of data can be best judged by those who utilize the data.
Redman (2001) also defined data quality as the extent to which data meet the specific requirements of specific customers. What one customer may find to be high quality data would be of low quality to another, Redman (2001) and Montgomery (2007) asserted that data is regarded as high quality "if they are fit for their intended uses in operations, decision making and planning." They added that high quality data are exact, accessible, complete, conformant, reliable, credible, relevant and timely.

### 2.3 Data Quality Dimensions

This is a recognized concept employed by data specialists to explain a characteristic of data that can be evaluated against defined standards so as to ascertain data quality (Stvilia et al., 2007; Dama et al., 2013). According to Weiskopf & Weng (2013) it is described as a context for measurement along with proposed measurement units. The mostly measured dimensions comprise completeness, consistency, timeliness, and uniqueness, though the ability to provide a method for measurement is restricted to the possible range of dimensions (Weiskopf & Weng, 2013; Sattler, 2009; Batini & Scannapieco, 2016). During this process, the dimensions to be measured are being selected by the data quality analyst who considers the tools, methods, and expertise required for the measurements. This process results in a collection of particular measures joined to add to data quality.

Recently, researchers suggest 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; Zozus et al., 2014). However, though there has not been any agreement on the dimensions of data quality there are cross cutting dimensions identified by the literature: Completeness, Uniqueness, Timeliness, Validity, Accuracy and Consistency (Batini &
Scannapieco, 2016; Glèlè Ahanhanzo et al., 2014; Ndabarora et al., 2014; Weiskopf & Weng, 2013; Sattler, 2009; Pipino, Lee, & Wang, 2002).

2.3.1 Completeness

The completeness of data can be defined as the rate at which the data comprises items essential to support the application it was anticipated for. Simply put, it determines if the data has loopholes from what was projected to be collected, and what was truly collected (Nektar Data Systems, 2016). According to Pipino and Wang, 2002 it is defined as “a measure of the presence of expected data items in a given dataset or collection”. In order for data to be submitted, issues with incomplete data must be fixed, on conditions that all estimated data is present. It is extremely challenging with data that uses paper since it is susceptible to human mistakes (Nektar Data Systems, 2016). On the contrary, smart phones and tablets serve as storage devices for gathering the same information as the paper forms (Brunette, 2013).

2.3.2 Accuracy

This determines if the data is exact, and accurately denotes what it should (Nektar Data Systems, 2016). It is described as “the closeness of data values to the truth or the veracity of the information received” (Chen et al., 2014). In other words, it can be defined as the rate at which data appropriately describes the "real world" object described (Dama et al., 2013). 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). Accuracy of a morbidity data is very critical in estimating the prevalence and incidence of the disease among a particular group of people. For instance, if it was estimated in 2012 that the incidence of hypertension in a village in Kumasi was 35% (Hana et al., 2012), then it should mean that real data captured in health facilities on hypertension should
reflect exactly same statistics. A review of population based prevalence of hypertension showed that lack of data trends of hypertension emphasises the need to acquire more data using rigorous means to inform policy on hypertension and cardiac disease in Ghana (Agyemang et al., 2012).

2.3.3 Accuracy of Hypertension data

A study was conducted to assess the accuracy of the pregnancy hypertensive disorders reporting in single and linked population health data set. A comparison was made with hypertension reporting in a sample of records from two data collections and information abstracted from the corresponding medical records, the datasets (birth and hospital discharge) compared with data abstracted from medical records. A data of 1200 women who were selected using random sampling was compiled from their folders by three experienced health workers. The compiled data referred to as ‘validation data’ were uploaded into an electronic dataset and combined with data from the linked birth and hospital population health datasets (PHDS) for analysis, (Roberts et al., 2008).

Another survey was conducted to assess how accurate hypertension database was in a health care facility. Data was collected from progress notes, lab results, and consult notes from the 3 years before the compilation date and also from available patient folders. Three criteria were used to determine if an individual was classified as having been diagnosed of hypertension. Firstly, if the patient diagnosed of hypertension and was recorded by a physician in the folder, secondly, if they had a prescription for an antihypertensive medication in the context of an elevated blood pressure reading and lastly if their recorded blood pressures was consistent with the diagnosis outlined in the Canadian Hypertension Education Program guidelines. The chart audit diagnosis was used as the reference using a binomial probability distribution. In the study,
sensitivity was defined as the percentage of individuals who were diagnosed of hypertension using administrative data. Specificity was also defined as the percentage of individuals who were not diagnosed of hypertension using the administrative data. Positive predictive value was also defined as the proportion of individuals noted to have hypertension in the administrative data whose diagnosis was confirmed by chart audit whiles negative predictive value was defined as the proportion of individuals identified as not having hypertension in the administrative data whose lack of a hypertension diagnosis was confirmed by chart audit. The study compared the accuracy of multiple algorithms for defining hypertension in administrative data to real-world patients whose charts we reviewed from family practices in Ontario. It was therefore confirmed by Tu, Campbell, Chen, Cauch-Dadek, & McAlister (2007) that the use of administrative data to define hypertension and conduct ongoing surveillance of prevalence and incidence is feasible and practically accurate in adult patients.

2.3.4 Timeliness

Data timeliness can be referred to as the expectation of the time data should be obtained and how effective it can be utilized. There is mostly a mismatch between the expectation and reality which often results in data not being used effectively (Nektar Data Systems, 2016). Vaziri & Mohsenzadeh (2012) contends that the timeliness of data can also be viewed as the “rate at which a particular set of data is current in relation to a specified time”. He added that data are well-timed when they are current and readily accessible.
2.3.5 Consistency

Thatipamula (2013) defines consistency as the extent to which data remains the same all things being equal. In other words, there should not be any substantive difference when there is comparison of two or more data.

2.3.6 Validity

This is ascertained by whether the data seeks to measure what it anticipates to measure (Vaziri & Mohsenzadeh, 2012). Thatipamula (2013) asserts that issues with validity mostly emanate from processes, rather than the final result of the processes. He added that there are always problems with the use of paper for collecting data as this makes invalid data very hard to change thus increasing the cost and bringing about waste. Chen et al., (2014) also contended when there is no use of papers in data collection, devices such as smart phones and tablets are used. This allows for easy changes to forms which are quickly implemented.

To determine data transfer of health services pertaining to maternity, completeness and accuracy dimensions were used by Amoakoh-Coleman et al., (2015). Similarly, the completeness and accuracy dimensions were used by Chahed et al., (2013). They evaluated the completeness and accuracy of immunization data laying emphasis on reported values for DPT3 using daily and monthly PHC reporting forms. For the purpose of the study completeness and accuracy will be of much concern.

2.4 Measuring data quality

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

2.5 Importance of Data Quality

Improvement in patient care has tremendously been attributed to the health care sector which is progressively utilizing the results of research on data quality (Mikkelsen & Aasly, 2005; Fung et al., 2008, Rizi & Roudsari, 2013). This buttresses the point made by NHS Connecting for Health, (2013) that patients are more likely to receive better care if performance data used to support quality improvement is of good quality and reflects actual performance. Also effective decisions are made when clinicians or health workers have access to reliable data. Furthermore, health facilities and organizations can greatly benefit from quality data to be more efficient and effective in planning and meeting the needs of patients and service users. Apart from the health sectors, data quality greatly benefits other areas in biology, chemistry and engineering (Fu et al., 2011; Yang et al., 2013; Oltean-Dumbrava, Watts & Miah, 2016).

2.6 District Health Information Management System–2

The Ghana Health Service (GHS) in collaboration with Oslo University developed and implemented an innovation which is an online dataset known as the District Health Information Management Systems, (DHIMS-2). This was an upgrade of an earlier version which was not web-based, hence the two (2) attached to this one. This innovation was implemented to provide a comprehensive HMIS solution to help health administrators of various health districts to efficiently report and analyze data when collected. Due to the centralization of this software, online data are easily updated whiles it makes it easier to manage the software. The options of this application make the data entry process very simple as compared to the paper-based forms.
It also aids in the analysis of quality data. The control panel of this software can also be customized to serve an intended purpose especially with monitoring and evaluation.

The interface of the DHIMS-2 is not complex or difficult to use. It can easily be used without any complications. However, training is still required to equip data entry personnel as well as end-users to utilize DHIMS-2 optimally for the desired outcome. This is based on three fundamental premises – knowing what one is looking for (whether data element or indicator), where one requires this data or information from (location- regional, district, sub-district or facility level) and when (period or point) time reference. This software has been made available to all 170 districts and utilized by health facilities and directors of district health to perform various kinds of functions such as collecting, collating, transmitting and analyzing routine health service data. Health facilities and District Health Directorates staffs that have the capability of managing the DHIMS-2 have been signed up as secure users. Presently, 5,563 registered users ranging from government staff, quasi-government, private and faith-based facilities submit their service report every month through this software.

With DHIMS-2, data is structured in a hierarchy and reports are produced in the same manner at the various districts. Data can be uploaded at the various facilities and the district levels facility level only and for a facility at the sub district or district level. Reports of administrative activities are entered using the administrative office facility corresponding to the sub-district, district or regional health directorate. To aggregate health data in Ghana, only the population data is uploaded at all the various levels. Facility -level organization units can be moved among sub-districts. In view of this, there would be no need to create them anew when there is a change in boundaries at the sub-district level. The only thing the user needs to have is an installed web browser on a computer and also have access to the internet. To access DHIMS-2 you do not
need any software to be installed on your computer. It works independent of the operating system on your personal computer.

Improvement in data quality and the simplification of how data was acquired was the main aim behind the launching of the DHIMS-2. Formerly, the management of health data was done by the DHIMS-2 which was used by the health facilities. They had to forward their collated data initially to the districts, through to the regional and finally to the national level before data were sent the national dataset. This required long processes and it was time consuming.

2.6.1 Data quality Issues in DHIMS-2

Comparatively, the processes for acquiring data has drastically reduced indicating that the quality of the dataset in DHIMS-2 has greatly improved regardless of the fact that it is still dependent on the PDC. Despite all the benefits that can be derived from the DHIMS-2, there are some issues that have been identified in its utilization. The issues have basically been grouped into three categories: organizational, technical and behavioral.

2.6.1.1 Organizational determinants

According to Aqil et al. (2009), the performance of the DHIMS-2 is likely to be affected by organizational factors such as resource accessibility, money matters, and dissemination of information and attainment of strategic goals. To measure how the goals of data quality can be strategically achieved, Ajzen (2005) outlined certain indicators which included utilizing HIS information, response from the staff and the community, answerability, obligations awareness and analytical. HMIS plays a key role in how the HIS performs. This is evident in the vision statements of HIS and how support services of HIS are established and maintained. Odhiambo-Otieno (2005a) adds that when the levels of support services are identified, the most pressing activities are likely to be performed all things being equal.
2.6.1.2 Technical determinants
These factors are associated with the specialized know-how to improve, control, and develop HIS processes and performance. They are referred to as how indicators are developed; how forms for collecting data are designed; how manuals are made, I.T. types and developing a software for analyzing and processing data. It was further indicated that, assuming the indicators are insignificant, forms for collecting data will be difficult to fill, and if the system’s software is not convenient and straightforward to use, performance of the HIS will be affected. In the same way, the use of information will be affected when data is not properly processed by the software on timely basis. This can therefore lead to analysis which provides meaningless conclusions for making effective decisions (Aqil et al. 2009).

2.6.1.3 Behavioral determinants
The performance and processes of HIS are directly affected by behavioral factors such as: the requirements of HIS users, determination, encouragement and level of expertise in the performance of HIS tasks. The performance of a task can be greatly affected by the feelings of an individual about the end results of a task or how confident the individual can be in performing a task, as well as how difficult and complicated the task is (Hotchkiss et al., 2010).

2.7 Previous studies on Data Quality
Quite a number of studies have been conducted on data quality in both developing and developed countries. Findings from a study conducted by Mavimbe, Braa & Bjune (2005) revealed that data quality was poor in different settings like Mozambique and Kenya. This confirms a study conducted by Odhiambo-Otieno (2005b). He posited that data quality was poor in different resource-limited settings and different technical, behavioral, and organizational factors have been identified. It was clearly seen that the information systems failed to produce the desired results. Da Silva & Laprega (2005) also asserted that in Brazil,
there was a limitation to information use and making decisions. Mapatano & Piripiri (2005) add that many factors have contributed to low performance of information systems. These factors included complexities in measuring indicators as a result of bad selections of denominators in DR. Congo. Other observations included mistakes in reports; lack of computerization, financial constraints and lack of support in Kenya. Another study conducted in Tanzania by Nsubuga, Eseko & Tadesse (2002) also identified loopholes in how cases were defined, report quality, response, management and data interpretation.

A study from Uganda revealed that there was low information use (24%), which was consistent with the limited observed skills level to interpret (41%) and use information (44%) (Aqil et al., 2009). Another study conducted on developments of HMIS in Uganda stated that for the past 5 years, there has been a remarkable improvement in timeliness of monthly reports of outpatient data from the districts to the central level from a national average of 21% in 2000, to 63% in 2002, and to 88% in 2004 (Peter et al., 2005). Again, limited knowledge of the usefulness of HIS data was found to be a major factor in low data quality and information use in Kenya (Odhiambo-Otieno 2005b).

A study conducted on utilization of HIS in Jimma Zone by Sultan, Challi and Waju (2011) indicated that 8 (26.7%), 57 (31.3%), and 54 (36.0%) units and/or departments of Health Posts, Health Center’s, and District Offices, respectively, tried to change data into information, with cumulative (overall utilization of health information) 32.9% units and/or departments of health facilities used their data/information for decision-making, planning, budget and M&E of their activities. There were also poorly coordinated processes and absence of capacity building activities reported.
A further study conducted by Helen in 2011 in Bahr Dar on assessment of HMIS implementation indicated that there was no incentive for information use or motivation to improve information culture. The same study showed that a little below half of the respondents (47.5%) lacked confidence to partake and make decisions for HMIS related activities. This meant that rules and regulations regarding the new HMIS were low; 65.7% of respondents reported inadequate technologies to utilize information; respondents who actually used information for making decisions were 45.6%, 35.3% also used it for future reference and 42.4% used to study trends, with 42.9% to pass report data to health office, (Teklegiorgis et al., 2016).

A report from a study conducted by Gebrekidan, Negus & Hajira (2012) on data quality and information use revealed that a limited utilization of information for making decisions regarding management of implementing programs was studied. It was observed that only 37% of the facilities made decisions based on the results obtained from routine health information. Likewise, supervision and feedback from the senior levels was inadequate as this led to issues of inadequate documentation, late and incomplete reporting, and inaccurate reporting. These findings indicate the rate at which data quality can be greatly affected by limited investment in infrastructure and human resource capacity as well as by the performance of the data aggregation and reporting units of the system. The study also identified there was an observed assigned focal person for HMIS at 25 (78%) health facilities with 7 (28%) of the focal persons having information technology training. In addition, regular budget allocation for HMIS running costs were found at 7 (22%) of the health facilities.

A study conducted by Gashaw (2006) on utilization of HMIS indicated that the overall utilization of HMIS was found to be 22.5%. It was found that the major factors that affect the
utilization of information were training, standard data collection methods, data processing, and reporting. According to the HMIS Task Force in 2011, despite the fact that the available information system facilitates the decentralization process, the information lacks quality and validity, and the use of information for decision-making and providing feedback to concerned parties was not commonly practiced yet.

A study was conducted on implementation of the new HMIS by the Ethiopian FMOH and HMN in 2009. It focused on resources of HMIS, quality of data, utilization of information, staffs’ motivation, etc. The study revealed that the availability of the HMIS registers, forms, and tools were of low standards and the availability of display charts recommended by the new HMIS and the monitoring and evaluation guidelines were found not to be impressive. Likewise, there was evidence of compliance with regular performance reviews and use of data for decision making. This study also revealed that reports based on timeliness ranged from 67% to 100% for hospitals; 86% to 100% for health centres; and from 75% to 86% for health posts, whereas completeness of reports varied from 93% to 96% in hospitals; 89% to 96% in health centres, and 83% to 91% in health posts (Woldemariam et al. 2010).

Generally, HIS performance is evaluated based on HIS input, process, and output and the availability of good data quality and continuous use of information is used to measure performance. This in turn is affected by technical, behavioral, and organizational factors. The PRISM Conceptual framework presented is based on the concept that HIS assessment is one of the HIS inputs which assesses its performance and identifies technical, behavioral, and organizational factors that affect data quality and information use which in turn affects the health system of a country. An HIS strategy will be developed based on the results of this assessment and an intervention will be made according to the HIS strategy.
CHAPTER THREE

METHODOLOGY

3.1 Introduction

This chapter presents the methodological framework that was used in this study. It outlines how data was obtained, managed and analysed. It includes the study area, study design, study population, sample size & sampling procedures, data collection tool and procedure, pre-test, data analysis and presentation.

3.2 Study Area

This study was conducted in the Greater Accra Region. It has 2 Metropolis, 8 Municipalities and 6 Districts. Greater Accra was the Region of choice for this study because for the past five years, it has recorded the highest morbidities in outpatient hypertension. The study included four Public Health facilities namely Maamobi General Hospital, Ga South Municipal Hospital, Taifa Polyclinic, and Kaneshie Polyclinic. These public health facilities which fall within the urban and peri-urban localities are situated in the Ayawaso Sub-Metropolis, Ga South Municipality, Ga East Municipality and Okai koi Sub-Metropolis of Greater Accra respectively, (Figure 3).
3.3 Study Design

A cross sectional study taking retrospective look of records of Outpatient hypertension morbidity data in the consulting room register was used. Through a mixed method approach quantitative assessment of quality of data in the DHIMS-2 and qualitative means to identify factors influencing data quality in respect of completeness and accuracy. A cross sectional study because data was collected at one point in time (Cresswell et al., 2003; Olsen, & St George, 2004; Neuman, 2006; Bernard, 2012).
Secondary data from the third quarter of 2016 records was reviewed. During the time of data collection, the specified period’s data in the DHIMS-2 was electronically locked. Four levels of data sources were reviewed.

a. Patient folders

b. Source data at health facility level (consulting room register)

c. Facility aggregate data (Monthly OPD morbidity returns form)

d. DHIMS-2 data (district, regional and national level data)

3.4 Study Population

The study population was all folders of hypertensive patients that were captured into the consulting room register during the third quarter of 2016.

3.4.1 Inclusion Criteria

Health facilities that made use of Paper-based Data Collection to manage morbidity data in respect of the period under review, which involved physical patient folders and consulting room registers. Folders belonging to patients who were managed as hypertensives during the specified period under review and were captured as such into the consulting room registers of the selected health facilities within the specified period of interest.

3.4.2 Exclusion Criteria

Health facilities that did not make use of Paper-based Data Collection to manage morbidity data in respect of the period under review but rather used Electronic-based Data Collection. No standardized EDC software has been recommended by GHS hence facilities that uses such
software cannot be included in this study. Again facilities that did not capture any hypertensive data into the DHIMS-2 within that three months of 2016 (July, August and September).

### 3.5 Sampling Technique

*Stage 1 selection of administrative units (Metropolis, Municipality and District)*

Among the Sub - Metropolis, Municipalities and Districts in the Greater Accra Region. The first ten ranking units were purposively selected since they had reported relatively high cases of hypertension over the past five years. These ten administrative units were listed on paper and coded as A, B, C, D, E, F, G, H, I & J. These alphabets were then written of ten pieces of papers and folded into a bowl. Without looking, four out of the Ten Administrative units were randomly selected.

*Stage 2 selection of health facilities*

In selecting the health facilities to be included in the study, the first ten high reporting facilities in each of the four selected administrative units were listed on paper and coded from A to J. Pieces of paper numbered A to J were folded into a bowl again and by means of simple random sampling, only one from each of the four administrative units was selected to participate in the study. Based on the inclusion criteria, only the health facilities that had hypertension data reported in the DHIMS-2 with regards to the specified period was selected, hence a facility that was randomly selected and fell within the exclusion criteria the next preferred one was included into the study. In selecting patient folders to be reviewed at the facility level, consulting room registers used within that specified period were all thoroughly assessed and all patients’ folders selected for the study. However, if a folder is not found, it was considered as non-existence.
Stage 3 selection of key informants

Due to the nature of this study, purposive sampling was used to recruit between four and ten health staff who were directly connected to documentation and management of hypertension data in the facility. The following inclusion criteria were considered in selecting the facility staff.

- The staff might have worked directly with hypertension data for the past 2 years of which the variables of interest were documented into folders and consulting room registers.
- The staff had to be directly involved in compilation of facility monthly morbidity form.
- The staff had to be directly engaged in data entry of morbidity form into the DHIM-2.

3.6 Study Variables

Per the focus of the study, only hypertension data recorded in the consulting room registers of the facilities was assessed.

Dependent variables

- New cases of hypertension in consulting room register.
- New cases of hypertension in patient folders.

3.7 Quality Control

A standardized checklist was prepared prior to the start of the fieldwork. This checklist was scrutinized by an experienced supervisor and first pre-tested in a health facility that met the inclusion criteria and the weaknesses identified were corrected. Data collectors were trained on how to use the checklist to avoid any errors during the data collection.
3. 8 Data collection

The investigator initially gathered the consulting room registers used in the months of July, August and September of 2016 from the selected facilities. Data for the months of July, August and September of the year was reviewed because similar studies done made use of the same months, hence for easier comparison. Using double data visual verification by two data collectors, cases of hypertension captured in the consulting room registers were enumerated on paper indicating patient ID numbers and date of attendance. Medical Records staff of the facility retrieved the identified folders and validation of the case as to whether on the date the case was captured, the diagnosis was new or old. This was determined by reviewing the history and previous attendances of the patient. The status of the diagnosis was then compared to what was captured into the consulting room register. This led the investigator to assess the first level of error, (Error 1). To assess the second level of error, (Error 2) the data collectors manually counted all the new cases of hypertension captured in that period and compared it to what had been recorded in the Monthly Outpatient Morbidity Form (Facility aggregate form) and used as reference. The next level of error (Error 3) was determined by comparing the recorded data in the facility aggregate form to the DHIMS-2 data.

The biostatistics department of the health facility provided the DHIMS-2 data. The three sources of data were compared to determine errors in data transfer from one source to the other. Error 4 was, however, determined by comparing the newly diagnosed hypertension cases in the patients folder to what exists in the DHIMS-2 database which was what was used for reporting hypertension cases from the Facility, District, Regional and National levels. A pretested data abstraction tool was used to collect the data of interest. To identify the factors affecting completeness and accuracy of the data, staff directly involved with managing hypertension data were interviewed using a semi-structured interviewer guide. Questions on the guide covered
knowledge acquisition or understanding on hypertension data management, coaching or supervision on how to capture hypertension data, data validation activities and the importance attached to ensuring quality hypertension data in the facility.

### 3.9 Data analysis

Completeness and accuracy of data transfer from primary source to other levels was computed. The formulae that was used was adapted from a similar study done by Amoakoh-Coleman & Kayode (2015). For completeness which is percentage of missing data for the variables (status of principal or additional diagnosis) was calculated as:

\[
\text{Completeness} = \left( \frac{\text{md}}{\text{di}} \right) \times 100
\]

Where \( \text{md} \) = total number of reported data; \( \text{di} \) = total number of data inspected

For accuracy of data transfer, four types of percentage errors in data transfer across the four data sources were calculated

\[
\text{Percentage Error} = \left( \frac{\text{Difference between both data sources}}{\text{total data entered}} \right) \times 100\%
\]

\[
= \left( \frac{\text{di} - \text{dj}}{\text{D}} \right) \times 100
\]

(Where “d“ is the actual number of data in the data source; \( \text{di} \), \( \text{dj} \) represent the sequence in data gathering sources with \( j \) always acting as the reference for \( i \); and \( \text{D} \) is the total data inspected for the variable).

Percentage Error 1 is a measure of the error involved in the transfer of data from the primary sources (folder) to the consulting room register. It estimates the deviation of the person who transferred the data from the folder into the register.
Percentage Error 2 is a measure of the disparity between the data in the consulting room register, and what is collated into the facility aggregate form. It estimates the deviation of the monthly OPD morbidity from the consulting room register.

Percentage Error 3 is a measure of the error that occurs during the transfer of data from the aggregate forms into the DHIMS-2 software. It estimates the deviation of the facility aggregate data from the DHIMS-2 data, mainly due to data entry errors.

Percentage Error 4 is a measure of the error that exist between the primary data source i.e patients’ folder and the DHIMS-2 software. It estimated the deviation of true new cases of hypertension and what sits in the DHIMS-2 which is what is used for taking decisions. The 95% confidence interval of completeness and percentage error was computed using the formula;

\[ y = p \pm z \sqrt{p(1-p)} \div d_1 \]

Where \( y \) = 95% confidence interval of the estimate; \( p \)=% of missing data or error rate; \( z = 1.645 \) (1-sided alpha level of 0.5); and \( d_1 \) = total number of data inspected.

To identify the factors associated to incomplete and inaccurate hypertension data, thematic analysis was used. Questions related to pre-determined thematic areas were asked staff who were responsible for managing data in the facility. The responses were coded, entered into an excel spreadsheet and unique themes that run through the responses were identified.

3.10 Ethical approval

This study was carried out in collaboration with the biostatistics department of the Greater Accra Regional Health Directorate. Consent was sought from the directorate to use the routine
secondary data from the selected facilities as well as what was available in the District Health Information Management System version II (DHIMS-2). Data is readily available to health workers who seek permission to use for research purposes. Health personnel involved with data management at the health facility and district levels were briefly consulted to share their experiences with data management. This helped identify some of the factors that contribute to errors in data management. Since there were no direct human interactions or follow up, there was no need for an ethical approval.

3.11 Pre-test

The data collection process was tested at the Ga West Municipal Hospital in the Greater Accra Region to identify the weakness in the process and address them before the actual study was done. However, that facility was not included in the original study. This was done to enable the investigator identify some bottlenecks and gaps that could invalidate the study.
CHAPTER FOUR

RESULTS

4.1 Introduction

This chapter consists of two sections: the results describing the data completeness and data accuracy (error rate) of hypertension data which estimates the quality of data from four health facilities in the Greater Accra Region. The other section describes the qualitative data extracted from the key informants which revealed factors associated with incomplete and inaccurate hypertension data. The results of the completeness and accuracy rates are presented in frequencies and percentages in the form of tables. Statistical tests on the relationship between the variables of interest are also presented in this section.

A detail description of the distribution of hypertension data in terms of percentage completeness and percentage error on sex and age groups of the clients that attended the four health facilities during the third quarter of 2016 is shown in Table 1. It reveals which sex and age groups were mostly affected or otherwise. However, percentage completeness and missing data among the four health facilities is also seen in Table 2, whereas, the percentage of error in transferring hypertension data from the primary data source as it flows to the DHIMS-2 database is also shown in Table 3.
Table 1: Distribution of percentage completeness and error rates of hypertension data on sex and age groups in four health facilities in Greater Accra Region

<table>
<thead>
<tr>
<th>Facility</th>
<th>Cases (N=931)</th>
<th>% Completeness</th>
<th>% Error 1</th>
<th>% Error 2</th>
<th>% Error 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taifa Poly clinic</td>
<td>n=215</td>
<td>65.9</td>
<td>76.3</td>
<td>22.1</td>
<td>97.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>69</td>
<td>21.1</td>
<td>24.5</td>
<td>7.1</td>
<td>31.2</td>
</tr>
<tr>
<td>Female</td>
<td>146</td>
<td>44.8</td>
<td>51.8</td>
<td>15</td>
<td>65.9</td>
</tr>
<tr>
<td>Age group(yrs)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20-49</td>
<td>99</td>
<td>30.3</td>
<td>35.1</td>
<td>10.2</td>
<td>44.7</td>
</tr>
<tr>
<td>50-69</td>
<td>99</td>
<td>30.3</td>
<td>35.1</td>
<td>10.2</td>
<td>44.7</td>
</tr>
<tr>
<td>70+</td>
<td>16</td>
<td>5.3</td>
<td>5.9</td>
<td>1.7</td>
<td>7.7</td>
</tr>
<tr>
<td>Kaneshie Poly clinic</td>
<td>n=381</td>
<td>65.1</td>
<td>89.0</td>
<td>7.8</td>
<td>96.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>109</td>
<td>18.6</td>
<td>25.5</td>
<td>2.2</td>
<td>27.7</td>
</tr>
<tr>
<td>Female</td>
<td>272</td>
<td>46.5</td>
<td>63.5</td>
<td>5.6</td>
<td>69.2</td>
</tr>
<tr>
<td>Age group(yrs)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20-49</td>
<td>170</td>
<td>29.0</td>
<td>39.7</td>
<td>3.5</td>
<td>43.2</td>
</tr>
<tr>
<td>50-69</td>
<td>182</td>
<td>31.1</td>
<td>42.6</td>
<td>3.7</td>
<td>46.3</td>
</tr>
<tr>
<td>70+</td>
<td>29</td>
<td>5.0</td>
<td>6.7</td>
<td>0.6</td>
<td>7.4</td>
</tr>
<tr>
<td>Ga South Hosp.</td>
<td>n=157</td>
<td>51.6</td>
<td>35.4</td>
<td>34.9</td>
<td>10.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>46</td>
<td>15.1</td>
<td>10.4</td>
<td>10.2</td>
<td>3.2</td>
</tr>
<tr>
<td>Female</td>
<td>111</td>
<td>36.5</td>
<td>25.0</td>
<td>24.7</td>
<td>77</td>
</tr>
<tr>
<td>Age group(yrs)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20-49</td>
<td>77</td>
<td>25.3</td>
<td>17.4</td>
<td>17.1</td>
<td>5.3</td>
</tr>
<tr>
<td>50-69</td>
<td>60</td>
<td>19.7</td>
<td>13.5</td>
<td>13.3</td>
<td>4.2</td>
</tr>
<tr>
<td>70+</td>
<td>20</td>
<td>6.6</td>
<td>4.5</td>
<td>4.5</td>
<td>1.4</td>
</tr>
<tr>
<td>Maamobi Gen. Hosp.</td>
<td>n=178</td>
<td>59.9</td>
<td>90.9</td>
<td>21.0</td>
<td>111.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>69</td>
<td>23.2</td>
<td>35.2</td>
<td>8.1</td>
<td>43.4</td>
</tr>
<tr>
<td>Female</td>
<td>109</td>
<td>36.7</td>
<td>55.7</td>
<td>12.9</td>
<td>68.5</td>
</tr>
<tr>
<td>Age group(yrs)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20-49</td>
<td>90</td>
<td>30.3</td>
<td>45.9</td>
<td>10.6</td>
<td>56.6</td>
</tr>
<tr>
<td>50-69</td>
<td>67</td>
<td>22.5</td>
<td>34.2</td>
<td>7.9</td>
<td>42.1</td>
</tr>
<tr>
<td>70+</td>
<td>21</td>
<td>7.1</td>
<td>10.8</td>
<td>2.5</td>
<td>13.2</td>
</tr>
</tbody>
</table>

Error 1 is a measure of the difference between the data in the folder and that in the consulting room register.
Error 2 is a measure of the discrepancy between the data in the CRR and that in the Facility morbidity form.
Error 4 is a measure of the discrepancy between the data in the folder and that in the DHIMS-2.
% Completeness is a measure of the number of cases that is accounted for in the DHIMS-2.
*Error 3 was omitted since 0% error was determined for all health facilities.
Source (Field Survey, 2018)

Out of the 931 new cases of hypertension reported into the database, 64% of the case were seen at the polyclinics and the remaining at the hospitals, however, the polyclinics covered greater
percentage completeness which means that relatively lesser percentage of missing data. Again in all the health facilities, women formed a greater section of the new cases of hypertension during that period. This implies that whatever the effects of poor data quality are in terms of completeness and accuracy, women are affected the most. As studies have reported, hypertension mostly affects people aged from 20 years and above and the result in table 1 also confirms this. Furthermore the study found that most of the cases were within the ages of 20 to 49 years hence the errors identified in the data transfer which affects quality of healthcare will therefore affect people within that age group as indicated in Table 1.

<table>
<thead>
<tr>
<th>Facility</th>
<th>Total cases counted (a)</th>
<th>Total cases reported (b)</th>
<th>% Completeness (b/a)*100 = c</th>
<th>% of missing data (100 – c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taifa P’oly (A)</td>
<td>326</td>
<td>215</td>
<td>65.9</td>
<td>34.1</td>
</tr>
<tr>
<td>Kaneshie P’(B)</td>
<td>588</td>
<td>381</td>
<td>65.1</td>
<td>34.9</td>
</tr>
<tr>
<td>Ga South (C)</td>
<td>299</td>
<td>157</td>
<td>51.6</td>
<td>48.4</td>
</tr>
<tr>
<td>Maamobi (D)</td>
<td>291</td>
<td>178</td>
<td>59.9</td>
<td>40.1</td>
</tr>
</tbody>
</table>

Mean 60.7 39.3

% Completeness is estimated by dividing the total cases reported by total cases counted and multiplied by 100. % missing data is estimated by subtracting the % completeness from 100.

Source (Field Survey, 2018)

4.3 Overall completeness of hypertension data among the four facilities in GAR

An estimation of the percentage of the missing data in the primary data source (CRR) ranged between 34.1% and 48.4%. There were variations between the polyclinics and the hospitals.
Among the polyclinics, the percentage completeness ranged from 65.1% to 65.9%, whereas that of the hospitals also ranged from 51.6% to 59.9%. The least percentage of missing data recorded were associated with the polyclinics and the highest linked to the hospitals though the later recorded relatively fewer cases of hypertension during the third quarter of 2016 in the DHIMS-2 database as indicated in Table 2. The mean completeness of hypertension data among the four health facilities was 60.69% (95% C.I = 49.32%–72.06%).

Table 3: Percentage Error of hypertension data in four health facilities in the Greater Accra Region during the third quarter of 2016

<table>
<thead>
<tr>
<th>Facility</th>
<th>1° data (folder)</th>
<th>2° data (CRR)</th>
<th>Facility data</th>
<th>DHIMS data</th>
<th>Data inspected</th>
<th>% Error 1</th>
<th>% Error 2</th>
<th>% Error 3</th>
<th>% Error 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taifa P’oly (A)</td>
<td>73</td>
<td>186</td>
<td>215</td>
<td>215</td>
<td>155</td>
<td>76.3</td>
<td>22.1</td>
<td>0</td>
<td>97.1</td>
</tr>
<tr>
<td>Kaneshie P’ (B)</td>
<td>83</td>
<td>357</td>
<td>381</td>
<td>381</td>
<td>310</td>
<td>89.0</td>
<td>7.8</td>
<td>0</td>
<td>96.9</td>
</tr>
<tr>
<td>Ga South (C)</td>
<td>150</td>
<td>181</td>
<td>157</td>
<td>157</td>
<td>165</td>
<td>35.4</td>
<td>34.9</td>
<td>0</td>
<td>10.9</td>
</tr>
<tr>
<td>Maamobi (D)</td>
<td>34</td>
<td>153</td>
<td>178</td>
<td>178</td>
<td>134</td>
<td>90.9</td>
<td>21.0</td>
<td>0</td>
<td>111.9</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>72.9</td>
<td>21.5</td>
<td>0</td>
<td>79.2</td>
</tr>
</tbody>
</table>

Error 1 is a measure of the difference between the data in the folder and that in the consulting room register (CRR).
Error 2 is a measure of the discrepancy between the data in the CRR and that in the Facility morbidity form.
Error 3 is a measure of the difference between the data in the Facility morbidity form and that in the DHIMS-2.
Error 4 is a measure of the discrepancy between the data in the folder and that in the DHIMS-2.
1° data is the new cases of hypertension counted in the patient’s folder (primary source).
2° data is the new cases of hypertension counted in the consulting room register.
DHIMS data is the reported new cases of hypertension in the database.
Source (Field Survey, 2018)
4.4 Accuracy of data transfer

In this study, four sources of data were compared for accuracy. The most reliable in terms of accuracy of hypertension data transfer, was from the facility aggregate form to the DHIMS-2 database (error 3) with a percentage error of 0.0% across all the facilities. This means that there was very slight or no error in transferring the facility aggregate data to the DHIMS-2. Percentage error 1 ranged from 35.4% to 90.9%, however, the mean error 1 rate among the four health facilities were statistically not different from each other and this was significant at [\( F (3, 8) = 2.82, p = 0.1069 \)]. Error 2 was relatively lesser across all the facilities compared to Error 1, and ranged from 7.8% in facility B to 34.9% in C, (Table 3). The mean error 2 rate was statistically not different from each other between the selected levels of facilities, [\( F (3, 8) = 0.43, p = 0.7344 \)]. Percentage error 4 was generally very high in three of the facilities, ranging from 10.9% in facility C to 111.9% in facility D, (Table 3). This means that considerably, the error that exists in the primary source data (folder) was very high compared to what exists in the DHIMS-2 database. There was a statistically significant difference in the mean error 4 rate between all the facilities as determined by one-way ANOVA [\( F (3, 8) = 8.83, p = 0.0064 \)]. A Tukey post-hoc test revealed that Error 4 was statistically significantly lower in facility C compared to facility A (-86.1 ± 21.9, p = 0.018). Error 4 was also statistically significantly lower in facility C compared to facility B (-85.9 ± 21.9, p = 0.019). However, the mean Error 4 was statistically significantly higher in facility D compared to facility C (100.9 ± 21.9, p = 0.008).

Accuracy of hypertension data at each of the four health facilities

For facility A, in terms of the error of transferring new cases of hypertension from patient’s folder to the consulting room register, defined in this study as Error 1, it made 76.3% (95%
CI= 22.0% - 130.6%). Error 2 was relatively lower in this facility at 22.1% (95% CI= -35.5% - 79.8%) and for Error 3, which deals with transfer of hypertension data from the facility aggregate form to the DHIMS-2, it made no or insignificant error, hence, the 0.0%. The situation is, however, different with error 4, which made a representation of 97.1%. This defines the error in transfer of new cases of hypertension data from patient’s folder to the DHIMS-2 database, (Table 3).

For facility B, the level of error committed in Error 1 was 89.0%. Error 2 was relatively lesser than all the other facilities at 7.8%. Similar to all the other facilities, Error 3 was also 0.0% with Error 4 being as high as 97.1%, (Table 3).

In facility C, Errors 1 and 4 were considerably lesser than all other health facilities at 35.4% and 10.9% respectively. However, this facility recorded the highest among the other facilities in respect of Error 2 at 34.9% and similar to the rest of the facilities being studied, Error 3 was also 0.0%, (Table 3)

Facility D recorded the highest level of error in terms of Error 1 (90.9%) as well as Error 4. The percentage error 4 was as greater as 111.9% and this is the measure of discrepancies with hypertension data as it exists in the DHIMS-2 compared to what really exists in the patient folders. A 100% error rate implies everything was wrong, the data that exist in the DHIMS-2 is highly at variance with what was inspected and verified in the patient folders. The over 100% error rate therefore is an indication that a huge discrepancy of hypertension data in the DHIMS-2, as against what was inspected for this particular facility. With Error 3 also being 0.0% just as all the other facilities, Error 2 was noticeably lesser compared to Errors 1 and 4 at 21.0%, (Table 3).
This study in whole, assessed new cases of hypertension captured into the web-based database to ascertain the veracity or otherwise of the cases that were seen at the OPD of the various health facilities. Out of the 931 new cases reported in the database, few were assessed to be true, (37%). In Taifa polyclinic for instance, only 34% of the cases were measured to be true as indicated in figure 4 above. Same applied to the Kaneshie polyclinic and Maamobi General Hospital with 22% and 19% respectively, however Ga South hospital had about 96% which indicates a high level of accuracy of the data, (Figure 4).

**Qualitative data analysis**

The semi-structured interview guide which sort to derive information from twelve (12) key informants on the issues affecting the quality of hypertension data found these recurring themes; Poor understanding due to lack of training or coaching on how to document
hypertension data, less priority on quality of hypertension data causing poor use of the data and inadequate supervision and data verification activities on hypertension data quality. These health staff comprised facility staff who had a minimum of 2 years and a maximum of 10 years of working experience either at the Medical records/Biostatistics or Out Patients Departments of the facility.

All the four health facilities studied were public facilities and classified under the secondary level of health care. Hence the total of 931 hypertension cases reported into the DHIMS-2 database by these health facilities during the third quarter of 2016 can be identified to the section of clients in the population who predominantly reside in urban and peri-urban settings and receive health care at the secondary level of health delivery in public facilities.

**Inadequate training on hypertension data management**

The responses given by the facility staff revealed that, when especially a nurse is assigned the role of capturing data into the secondary data source i.e. consulting room register, no specific training or tutorials is given focusing on the peculiarity of documenting chronic conditions of which hypertension is key. “We are normally told to fill every indicator in the register especially with regards to malaria cases but for hypertension and others they do not give further explanation so we capture them as we see in the patient’s folder.” *(Key informant 7).*

Secondly, the gap in knowledge sharing among working colleagues in some of the facilities lead to the situation where one individual is given some form of on the job training as to how to ensure completeness in data capture and accuracy in transferring the data from one source to the other, however, when a new person assumes that role of managing data, some form of anomalies are committed, since that person lacks the understanding to carry out the task.
Poor data management

Among the different cadre of health staff who recounted their experience with managing hypertension data, almost all of them identified the fact that, in the routine health information management systems at the facility levels, no form of monitoring is done to ask questions in relation to the quality of hypertension data in the system. However in facility C and D where a special clinic is operated solely for diabetics and hypertensives, the staff there revealed that sometimes during annual performance review sections, doubts about the accuracy of hypertension data reported in the DHIMS-2 are raised, yet no effort is made to verify the claim. “For a disease like malaria, because there is a National Control Program, you normally see officials from the municipal or regional level coming to validate the data but such is not seen with hypertension” (key informant 4).

Inadequate validation of hypertension data

The health information management systems at the facility level also a component of verifying reported data to check underestimation or overestimation. This exercise mostly done on monthly basis focuses on some selected indicators of interest to the facility, however, hypertension data quality is missing out. Almost all the key informants spoke to the fact that although the validations are held to improve data quality in the facility, hypertension morbidity is not covered.

“We normally validate antenatal and neonatal care and sometimes the epidemic prone diseases but hardly do we concentrate on hypertension issues, maybe now we will include it.” (Key informant 5).
Besides the direct information given by the respondents, we also highlighted certain key practices and conditions that affect data quality in general. For instance in the polyclinics, there was lack of human resource, relatively fewer people burdened with the task of collating data from several sources as compared to the hospitals. “This is not the only work I do routinely and sometimes when I go on leave, I return with the previous months work waiting for me so it is not easy” (Key informant 2). This challenge was common in all the four facilities, however the hospitals suffered less in relation to human resource challenges. Moreover, it is interesting to note that the volume of cases seen by these two different levels of health facilities are almost the same though human resource allocation are not the same.

Again, we gathered through the interview, that though Standard Operating Procedures, (SOPs) were found in all these facilities and existed to guide on data capture and collation, staff assigned to these responsibility were not given any hint on the need to consult from such sources although the book clearly defines what should be captured as a new case of hypertension or old case.

The qualitative analysis revealed several factors that affect quality of routine health information such as the less attention or importance given to hypertension data, lack of capacity building for health facility staff and inadequate validation activities focus on hypertension. It was also evident that the staff did not have regular training programs for development of specific health information skills. There were adequate data management processes in place to ensure the production of good quality data at the health facilities but there was less concentration on the non-communicable diseases apart from them being the leading cause of death.
CHAPTER FIVE

DISCUSSIONS

The study sought to assess the quality of hypertension data in the DHIMS-2 database within four health facilities in the Greater Accra Region by estimating the completeness and accuracy rates of the data as it exists in primary data sources and reporting tools.

Most of the hypertensives were between the ages of 20-49 years representing about 46%. The female proportion of hypertensives, which formed the majority, was 69%. Studies have reported that age is significantly associated with hypertension (Kearney et al., 2005) and as shown in (Table 1), most of the cases were aged between 20 to 69 years. Being a female aged 20 years and above with hypertension, therefore increases one’s chances of having their data not accurately captured in the DHIMS-2. Again, being managed as a hypertensive in a polyclinic also increases one’s chance of having their data not accurately reported in the database as indicated in Table 2.

In this study, the overall completeness rate of hypertension data in all the primary source data sampled among the four health facilities was 60.7% (95% C.I= 49.3% - 72.1%) indicating a percentage missing data of 39.3% which is far above the acceptable range compared with reported ranges of other studies done. For instance, in Nigeria, a study by Adejumo in 2017 found the overall average completeness of monthly summary forms and that of their DHIS to be 77.3% whilst for paper records, completeness rates were identified in Tanzania to be 64.2% (Simba & Mwangu, 2005). Moreover, the completeness rate for HIV data in the DHIS were found to be 50.3% in South Africa (Mate et al., 2009). The 39.3% of missing hypertension data was well above the 5.7% and 3.1% for maternal health and neonatal health indicators respectively reported in other studies in the DHIMS-2 database.
In public health, completeness of hypertension data in the primary source data (CRR) is very vital to estimating precisely the incidence or prevalence of the condition among the population. Hence basing on the logic and provision that, the DHIMS-2 database is the recognized source for health information by the Ghana Health Service, care must be taken in the use of such information. It was observed during the visual data verification that, accurate documentation of hypertension in the primary source data was a big challenge therefore rendering most of the data invalid to be counted and reported onto the facility aggregate form. In adhering to good data collection standards from primary source data, incomplete data was not counted since they give room to ambiguity and inconsistency. Hence as part of our data collection precautions, we only counted data which had been completely documented. Facility C which recorded the highest had as much as 48.4% of missing data, which virtually meant that almost half of the hypertension data captured into this facilities primary source data i.e. CRR, during the specified period were not valid to be accounted for.

The results of this study indicate that, there are varied deviations in the completeness and accuracy of data transfer of primary source data in almost all the facilities and this should be cautiously considered when making use of the data. Fundamentally these discrepancies are hinged to the fact that recording of data into these sources is basically manual and paper – based. As determined by another study, primary source to database error rate depend extensively on the aggregate of structured data collection in a clinical setting and also on the density of the medical record (Nahm, 2008). Nonetheless, in this study, once the data are captured in the facility aggregate forms, they are correctly transmitted to the web-based software, DHIMS-2. It is important to also note that, the Ghana Health Service in collaboration with other partners such as CHIM and NMCP, developed Standard Operating Procedures and Guidelines for Data Verification and Analysis to be used by professionals in ensuring high
quality of service data to inform decision making towards the strengthening of health systems in Ghana. All these instruments clearly define what quality data is and prescribe how to achieve accuracy and completeness. Hence there are approved standardized procedures in ensuring high quality health service data in all health facilities. It was therefore clear as to what acceptable standards of data quality are for Ghana. Granting the fact that there are no specific global and uniformly defined guidelines on the subject, the quantification of data quality using percentage error from the sources of data is acceptable, as done in several studies, hence same method being applied in the percentage errors calculated in this study.

In a related study in the UK, to identify methods to defining hypertension in electronic medical records by comparing the prevalence and treatment rate of hypertension in The Health Information Network database (THIN) with results from Health Survey for England (HSE), the use of diagnosis code alone underestimated the prevalence by 14.0% whilst the combination of antihypertensive drug prescriptions and abnormal blood pressure also overestimated the prevalence by 38.4% (Peng et al., 2016).

In Mozambique, completeness of manual data was assessed to be between 37.5% and 52.1% for primary health care data (Gimbel, 2011). Another study done in Mali and Senegal also established a mean completeness from 94.0% to 97.0% by assessing the completeness of maternal and perinatal care services using information system data (Dumont, 2012). A similar study in Ghana to assess the completeness of routine maternal health service data as well as its accuracy in transferring to various levels established that the completeness of the antenatal variables sampled was 94.3% (% missing data of 5.7%) for facility level aggregate data. By reviewing the first quarter of 2012 data in that cross sectional study, they also found the error in transferring data from primary source data to aggregate form and DHIMS-2 to be 1.0%
Likewise a similar study done in seven out of the ten district in Greater Accra by assessing seven neonatal health indicators to quantify the completeness and accuracy of the data sampled established that the overall error rate of the DHIMS-2 database was 0.68% and the percentage of missing data was 3.1% (Kayode et al., 2014). Both studies made their conclusions about data quality by comparing their work to earlier works done in this field.

Over the years the consensus on what should be regarded as an acceptable error rate for high quality data has not yet been firmly established, which is why this value differs across clinical and pharmaceutical fields (Fong, 2001). Largely, the majority of the experts in this field has agreed that 0.5%, 0–0.1% and 0.2–1%, should be measured as the acceptable error rate in respect of the overall critical and non-critical variables as proposed by the Society for Clinical Data Management in 2000. However, it is important to note that visual data verification, approximately, misses 15% of errors (Fong, 2001). Therefore for 0.5% as an acceptable error rate, the error rate of the primary entered database should, therefore, not be more than 3.3% in order to attain an acceptable quality level (Suresh et al., 2012).

It must be noted that in measuring the error rates, the primary source data used was the patient’s folder and the consulting room register became the secondary source data. Commonly in all the health facilities, the facility aggregate form or monthly OPD morbidity form and the DHIMS-2 data (error rate 3) were perfectly identical when compared, hence error rates of 0.0% throughout. This reveals that, once the data is reported onto the monthly aggregate form, persons responsible for entering the data into the DHIMS-2 do not make any mistake. This is a very good practice and highly commendable.
In a sharp contrast to Error rate 3, the study found an overall of 21.5% (95% C.I= 4.49% – 38.48%) error rate in the transfer of hypertension data from consulting room registers to the monthly OPD morbidity form (facility aggregate form). Facility B obtained the least error rate in this regard at 7.8% and highest was facility C, 34.9%. This data transfer activity is exclusively designated to the Biostatisticians in the facility. One major challenge expressed by some of these professionals when investigated was the fact that most of the hypertension diagnosis stated in the consulting room registers had omissions to the status of diagnosis being new or old hence could not be accepted into the aggregate form. This was a typical issue of completeness rate causing this error. Comparing the mean error rate 2 to the 3.3% of acceptable quality level standards determined in earlier studies, this was quiet high and does not render the data to be very reliable. Nonetheless, a spearman’s correlation was run to assess the relationship between completeness rate and error rates 1, 2, and 4 using July to September 2016 hypertension data within the four facilities. There was no statistically significant correlation between completeness rate and the error rates in all the instances. Therefore the value of the completeness rates could not predict the extent of accuracy of the data as determined in this study.

In assessing the accuracy of data transfer from the primary data source (folder) to the secondary data source (CRR), the study established that the overall error rate obtained was 72.9% (95% C.I=52.20% - 93.62%). We observed that, out of the 764 cases that were documented in consulting room registers for having been diagnosed as new cases of hypertension in the third quarter of 2016, 340 of these cases actually happened to be genuinely new hypertensives. This therefore was what resulted in the high error rate estimated which is far above the recommended error rate as certified to exist in good clinical practice. Enquiries in respect to this error revealed that in almost all the four facilities, the data transfer was done by a nurse
who had received some level of training in that aspect. It, however, appeared that, much scrutiny was not done to know the exact status of the patient before transferring the data into the register by the persons responsible. For some of these errors, patients who had either attended the facility on account of review or though known hypertensives used new folders and were classified as new hypertensives. These were issues that with little devotion on the part of the trained professionals could have been dealt with. It is, however, important to also note that, in instances where the physician specifically documented the patients’ diagnosis as “known hypertensive”, no misclassification of the status of diagnosis was made, but when the diagnosis was not specified, the classification was then left to the discretion of the nurse transferring the data which in most cases resulted to this particular error.

One of the motivations for this study, was to determine if the reported new cases of hypertension in the DHIMS-2 database was a true reflection of incidence of the disease in the population. The study therefore measured this by comparing the diagnosis of patients with the condition as it existed in folders to reported values in the web-based database i.e. (error 4) during the specified period of the study. The mean error 4 rate was 79.2% (95% C.I= 49.02% - 109.42%), this obviously was far from an acceptable rate of error (0.5%) for high quality data and the average rates of forty-two source database validation studies found to be 9.76% (Davis et al., 2002) (Nahm, 2008). In another study an error rate of 4.0% was detected when manual records on diabetic cases from primary data sources were compared to an electronic source (Newgard et al., 2012). Apart from facility C which recorded 10.9% of error rate 4, the three other facilities deviated hugely with facility D having over 100% error. This implies that for facilities A, B and D the incidence of hypertension data reported in the DHIMS-2 cannot reflect the true situation on the ground, hence the data is not reliable. The provision put in place at facility D was that, diabetic and hypertension clinic coordinates the capture of new cases of
hypertension into the facility form hence quiet circumspect compared to the other facilities. These pronounced discrepancies established in the DHIMS-2 database could be attributed to varying poor data management practices at the facility levels. As emphasized by Davis and LaCour in 2002 that an original data must be accurate in order to be useful and documentation should reflect the event as it actually happened, same cannot be vouched by the hypertension morbidity data in the DHIMS-2 database. This assertion implies that for a data to be useful it must be accurate and valid as indicated by the W.H.O guide for improving data quality in developing countries. It is also important to note that the picture of incidence of hypertension data in the OPD morbidity dataset of the DHIMS-2 may not correctly reflect the situation within the population due to the less attention, complexity of data compilation and use of the data in taking decision as compared to other maternal and neonatal indicators which mostly has some sort of national interests.

5.1 Factors influencing data quality in the four health facilities

With the development of SOPs to maximize accuracy, completeness, integrity and traceability of data as well as guidelines for ensuring quality data in the Ghana Health Service since the last five years, it is expected that much should be achieved. To identify the factors affecting specifically the completeness and accuracy of hypertension data in the facilities visited, we purposively engaged individuals entrusted with the responsibilities of managing the data. This included three health information officers, three biostatistician assistants, six nurses who recounted their experience with documentation and reporting of hypertension data. Though a structured interview guide was not used to extract the information from them, all the eleven professionals contributions can be narrowed into these thematic areas;

- Lack or inadequate training on how to document hypertension data: generally, coaching and on the job training is given intermittently to nurses who are directly connected to
capturing hypertension data into the registers at the facility level, however, due to regular rotation and changing of the responsibility of nurses from one duty post to the other, persons who get to be trained at one point in time, are reassigned from that responsibility and hand over to persons with less training and understanding, hence affecting the integrity of data capturing process. Again, no form of specific training on hypertension documentation was given to persons who managed the data.

- Poor data management process and less importance attached to hypertension data: in all the four facilities, there was the lack of tenacity and interest to ensure high completeness and accuracy of hypertension data, as seen in other programs like malaria control where there is a focal person with the prime responsibility of coordinating and reporting on malaria to higher levels hence some importance attached to the data, this was not seen in managing hypertension data.

- Inadequate supervision and data verification activities on hypertension data quality: though the facility biostatisticians hold the responsibility of supervising data management activities and establishing validation systems to improve data quality, it was obvious that it was not regularly done and in most cases less attention was given to accuracy and completeness of hypertension data. It was evident that though SOPs and Guidelines exist to ensure accuracy and completeness of data, if an indicator is not prioritized, or given much attention, the quality of that data gets undermined.

- Lack of dedicated staff to manage hypertension data: though several professionals are directly involved with capturing and transferring hypertension data, this study identified that lack of understanding on the exclusive nature of dealing with data of
chronic diseases like hypertension resulting in poor data quality, hence having a dedicated staff to manage the data solely could arrest the problem.

- Though the current consulting room register has a provision for classifying a diagnosis as either a new case or an old case, there were many instances where hypertension diagnosis was not classified (missing bits of data), hence lacking the merit of being included into the facility monthly morbidity form. In facility C for instance, the officer in charge of the data, often made use of the folder numbers of the cases to determine the outcome instead of relying on the status of diagnosis. This is certainly not reliable because a patient attending a facility for the first time might be a known hypertensive and such a case should not be counted into the morbidity form.

5.2 Study strengths and limitations

Though the region of study was purposively selected, the metropolis and municipalities were randomly involved. The health facilities recruited from these administrative units were also randomly sampled based on pre-specified criteria in other to avoid selection bias. Due to the study time frame and limited resources, only few districts and health facilities in the region were engaged hence limits the generalization of the results in all health facilities. None of the private health facilities were included in the study hence results may not be applicable to private health facilities in the region. In almost all the health facilities the data was verified separately by two people hence reducing the probability of committing error during the verification process. The study extended the scope of the validation to the primary source data i.e. patient’s folder as compared to other studies which mostly made use of registers hence boosting the validity of the work. The hypertension data as a single morbidity indicator cannot be used to generalize the entire accuracy of the monthly morbidity form and DHIMS-2, however the
results can be rationalized in the presence of the same underlying mechanisms affecting the
data quality.
CHAPTER SIX

CONCLUSIONS AND RECOMMENDATIONS

6.1 Conclusions

The completeness of hypertension data in the DHIMS-2 database was relatively complete as over 50% of data captured into the consulting room registers are accounted for in the web-based software. In this regard, care must be taken in using the data in the database in taking decisions and making inferences. The accuracy of hypertension data in the DHIMS-2, however, does not reflect what exists in primary sources and this disparity is quite wide. The data in the database should not be represented as the true incidence of the disease among the population hence a more refined means should be adopted to know the facts. Factor affecting data quality in terms of completeness and accuracy in our hospitals and polyclinics were related to less priority or interest placed on quality of hypertension data, inadequate data verification activities on hypertension, poor understanding on hypertension data management and lack of dedicated staff to manage hypertension data.

6.2 Recommendations

- Public health facilities especially Polyclinics and Hospitals should operate hypertensive and diabetic centers and exclusively collate the hypertension data instead of doing that through the general pool of morbidities. This can immediately reduce the errors in capturing and transferring the data.

- Regular validation of hypertension data: all facility validation teams must add hypertension morbidity to their list of indicators to be validated every month.
Facilities must task persons to act as hypertension focal persons who will coordinate, monitor and supervise data on hypertension and if possible other chronic conditions like Diabetes to ensure data quality. As revealed in a similar study, using dedicated staff to manage data improves the quality of data (Amoakoh-Coleman et al., 2015).

The program managers of non-communicable diseases at the national level and the center for health information management must institute a hypertension e-tracker database to uniquely monitor hypertension cases reported at health facilities just as same is done for HIV and TB cases.

Training of data managers, physicians and nurses on the exclusive nature of managing chronic related diseases data since they are mostly handled as acute diseases hence leading to errors in capturing them.


Hana, C. O. I. N. G., Cook-huynh, M., Ansong, D., Steckelberg, R. C., Boakye, I., Seligman, K., Amuasi, J. H. (2012). And Adults From A Rural It is important to possess baseline information on non-communicable diseases, such, 22, 347–352.


APPENDICES

APPENDIX 1

Data Abstraction Tool for estimation of Completeness of Data

<table>
<thead>
<tr>
<th>Months</th>
<th>Total cases of hypertension counted in CRR (a)</th>
<th>Number of new hypertension cases recorded into DHIMS-2 (b)</th>
<th>% completeness of cases (b/a * 100%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>July</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>August</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>September</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX 2

Data Abstraction Tool for estimating the Error rate (Accuracy)

<table>
<thead>
<tr>
<th>Month</th>
<th>Primary data as counted (folder)</th>
<th>Secondary data as counted (CRR)</th>
<th>Facility aggregate data</th>
<th>DHIMS-2 data</th>
<th>Total Data Inspected</th>
<th>Error rate-1 %</th>
<th>Error rate-2 %</th>
<th>Error rate-3 %</th>
<th>Error rate-4 %</th>
</tr>
</thead>
<tbody>
<tr>
<td>July</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>August</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>September</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX 3
Semi-structured Interview guide on hypertension data management.

Health facility data
1. Facility ownership: Private □ Public □
2. Facility type: Primary □ Secondary □ Tertiary □
3. Facility location: Urban □ Peri-Urban □ Rural □

Health staff background data
4. Cadre of staff: ………………………
5. Department/Unit of staff: OPD □ Records □ Biostatistics □
6. Number of years worked: (2-4) □ (5-7) □ (8-10) □

Questions
7. What specific role do you perform in hypertension documentation?
8. Can you describe the training you received before that role?
9. Are you regularly supervised or given supportive supervision? How often?
10. Briefly describe how you manage hypertension data at your unit.
11. What kind of decision is made with hypertension data?
12. Kindly describe any means of data verification to ensure accuracy and completeness of the data?
13. In your own view, do you think the quality of hypertension data is of priority to facility managers or other higher entities?
14. What do you think are challenges faced with managing hypertension data?
APPENDIX 4

GHANA HEALTH SERVICE ETHICS REVIEW COMMITTEE

In case of reply the number and date of this Letter should be quoted

My Ref. GHS/RDD/ERC/Admin/App  Your Ref. No.

Emmanuel Tetteh Doku
University of Ghana
School of Public Health
Legon, Accra

The Ghana Health Service Ethics Review Committee has reviewed and given approval for the implementation of your Study Protocol.

<table>
<thead>
<tr>
<th>GHS-ERC Number</th>
<th>GHS-ERC: 154/12/17</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project Title</td>
<td>Accuracy and Completeness of Hypertension Data in DHIMS 2 within Health Facilities in Greater Accra Region</td>
</tr>
<tr>
<td>Approval Date</td>
<td>19th December, 2017</td>
</tr>
<tr>
<td>Expiry Date</td>
<td>18th December, 2017</td>
</tr>
<tr>
<td>GHS-ERC Decision</td>
<td>Approved</td>
</tr>
</tbody>
</table>

This approval requires the following from the Principal Investigator

- Submission of yearly progress report of the study to the Ethics Review Committee (ERC)
- Renewal of ethical approval if the study lasts for more than 12 months
- Reporting of all serious adverse events related to this study to the ERC within three days verbally and seven days in writing.
- Submission of a final report after completion of the study
- Informing ERC if study cannot be implemented or is discontinued and reasons why
- Informing the ERC and your sponsor (where applicable) before any publication of the research findings.

Please note that any modification of the study without ERC approval of the amendment is invalid.

The ERC may observe or cause to be observed procedures and records of the study during and after implementation.

Kindly quote the protocol identification number in all future correspondence in relation to this approved protocol

SIGNED: ......................................................

DR. CYNTHIA BANNERMAN
(GHS-ERC CHAIRPERSON)

Cc: The Director, Research & Development Division, Ghana Health Service, Accra