FACTORS AFFECTING DATA QUALITY IN UGANDA HEALTH MARKETING GROUP SUPPORTED PRIVATE CLINICS IN KAMPALA DISTRICT

BY

MILTON BARYAMUREEBA

JAN15/PM&E/0366U

DEPARTMENTOF PUBLIC ADMINISTRATION,MANAGEMENT AND POLICY:
SCHOOL OF BUSINESS AND MANAGEMENT

SUPERVISOR

DR.WILLY NGAKA

A PROPOSAL SUBMITTED TO THE DEPARTMENT OF PUBLIC
ADMINISTRATION, MANAGEMENT AND POLICY: SCHOOL OF BUSINESS AND
MANAGEMENT
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS   FOR THE AWARD OF
THE DEGREE OF MASTERS IN PROJECT MONITORING   AND EVALUATION OF
UGANDA TECHNOLOGY AND MANAGEMENT UNIVERSITY (UTAMU)

APRIL 2015

DECLARATION

I hereby declare that this proposal is my own and has not been submitted for any degree or examination in any other university. In all cases where other people's ideas were used, they have been duly acknowledged by complete references.
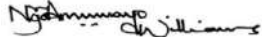
Full name of Student: **MILTON BARYAMUREEBA**

Signature…………………………………………… *Date*……………………………………………

APPROVAL

This proposal has been submitted with my approval.

Supervisor. **DR.WILLY NGAKA**

April 15, 2015

Signature…………………………………….. Date…………………………………..

TABLE OF CONTENTS

CHAPTER ONE

INTRODUCTION

1.1 Introduction

This study investigates Factors affecting Data quality in private clinics supported by Uganda Health Marketing Group (UHMG) whereby it takes the data quality as the dependent variable (DV) and the factors affecting data quality as the Independent Variables (IV). Data quality is a perception or an assessment of data's fitness to serve its purpose in a given context. This chapter presents the background of the study, problem statement, purpose of the study, specific objectives, research questions, hypothesis, scope of the study, conceptual framework as well as the operational definitions.

1.2 Background of the study

1.2.1 Historical background

Before the rise of the inexpensive server, massive mainframe computers were used to maintain name and address data so that mail could be properly routed to its destination. The mainframes used business rules to correct common misspellings and typographical errors in name and address data, as well as to track customers who had moved, died, gone to prison, married, divorced, or experienced other life-changing events (Olson, 2003). Government agencies began to make postal data available to a few service companies to cross-reference customer data with the National Change of Address registry (NCOA). This technology saved large companies millions of dollars in comparison to manually correction of customer data. Large companies saved on postage, as bills and direct marketing materials made their way to the intended

customer more accurately. Initially sold as a service, data quality moved inside the walls of corporations, as low-cost and powerful server technology became available (Olson, 2003).

In 1964, Zero Defects (or ZD) was a data management-led program to eliminate defects in industrial production data that enjoyed brief popularity in American industry from 1964 to the early 1970s (Halpin, 1966). Quality expert Philip Crosby later incorporated it into his "Absolutes of Quality Management" and it enjoyed a renaissance in the American automobile industry—as a performance goal more than as a program in the 1990s. Although applicable to any type of enterprise, it has been primarily adopted within supply chains wherever large volumes of components are being purchased (common items such as nuts and bolts are good examples).

In 1990's, most of the companies all over the world begun to set up data governance teams whose sole role in the corporation was to be responsible for data quality. In some organizations, this data governance function was established as part of a larger Regulatory Compliance function - a recognition of the importance of Data/Information Quality to organizations because problems with data quality don't only arise from incorrect data; inconsistent data is a problem as well hence necessitated eliminating data shadow systems and centralizing data in a warehouse is one of the initiatives a company can take to ensure data consistency (Olson, 2003).

By the start of the year 2000, enterprises, scientists, and researchers had started to participate within data curation communities to improve the quality of their common data. The market was going some way to providing data quality assurance. A number of vendors made tools for analyzing and repairing poor quality data in situ," service providers cleaned the data on a contract basis and consultants advised on fixing processes or systems to avoid data quality problems in the first place (Redman, 2004). Most data quality tools offer a series of tools for

improving data, which may include data profiling, data standardization, geocoding, matching or Linking, monitoring - keeping track of data quality over time and reporting variations in the quality of data as well as the batch and real time because once the data is initially cleansed (batch), companies often want to build the processes into enterprise applications to keep it clean (Redman, 2004).

This thereafter necessitated the formation of the International Association for Information and Data Quality (IAIDQ) which was established in 2004 to provide a focal point for professionals and researchers in the field of data quality. This was also coupled with the introduction and certification of ISO 8000 which is the international standard for data quality in the whole world (Olson, 2003).

1.2.2 Theoretical background

This study will dwell specifically on the systems theory of data quality (Ivanov, 1972) as well as the general theory of data quality. A systems-theoretical approach influenced by American pragmatism expands the definition of data quality to include information quality, and emphasizes the inclusiveness of the fundamental dimensions of accuracy and precision on the basis of the theory of science (Ivanov, 1972).

The theory states that all things, living and nonliving, could be regarded as systems and that systems have properties that are capable of being studied and can affect the quality of the outcome both in the short and long run. Since the theory defines a system as an organized whole made up of components that interact in a way distinct from their interaction with other entities and which endures over some period of time, this interaction brings about exchange of information and when manipulated effectively leads to a quality outcome or result.

3

Therefore, the systems theory of data quality brings out clearly that data quality is determined by the efficiency in the exchange of information between the system and its environment and this is regulated by a process called feedback, a method of evaluation used to determine whether the system's outputs are consonant with the perceived outcomes (goals) that the system has established for itself (Ivanov, 1972).

The theory is advantageous in pure scientific situations because all aspects of systems iterated by the theory can be carefully controlled for environmental effects in data management (Ivanov, 1972).

The systems theory of data quality however assumed a single dimension cause-and-effect relationship between social units within the environment and also had some difficulty with the single-dimension relationship and felt that systems theory did not fully capture the complex dynamics that occur within social systems.

The General Theory of Data Quality maintains that true data quality standards are enterprise-wide standards providing an objective data foundation. True information quality standards must always be customized to meet the subjective needs of a specific business process and/or initiative. Both aspects of this shared perspective of quality must not only be incorporated into a single sustained program that enforces a consistent enterprise understanding of data, but that also provides the information necessary to support day-to-day operations.

General Theory of Data Quality tenants that data duality is improved by measuring and modifying the Process, not sifting the Output to identify failures that need to be reworked or thrown away. The theory emphasizes the same concept by distinguishing two types of analysis through enumeration which is the act of classifying data then counting statistically analyzing

outcome data. The only short coming with the General Theory of Data Quality is that too often quality failures are not traced to their real cause in the process.

1.2.3 Conceptual background

Data Quality is a process required for the integrity of the data management by covering gaps of data issues (Juran, 2010). Data are of high quality if, "they are fit for their intended uses in operations, decision making and planning." (Juran, 2010). Alternatively, data are deemed of high quality if they correctly represent the real-world construct to which they refer.

Data quality is multidimensional, and involves data management, modeling and analysis, quality control and assurance, storage and presentation. As independently stated by Chrisman (1991) and Strong et al. (1997), data quality is related to use and cannot be assessed independently of the user.

Data is defined as distinct pieces of information, usually formatted in a special way (Akash, 2011). Data can exist in a variety of forms such as numbers or text on pieces of paper, as bits and bytes stored in electronic memory, or as facts stored in a person's mind.

Beynon-Davies (2009) defines data as a set of values of qualitative or quantitative variables; restated, pieces of data are individual pieces of information.

Private clinics refer to clinics owned by private individuals set up with the main purpose of making profits. A private clinic is a health care facility that is primarily devoted to the care of outpatients through primary health care needs of populations in local communities but in its total management is in private ownership.

1.2.4 Contextual background

Globally, reliable and accurate public health information is essential for monitoring health and for evaluating and improving the delivery of health-care services and programmes (AbouZahr, 2005). As countries report their progress towards achieving the United Nations Millennium Development Goals, the need for high-quality data has been neglected. Furthermore, funding and support for public health activities, such as immunization programmes, remain contingent on demonstrating coverage using routine statistics (Doyle, 2009). However, assuring the quality of health information systems remains a challenge.

In Africa, studies of public health information systems frequently document problems with data quality, such as incomplete records and untimely reporting (Makombe, 2008). Yet these systems are often the only data sources available for the continuous, routine monitoring of health programmes. Efforts have been made to improve the quality and management of public health information systems in developing countries. Two examples are the Health Metrics Network, an international network that seeks to improve the quality of health information from various sources, and the Performance of Routine Information System Management (PRISM) framework, which was developed as a method for assessing the strengths and weaknesses of routine health information systems (Hotchkiss, 2010). Other initiatives, such as the Data Quality Audit, have been used by the GAVI Alliance to improve the monitoring of immunization coverage (Doyle, 2010). However, the complex nature of health information systems and the demands placed upon them have complicated efforts to improve the quality of routine data (Barron, 2010).

Studies done in Kenya on PMTCT programme showed that one unexpected complication that arose during the study could have reduced the effect of the data improvement intervention. The PMTCT programme in Kenya is relatively dynamic and the names and definitions of the data

elements used for monitoring are frequently changed (Kimaro, 2005). Several challenging changes occurred during the study. For example, the data element used in the DHIS to record whether a baby had undergone a polymerase chain reaction test for HIV at 6 weeks was initially titled "HIV 1st test of baby born to HIV-positive woman" but was later changed to "HIV PCR test of baby born to HIV-positive woman at 6 weeks or later". Such changes were made without the district offices providing definitions to the clinics. This could have caused considerable confusion at individual facilities and compromised the quality of reporting on that particular data element (Kimaro, 2005).

Despite these limitations, the improvement in PMTCT data quality observed in this study is encouraging, for it suggests that similar approaches could improve the quality of the data needed for decision-making and resource allocation in other public health programmes (Kimaro, 2005). The rationalization of data collection tools, clear definitions of data elements, continuous feedback on data quality and intermittent but regular data audits are effective ways of improving data quality. However, while this study shows that public health information can be improved, the final result falls short of what we should accept from our health information systems.

In hospitals in Uganda, health care data collected provide government authorities like ministry of health with information required to not only review the services of all hospitals under their control, but also to plan for the future. In addition, the use of a disease classification system at primary health care level enables the government to collect data on the health status of the community and provide detailed national health statistics. In some countries, the ministry of health determines whether hospitals are required to supply information only on the main conditions or on all diagnoses treated and procedures performed (Kwesiga, 2001).

For most private clinics in Uganda, many clinicians assume that the data contained and portrayed in their health systems is absolute and error free or that the errors are not important. But error and uncertainty are inherent in all data, and all errors affect the final uses that the data may be put to. Clinics and most health units don't take time to examine the information quality chain responsible for species-occurrence data and their documentation is not consistent to data management principles, thus making it hard for them to be able to know and understand the data and determine their "fitness for use" (Kwesiga, 2001).

Most clinics rush to submit forged data sets upon request and this normally contains acute problems traceable right from entry to conversion. In addition to forging data sets, most of the clinics avail row data in form of health reports which are sometimes written in ink and these data sets are very hard to integrate in case they are needed to provide some meaningful information on health issues in such clinics or health centres. This therefore in addition to threatening patient safety, poor data quality increases healthcare costs and inhibits health information exchange, research, and performance measurement initiatives (Ministry of Health Report, 2006).

Worse still, some of the clinics have a tendency of waiting for the time periods when this information is needed and normally, compilation of data sets begin one or two months towards the dates when they know that officials from UHMG or Ministry of Health will come in collecting this data. This implies that such data sets have loopholes given that it has not fully represented the time period in which they are supposed to be compiled. This therefore leaves a lot to be desired given the urgency and the need for the data sets as they are needed to address public health concerns in certain regions.

1.3 Problem statement

Healthcare data and its transformation into meaningful information is a central concern for consumers, healthcare providers, the healthcare industry, and the government. Standards, technologies, education, and research are required to capture, use, and maintain accurate healthcare data and facilitate the transition from paper to electronic systems in order to effectively formulate policies regarding health especially in public domain. It is on this note that UHMG supports private clinics through training, mentoring and provision of data gathering tools to so that they can collect, analyse, and report to Ministry of Health through the District Information System and then to UHMG. Despite all these efforts, data from these clinics is usually inaccurate, late, incomplete and even getting these reports is a struggle. This data therefore makes it hard for the stakeholders to use it to make informed decisions and therefore find it hard to improve programme performance.

The above therefore depicts that the essentials of data management especially the clinical coding procedure are often neglected issues with health clinics databases and too often, health related data are used uncritically without consideration of the error contained within, and this can lead to erroneous results, misleading information, unwise decisions and increased costs. The study therefore intends to establish the different factors that affect the data quality in the private health sector.

1.4 Purpose of the study

The purpose of the study will be to establish the factors affecting Data Quality in Uganda Health Marketing Group (UHMG) Supported private clinics in Kampala District.

1.5 Specific objectives

This study will be guided by the following objectives

i. To examine the effect of internal factors on Data Quality in UHMG Supported private clinics.

ii. To find out how external factors affect Data Quality in UHMG Supported private clinics.

iii. To find out how to improve data quality in UHMG Supported private clinics in Kampala.

1.6 Research questions

This study seeks to answer the following questions

i. How do internal factors affect Data Quality in UHMG Supported private clinics?

ii. What is the effect of external factors on Data Quality in UHMG Supported private clinics?

iii. How can data quality be improved in UHMG Supported private clinics in Kampala?

1.7 Hypothesis

i. Internal factors affect Data Quality in the private clinics

ii. There is a relationship between external factors and Data Quality in the private clinics

1.8 Conceptual framework

FACTORS AFFECTING DATA QUALITY (IV)   **DATA QUALITY (DV)**

Internal factors

- Accuracy
- Reliability
- Timeliness
- Completeness
- Consistency
- Precision

External factors(Use+Value)

- Relevance
- Content
- Importance
- Sufficiency
- Understandability
- Freedom from bias
- Interpretability

- Timely
- Accurate
- Reliable
- Complete
- Consistent
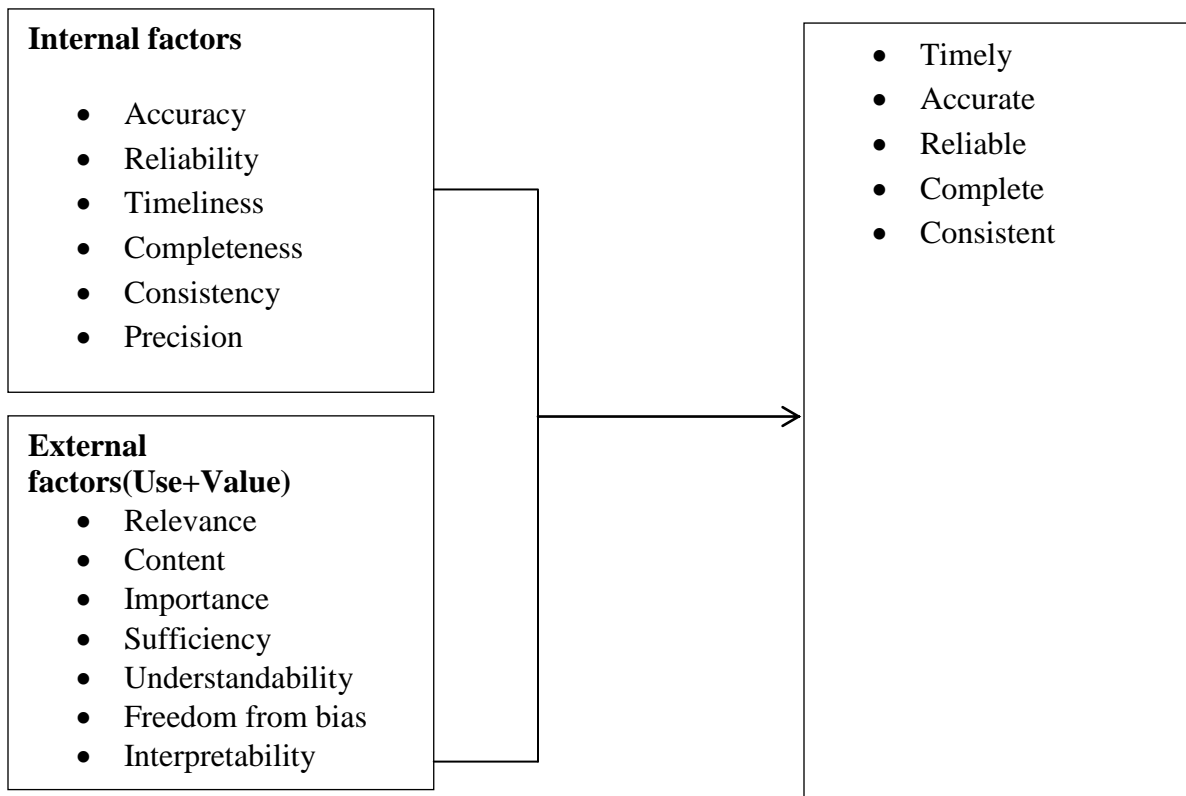
Source: Adopted with Modification from Wang & Storey, (1996).

The conceptual framework depicts the effects internal factors and external factors have on data quality. Internal factors are data related in terms of accuracy, reliability, timeliness, completeness, consistency and precision and system related. External factors are measured in terms of use and value.

1.9 Scope of the study

1.9.1 Content scope

The study will focus on factors like internal factors and external factors viz-a-viz data quality in private owned clinics supported by UHMG. Internal factors and external factors are independent variables while data quality is the dependent variable.

1.9.2 Geographical scope

The study will be conducted in the private clinics supported by UHMG in Kampala district. There are 30 private clinics supported by UHMG in Kampala.

1.9.3 Time scope

The study will focus on a time frame from 2012 to 2014. This will be a reference period for the study. This will be a guidance period which will give a clear picture on how different factors have affected data quality taking into consideration that this is the period when UHMG increased its interventions in private clinics in Kampala.

1.10 Justification of the Study

Generally, data quality in the Health sector in Uganda has remained a very big issue and the private health sector has not been spared. UHMG supports private clinics training, Mentoring and provision data gathering tools to private clinics so that they can collect, analyse, and report of Ministry Of Health through the District Information System and then to UHMG but despite all these efforts, data from these clinics is usually inaccurate, late, incomplete and even getting it these reports is a struggle. This data therefore makes it hard for the stakeholders to use it make informed decisions and therefore improve programme performance.

Furthermore, it should be noted that UHMG is funded by USAID and donor money must be accounted for through quality data with the highest level of preciseness and failure to get quality data would mean disaster to the organization since funding can easily be stopped by the funder. It is against this background that the researcher wants to find out the where the problem lies and what can be done to solve the problems.

1.11 Significance of the Study

The findings may benefit the Ministry of Health as it will highlight gaps within data management which need to be closed in order to receive correct and accurate data from private clinics for improved health service delivery for all the Ugandans using private clinics.

UHMG will use the findings of this study to come up with strategies to bridge the gaps that will be identified Findings and recommendations from the study are envisaged to be useful in improving the quality of data in UHMG supported clinics and the general private Health sector services in and outside the community/ area of study.

The study findings could add new concepts and knowledge to the existing body of knowledge of Data quality. The study findings are expected to provide up to date literature to the academicians who may wish to carry out similar or related study. The study findings are hoped to stimulate further research in data quality issues.

1.12 Operational Definitions

For purposes of this study, the concepts below are defined as assigned thereof not necessarily reflecting their ordinary or dictionary meanings. These are:

Data: This refers to raw observations or acts and statistics collected together for reference or analysis.

Data Quality: Data are of high quality if, "they are fit for their intended uses in operations, decision making and planning." (Juran, 2009). Alternatively, data are deemed of high quality if they correctly represent the real-world construct to which they refer.

Private clinics: These are clinics owned by private individuals set up with the main purpose making profits.

Data cleaning refers to the process of "fixing" errors in the data that have been identified during the validation process.

Validation is a process used to determine if data are inaccurate, incomplete, or unreasonable. The process may include format checks, completeness checks, reasonableness checks, limit checks, review of the data to identify outliers (geographic, statistical, temporal or environmental) or other errors, and assessment of data by subject area experts.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

This chapter presents the scholarly material regarding the study. Theories regarding data quality will also be reviewed and it is arranged according to the study objectives. The gaps identified in the literature review are also indicated.

2.2 Theoretical Review

This study will dwell specifically on the systems theory of data quality (Ivanov, 1972) as well as the general theory of data quality.

The theory states that all things, living and nonliving, could be regarded as systems and that systems that have properties that are capable of being studied can affect the quality of the outcome both in the short and long run. Since the theory defines a system as an organized whole made up of components that interact in a way distinct from their interaction with other entities and which endures over some period of time, this interaction brings about exchange of information and when manipulated effectively leads to a quality outcome or result.

Therefore, the systems theory of data quality brings out clearly that data quality is determined by the efficiency in the exchange of information between the system and its environment and this is regulated by a process called feedback, a method of evaluation used to determine whether the system's outputs are consonant with the perceived outcomes (goals) that the system has established for itself. In addition to this internal feedback, the system also has a method of

measuring responses from the external environment. In both situations, if the system perceives a variance between output and outcome, it can alter the process by varying the level of inputs (Ivanov, 1972).

. The theory is advantageous in pure scientific situations because all aspects of systems iterated by the theory can be carefully controlled for environmental effects in data management (Ivanov, 1972). The systems theory of data quality however assumes a single dimension cause-and-effect relationship between social units within the environment and also has some difficulty with the single-dimension relationship and feels that systems theory doesn't not fully capture the complex dynamics that occur within social systems. The theory is advantageous in pure scientific situations because all aspects of systems iterated by the theory can be carefully controlled for environmental effects in data management (Ivanov, 1972).

The General Theory of Data Quality maintains that true data quality standards are enterprise-wide standards providing an objective data foundation. True information quality standards must always be customized to meet the subjective needs of a specific business process and/or initiative. Both aspects of this shared perspective of quality must be incorporated into a single sustained program that enforces a consistent enterprise understanding of data, but that also provides the information necessary to support day-to-day operations.

General Theory of Data Quality tenants that data duality is improved by measuring and modifying the Process, not sifting the Output to identify failures that need to be reworked or thrown away. The theory emphasizes the same concept by distinguishing two types of analysis through enumeration which is the act of classifying data then counting statistically analyzing

outcome data. The only short coming with the General Theory of Data Quality is that too often quality failures are not traced to their real cause in the process.

2.2 Internal factors and Data Quality

Clinical coding procedure is used to ensure data quality in health records. Whether undertaking simple coding for primary health care services or for more sophisticated hospital health care services, a thorough knowledge of the classification systems' key components of data quality – accuracy, validity, reliability, completeness and timeliness are important (Hyde, 1992). No study however has been done in Uganda to this effect regarding UHMG supported clinics, a gap this study intends to fill.

A randomized study of 60 selected patient records with 1,891 notes from the Veterans Health Administration's computerized patient record system found that 84 percent of notes contained at least one documentation error, with an average of 7.8 documentation mistakes per patient (Weir, 2003).

Processes that manipulate the data inside the health care databases affect the data quality. Some of these processes are routine, while others are brought upon by periodic system upgrades, mass data updates, database redesign, and a variety of ad-hoc activities. Unfortunately, in practice most of these health procedures lack time and resources, as well as reliable meta data necessary to understand all data quality implications. It is not surprising, then, that internal data processing often leads to numerous data problems which reduce data quality (Arkady, 2007).

There are aspects in health data management that cause accurate data to become inaccurate over time, without any physical changes made to it (Abdelhak, Grostick & Hankin (2001). The data values are not modified, but their accuracy takes a plunge. This usually happens when the real world object described by the data changes, but the data collection processes do not capture the change. The old data turns obsolete and incorrect.

Jones (2003) argued that good data entry health forms and instructions somewhat mitigate data entry problems. In an ideal fantasy world, data entry is as easy to the user as possible: fields are labeled and organized clearly, data entry repetitions are eliminated, and data is not required when it is not yet available or is already forgotten.

Health data which is continuous reduces data quality. The information must ensure continuity between those caring for the patient today and those who will care for the patient in the weeks or years to come (Taulbee, 2000). Effective health information exchange can reduce or eliminate duplication of diagnostic tests, redundancy of processes to obtain information, and the risk of treatment errors. This leads to higher quality patient care, cost savings, and helps to eliminate duplicative processes.

Data processing is at the heart of all operational systems. It comes in many shapes and forms – from regular transactions triggered by users to end-of-the-year massive calculations and adjustments. In theory, these are repetitive processes that should work "like a clock." In practice there is nothing steady in the world of computer software. The first part of the problem is the change in the programs responsible for regular data processing. Minor changes are as regular as normal use. These are often not adequately tested based on the common misconception that small changes cannot have much impact but they reduce data quality in the long run (Hall, 2004).

Burger (2007) argued that timeliness affects data quality. More and more data is exchanged between the systems through real-time (or near real-time) interfaces. As soon as the data enters one database, it triggers procedures necessary to send transactions to other downstream databases. The advantage is immediate propagation of data to all relevant databases. You can close your eyes and imagine the millions of little data pieces flying from database to database across vast distances with lightning speed, making our lives easier.

Further more, a more subtle problem is when processing is accidentally done at the wrong time. Then the correct program may yield wrong results because the data is not in the state it is supposed to be. A simple example is running the program that calculates weekly compensation before the numbers from the hours tracking system were entered. There, timeliness is a very important aspect of data quality management for better report writing and interpretation (Volmink, 2007).

According to AbouZahr (2005), wrong precision with data sets have affected data quality in many organizations. This is worsened by poor data quality specifications which often do not reflect actual data requirements. As a result, data may be brought in compliance with some theoretical model but remain incorrect for actual use. A limitation to this study is that it was not carried out in Uganda and given the fact that Uganda has different socio-economic status with where the study was done, it leaves a gap that this study intends to fill, thus making the study inevitable at UHMG supported clinics.

In study done by Moyo (2005) in Zimbabwe showed that data reliability considerations consist of whether the record is cohesive in terms of the field contents and whether the information

makes sense or is usable in a real world context. This can be considered at any of the steps in the lifecycle of a record – original source, production of an export, import into another system, downstream processing. A record with good integrity will have data in all appropriate fields and the data will conform to best current practice standards. Data values should be within specified bounds but once it loses this metric, it affects data quality in different dimensions.

The quality of the data is directly proportional to the amount of time spent to analyze and profile the data and uncover the true data content (Hotchkiss, 2010). It should be noted that in most cases, the source data itself is never perfect. Existing erroneous data tends to mutate and spread out during conversion like a virus.

Consistency specifies that two data values drawn from separate data sets must not conflict with each other, although consistency does not necessarily imply correctness. Even more complicated is the notion of consistency with a set of predefined constraints. More formal consistency constraints can be encapsulated as a set of rules that specify consistency relationships between values of attributes, either across a record or message, or along all values of a single attribute. A deviation from consistent data set reduces data quality (Mate & Bennett, 2009).

2.3 External factors and Data quality

According to Arkady (2007), processes that bring data into the database from outside either manually or through various interfaces and health data integration techniques affect health data quality. Some of these incoming data may be incorrect in the first place and simply migrate from one place to another. In other cases, the errors are introduced in the process of data extraction, transformation, or loading. High volumes of the data traffic dramatically magnify these

problems. It is however not clear whether such scenarios are existent within the private clinics serviced by UHMG a gap this study intends to fill.

The Health Records Congress proceedings held in Vancouver revealed that often the desired data regarding certain health issue being investigated may not exist or may not be readily available and this leads to adoption of otherwise data known as "surrogate " data which affects data quality. A valid relationship must exist between the surrogate and the phenomenon it is used to study but, even then, error may creep in because the phenomenon is not being measured directly. Hence such data may lack some relevance in it and its intended purpose (Liu Aimin, 1992).

In most cases, the methods of formatting digital information for transmission, storage, and processing may introduce error in the data (Shaw, 2008). Conversion of scale, projection, changing from raster to vector format, and resolution size of pixels are examples of possible areas for format error. Multiple conversions from one format to another may create a negative effect to data quality similar to making copies of copies on a photo copy machine.

According to Brouwer, (2006), Sufficiency challenges in data management in health centres reduce data quality in many organizations. Wherever possible data is collected, sufficiency systems and records are also created to ensure it is as accurate and complete as possible. The most difficult situation is where there is no documented data which means such data is not sufficient enough to come up with a conclusion about the findings or report and this compromises data quality (Brouwer, (2006).

Redman, (2001) notes that understandable information leads to data quality reporting in many organizations. In order for information to be understood very well, it constitutes a measure of quality in each particular context. To service the widest range of applications, users should be

able to evaluate the fitness for use, or "usability", of data which enables them to understand it better..

According to Berendsohn (2000), there are many data quality principles that apply when dealing with species data and especially with the spatial aspects of those data. These principles are involved at all stages of the data management process. A loss of data quality at any one of these stages reduces the interpretability and uses to which the data can be adequately put. The views by Berendsohn (2000) may be correct but no study has been done in Uganda to this effect which makes this study inevitable.

Poor records documentation affects data quality in health units (Taulbee, 2000). Documentation must support the code assignment for accurate billing for patient care and payment of claims. Documentation will justify the patient's admission status, continued stay, and any therapies, treatments or procedures that are provided. Documentation must be specific and timely in support of accurate claims reporting, appropriate reimbursement, and provider accounts receivable (AR) goals. Inaccurate reporting of data has negative implications to the patient as well as to provider report cards and overall accountable care scores which relate directly to reimbursement.

Government policies also greatly impact on data quality: In the report by the Uganda Ministry of Health; Assessment of the Health Information System in Uganda (MOH, 2007), the key findings and recommendations were on all the six main components of HIS development, namely: HIS resources, core health indicators, data sources, information management, information products or data availability, and dissemination and use. Though core health indicators were found to be well defined and comprehensively captured through both routine facility-based and population-based

data sources, there were severe inadequacies identified in terms of capacity (skill and infrastructure), resources to support data capturing and management, availability and appropriate disaggregation, dissemination and utilization. Lower administrative levels chronically lack adequate capacities to capture data on vital events such as births and deaths that occur in their communities, and yet this is key information for bottom-up planning. In most districts, the challenges of harmonization and streamlining data sources still constrain provision of quality data for planning, implementation and performance evaluation. Investment in infrastructure, such as ICT will facilitate improvement in data flow and management through internet connectivity and automated data-warehousing.

2.4 Improving Data Quality in clinics

Data Profiling as one way in improving data quality is to uncover your data defects through data profiling, sometimes called data archeology, which is the process of analyzing the data for correctness, completeness, uniqueness, consistency, and reasonability. Whether you're collecting ad-hoc data for supporting a management decision, building a past performance database for future business development, or harvesting results from completed projects to innovate and improve business, it all starts with data quality. Put some forethought into your next data collection project or, if you already maintain data in your day-to-day job, consider taking a moment to pause and start building that data dictionary (Taulbee, 2000).

Redman, (2001) noted that there should be efficiency in data Cleansing. After the extent of "dirty data" is known, the easiest place to start the data quality improvement process is by cleansing operational data at the time it is moved into databases where it is used for cross-organizational reporting. However, data cleansing is a labor-intensive, time-consuming, and expensive process,

and cleansing all the data is usually neither cost-justified nor practical. On the other hand, cleansing none of the data is equally unacceptable. It is therefore important to carefully analyze the source data and to classify the data elements as critical, important, or insignificant to the business. Then, concentrate on cleansing all the critical data elements, and as time permits, cleanse as many of the important data elements as practical, leaving the insignificant data elements unchanged. In other words, you do not need to cleanse all the data, and you do not need to do it all at once.

Data defect prevention is how to prevent future "dirty data" from being entered. That begins by identifying the root causes for the data defects. The owners of the operational systems should plan to improve their programs and edit checks, unless the effort is unreasonably high. For example, if the corrective action requires changing the file structure, which means modifying (if not rewriting) most of the programs that access that file, then the cost for such an invasive corrective action on the operational system is probably not justifiable especially if the bad data does not interfere with the operational needs of that system. This type of decision cannot and should not be made by IT alone. Downstream information consumers must negotiate with the data originators about justifying and prioritizing the data quality improvement steps (Arkady (2007).

Metadata administrators are people responsible for loading, linking, managing, and disseminating metadata to facilitate the common understanding of data and to encourage data reuse. Metadata is the contextual information about the data. Metadata components include data names, data definitions, business rules, data content (domains), data type, data length, data owner, data transformations, degree of cleanliness, and so on.

Usage of data quality stewards increases data quality. These people are charged with preventing the propagation of inferior quality data throughout the enterprise, and thus, the decision-making processes. Therefore, it is their responsibility to perform regular data audits on business data, metadata, and data models, and to be involved in data reconciliation efforts by helping to identify and resolve the root causes of data quality issues. The findings of the audits and reconciliation efforts should feed back into a continuous data quality improvement cycle.

According to Moyo (2005), Data quality training should be instituted to address poor data entry habits. Not all data rules can be enforced through edit checks or by the features of relational databases, such as strong data typing, referential integrity, use of look-up tables, and the use of stored edit procedures. Many data violations can still occur because of human error, negligence, or intentionally introduced errors.

2.5 Conclusion

The studies reviewed above stipulated different views on how internal and external as well as organizational factors affected data quality. This was both in health sector and outside the health sector. It should however be noted that most these studies were not done in Uganda but this study is unique only to Uganda. More so, these studies were not done in Uganda Health Marketing Group which also makes it an important step to closing this gap for this particular study.

CHAPTER THREE

METHODOLOGY

3.1 Introduction

This chapter describes the methods that will be followed in conducting the study. It gives details regarding research design to be used, population of the study area, sample size and sampling techniques, a description of data collection instruments to be used, as well as the techniques that will be used to analyze data.

3.2 Research design

The study will adopt a case study design in which a cross-sectional descriptive survey designs will be used by way of methodological triangulation, and it will adopt both qualitative and quantitative approaches. The case study research design is intended to enable the researcher to conduct an intensive and descriptive analysis of a single entity, UHMG, with the hope that the findings will be applicable to other health service providers in matters regarding data quality in health circles.

Cross sectional studies is selected in this study because it emphasizes detailed contextual analysis of a limited number of events or conditions and their relationships. The researcher will employ both quantitative and qualitative research approaches because they complement one another. Using both will help cover more areas while using only one approach may be defective (Barifaijo, Basheka and Oonyu, 2010).

3.3 Study Population

A population is a complete set of elements (persons or objects) that possess some common characteristic defined by the sampling criteria established by the researcher (Banerjee, 2010). In this study, a population of 30 clinics supported by UHMG of which the overall population is 130 people who will be used for the study. The population includes only the staff and management of the respective clinics supported by UHMG as well some officials from the Ministry of Health. These categories of population are chosen because they are presumed to be the type of people who have participated in data management and processing within their clinics and beyond and therefore are in position to give accurate and reliable information about the study.

3.4 Sample size and selection

A sample is simply a subset of the population. Sampling is the process of selecting sufficient numbers of elements from the population so that a study of the sample and its characteristics would make it possible for the researcher to generalize such characteristics to the population elements (Sekaran, 2000). The total sample size of 111 will be selected by use of Krejcie and Morgan sampling determination table.

Table 3.1 Category of respondents

Category of Respondents	Target population	Sample Size	Sampling formula	Sampling technique
Management of the clinics(In Charge)	30	28	Krejcie and Morgan table	Simple Random sampling
Staff	90	73	Krejcie and Morgan table	Purposive sampling
MOH officials	10	10	Krejcie and Morgan table	Purposive sampling
Total	**130**	**111**		

Source: RV. Krejcie and Morgan (1970)

3.4 Sampling techniques and procedure

3.4.1 Purposive sampling

Purposive sampling is one that is selected based on the knowledge of a population and the purpose of the study. The subjects are selected because of some characteristic like knowledge of the subject matter being researched on among others. In this method, the researcher will target a specific group of respondents in the selected group of respondents at the clinics supported by UHMG because they are believed to be reliable and knowledgeable about the topic and so they are in position to give dependable and detailed information about the topic of investigation (Sekaran, 2000). This will be used on the management of the respective clinics supported by UHMG.

3.4.2 Simple random sampling

Simple random sampling will be used to select the staff of the respective clinics supported by UHMG who will participate in the study. Simple random sampling will be done by using pieces of papers written on numbers from 1 to 300 and respondents will randomly pick the papers and whoever picks an odd number will be considered to be part of study until the sample size is attained. This method will allow each member in each of the above categories an equal and independent chance of selection thereby reducing bias (Mugenda and Mugenda, 2003).

3.5 Data Collection Methods

Data for this study will be derived from both primary and secondary sources. To investigate the variables of the study exhaustively, the researcher will use a combination of data collection methods by way of methodological triangulation. This will be done to enable the various methods to complement one another, thereby making up for the weakness in each method. As a

result, the researcher will be able to capture a more comprehensive variety of information, to reveal more discrepancies in the data collected and to eliminate more biases than will have been possible if the researcher had a single method (Mugenda and Mugenda, 2003). Primary data will be obtained using two methods: the questionnaire survey method and interviews. Secondary data will be obtained by means of documentary review.

3.5.1 Questionnaire survey

The selection of the questionnaire survey method will be guided by the nature of data to be collected, the time available and the objectives of the study (Touliatos and Compton, 1988). This method will be used on all respondents who will be selected to participate in this study and whose particulars appear in table 3.1 above. One of the reasons why this method is preferred is because the study involves variables that cannot not be observed and can only be derived from respondents' views, opinions and feelings (Touliatos & Compton, 1988).

3.5.2 Interviews

The interview method will be used on some few respondents in order to supplement the data obtained from the questionnaires. The sixteen will be administrators/management of UHMG supported clinics. The reason why the interview method is preferred for these respondents because the researcher is intending to capture in-depth, accurate and sensitive information which cannot not be obtained using the questionnaire method (Mugenda and Mugenda, 2003).

3.5.3 Documentary review

Document analysis will be used in studying the already existing literature and documents in order to either find gaps that can be filled by the study or evidence that can support or contradict the quantitative and quantitative findings. To exhaustively investigate the study the researcher

will use triangulation to capture a variety of information, reveal discrepancies that a single technique may not reveal (Mugenda and Mugenda, 2003).

3.6 Data collection instruments

The study will use three research instruments to collect primary data: a questionnaire, an interview guide and documentary review checklist. The selection of these instruments will be guided by the nature of data to be collected, the time available and the objectives of the study (Touliatos & Compton, 1988). Besides, the two instruments are effective and popularly used tools for collecting data in research surveys (Kothari, 2003).

3.6.1 Questionnaires

Self -administered questionnaires with structured questions will be used in data collection. The instrument is adopted because it is time saving since it enables respondents to freely tick their opinions from predetermined ideas and many respondents can fill a questionnaire at the same time and in the absence of the researcher. Moreover, because the study requires some confidentiality, respondents can freely participate without fear because they can complete the questionnaires in total privacy or with minimum guidance. As a result, this method will generate in-depth and relatively reliable data from a large number of individuals in a relatively short time and at minimal financial cost (Mugenda and Mugenda, 2003).

3.6.2 Interviews Guide

Interviews will be conducted with some few people especially the management and officials from Ministry of Health. An interview guide as a checklist to guide the interviewer will be used in the interview process to ensure uniformity and consistency of the information that will be provided. The interview guide allowed probing for questions in addition to pre-determined

questions so as to elicit detailed and precise data. This will help in digging deep into the issues under investigation (Kothari, 2003).

3.6.3 Document checklist

Document checklist will be used to gather relevant information from secondary sources with an intention of gathering related information about the factors affecting data quality in clinics and health centres. Data will be collected from; reports, Registers and Quality Assurance reports that are done by UHMG Quarterly.

3.7 Validity and Reliability of Instruments

3.7.1 Validity

Validity is the accuracy and meaningfulness of inferences, which are based on the research results (Mugenda, 1999). Validity also refers to the ability to produce findings and information that are in agreement with theoretical or conceptual values (Mugenda, 1999). Validity of instruments will be ascertained by first of all discussing the questionnaire and interview schedule drafts with the supervisor. The content validity of the instrument will be found worthy executing for the pilot run and thus the study. After constructing the questionnaire the researcher will contact the supervisor and three other experts in order to get expertise judgment on the validity. According to Gay (1996) construct validity over an instrument is refined based on expert advice. The following formula will be used to test validity index.

$$CVI = \frac{\text{Number of items regarded relevant by judges}}{\text{Total number of items}}$$

3.7.2 Reliability

Reliability is a measure of the degree to which a research instrument yields consistent results or data after repeated trials (Mugenda, 1999). It is also the ability to produce accurate results. The reliability of instruments will be established basing on the preliminary results derived from the pilot study. The study instruments will be used for a pilot or pre-test basis on selected group of people in some 4 selected clinics in Wakiso district and the results realized will be discussed with the supervisor and the content reliability of the instruments will be found worth using for data collection or not.

3.8 Data analysis

3.8.1 Quantitative data analysis

Data from the questionnaires will be arranged, coded, edited for consistency and easiness and later entered using Statistical Package for Social Scientists (SPSS). The entered data will later be analyzed and the relationship between the factors that affect data quality in UHMG supported clinics using Pearson's correlation coefficients. The correlation coefficient always takes a value between -1 and 1, with 1 or -1 indicating perfect correlation. A positive correlation indicates a positive association between the variables (increasing values in one variable correspond to increasing values in the other variable), while a negative correlation indicates a negative association between the variables (increasing values in one variable correspond to decreasing values in the other variable). A relationship value close to 0 indicates no association between the variables.

Since the method for calculating the correlation coefficient standardizes the variables, changes in scale or units of measurement will not affect its value. For this reason, the correlation coefficient is often more useful than a graphical depiction in determining the strength of the association

between two variables. Furthermore regression analysis using SPSS will be used to analyze how (the extent to which) these factors under investigation affect data quality. Data from questionnaires will be presented in form of frequency tables, pie charts and bar graphs.

3.8.2 Qualitative data analysis

Regarding qualitative data, the different answers from the respective respondents will be categorized into common responses. Qualitative data will be descriptive and obtained from interviews, open ended questions. This data will be presented in accordance with the objectives of the study and will help to substantiate findings from quantitative data. Some themes and appropriate response from interview will be stated to support the quantitative findings in form of direct quotations from the respondents as noted by (Kothari, 2003).

3.9 Measurement of Variables

According to Bell (1997), different variables can be measured at different levels. The researcher will use the nominal scale of measurement which applies to some common set of characteristics such as sex, age, level of education, category of respondent among others. Numbers will be assigned only for purposes of identification but not for comparison of variables. The ordinal measurement will be used to categorize and rank the variables being measured for example the use of statements such as greater than, less than or equal to, Amin (2005). The Likert scale will be used to collect opinion data and this will be used to measure the respondents' beliefs on how the factors under investigation affect data quality in UHMG supported clinics in Uganda using the five scales, that is, 5=Strongly agree; 4=agree; 3=undecided; 2=disagree; 1= strongly disagree.

3.10 Procedure for Data collection

The researcher will obtain an introductory letter from UTAMU after having developed a research proposal under the guidance of the supervisors that will introduce him to the relevant respondents at the respective UHMG supported clinics. The researcher will seek permission from the relevant clinic authorities to allow her conduct a study. The researcher will construct questionnaires for data collection and obtain an accompanying letter to assure the respondents that any information they give will be kept confidential and it is for academic purposes. The researcher will then proceed to administer the questionnaires, conduct interviews with the target population. The questionnaires will be administered by the researcher himself, filled by the respondents and returned to the researcher there and then. While for the interviews, the researcher will take some time and interview some staff face to face and fill in the interview guides.

3.11 Ethical Considerations

Permission to do the study will be sought from UTAMU. The researcher will first seek the consent of the respondents to conduct the study in UHMG supported private clinics. Strict confidentiality will be observed. Names of study participants will not be recorded on questionnaires and interview guides. Filled questionnaires will be kept under lock and key and only the researcher will have access to the keys.

3.12 Anticipated Limitations

Some volunteers to participate in the study might withdraw from a study amidst administering respondents because they have to attend to the patients. This will be rectified by convincing the respondents to be patient and respond to the research instruments.

Some respondents may be busy with attending to patients. This will be solved by informing the clinics in advance such that they are aware about the study and on arrival, they would spare some time to participate in the study.

References

Abdelhak M, Grostick S, Hankin MA, Jacobs E. (2001). *Health Information: Management of a Strategic Resource.*, WB Saunders Company, New Jersey, USA.

Arkady M., (2007). Data Quality Assessment, (edn) Technics Publishers, London, UK.

Amatayakul M. (2000). Improving Coding in Light of New Demands for Data. *Proceedings of the 11th Health Records Congress,* 105–109. Vancouver, Canadian Health Records Association, Canada.

AbouZahr C. (2005). Health information systems: the foundations of public health. *Bull World Health Organ* 2005; 83: 578-83 pmid: 16184276.

Akash M. (2011). "Classifying data for successful modeling". (5th Edition), Palgrave Macmillan Publishers, Basingstoke, UK.

Beynon-Davies,. P. (2002). *Information Systems: An introduction to informatics in organisations*. Palgrave Macmillan, Basingstoke, UK.

Burger E.H, (2007). Errors in the completion of the death notification form. Prentice Hall, Boston, USA.

Chapman, A. D. (2005). *Principles of Data Quality*, version 1.0. Report for the Global Biodiversity Information Facility, Copenhagen. Retrieved from(as of March 2015) http://www2.gbif.org/DataQuality.pdf

Doyle V. (2009). Accuracy and quality of immunization information systems in forty-one low income countries. *Trop Med Int Health* 2009; 14: 2-10 (Edn)McGraw-Hill Publishers, New York, USA.

Halpin, James F. (1966). *Zero Defects: A New Dimension in Quality Assurance*. New York: McGraw-Hill, USA.

Hyde LA. Developing a Quality Management Program for Coded Data. *Journal of AHIMA*, 63/1: 50–52, 1992.

Huang, K.-T., Yang, W.L. and Wang, R.Y. (1999). *Quality Information and Knowledge*. New Jersey: Prentice Hall.

Jones L., (2003) *The Quality of Clinical Documentation and Subsequent Effect on DRG Assignment: A Report on the Findings of the DRG Documentation Study:* Case Mix Development Program. Commonwealth Department of Health, Housing and Community Services, Canberra, Colombia.

Taulbee, S.M. (1996). *Implementing data quality systems in biomedical records* pp. 47-75 in Gad, S.C. and Taulbee, S.M. Handbook of data recording, maintenance, and management for the biomedical sciences. Boca Raton: CRC Press.

Redman, T.C. (2001). *Data Quality: The Field Guide*. Digital Press, Boston, USA.

Liepins, G. E. & Uppuluri, V. R. R. (1990). Data Quality Control: Theory and Pragmatics. Marcel Dekker, Inc., New York, USA.

Liu A,. and Jia-run M,. (2000). The Accuracy and Speed of ICD-9 Coding. *Proceedings of the 11th Health Records Congress,* 119–124. Vancouver, Canadian Health Records Association, USA.

National Centre for Classification in Health. *Australian Coding Standards, Vol. 5, International Statistical Classification of Diseases and related Health Problems, 10th Revision, Australian Modification (ICD-10-AM)*. Sydney, NCCH, 1998.

National Centre for Classification in Health. (2001), *Performance Indicators for Coding Quality (PICQ).* Quality and Education Division, Sydney, Australia.

World Health Organization.(2003). *International Statistical Classification of Diseases and Related Health Problems,10th Revision, Vol. 2.* Geneva, Switzerland.

Juran J.M. (2010). Quality Control Handbook, 6th Edition, McGraw-Hill Publishers, New York, USA.

Kimaro H.C, (2005). Analysing the hindrance to the use of information and technology for improving efficiency of health care delivery system in Tanzania. *Tanzan Health Res Bull*, McGraw-Hill Publishers, New York, USA.

Makombe S.D., (2008). Assessing the quality of data aggregated by antiretroviral treatment clinics in Malawi. *Bull World Health Organ* 2008; 86: 310-4, (edn) New York: McGraw-Hill, USA.

Olivieri, S., Harrison, J. and Busby, J.R. (1995). Data and Information Management and Communication. pp. 607–670 in Heywood, V.H. (ed.) *Global Biodiversity Assessment*. Cambridge University Press, London, UK.

Otwombe K.N, (2007). Improving national data collection systems from voluntary counseling and testing centres in Kenya. Technics Publications, LLC;London, UK.

Shaw V. (2005). Health information system reform in South Africa: developing an essential data set. *Bull World Health Organ* , New York: McGraw-Hill, USA.

Wang, R.Y., Storey, V.C., Firth, C.P., (1995). A frame-work for analysis of data quality research, *IEEE Transactions on Knowledge and Data Engineering* 7: 4, 623-640. Cambridge University Press, UK.

Weir, C.R,. (2003). " Evaluation of Data Input Errors.", (2nd Edition), Prentice Hall Publishers, Boston, USA.

http://www.who.int/healthmetrics/library/countries/hmn_uga_his_2007_en.pdf Retrieved from(as of March 2015)

APPENDIX I: QUESTIONNAIRE FOR CLINIC STAFF

Research Questionnaire to analyze the Factors that affect Data Quality in Private Clinics, a Case Study of UHMG Supported Private Clinics in Kampala district.

Dear Respondent,

I am a student of UTAMU carrying out an academic research for an award of a masters degree in Project Monitoring and Evaluation. Please kindly answer the questions as they relate to you as possible. All data collected will be treated with confidentiality and analyzed for academic purposes.

N.B:

➢ The exercise is purely for academic purposes. Therefore, any information given shall be treated with due confidence.

➢ The researcher will maintain anonymity in quoting specific statements unless permitted otherwise by the person(s) concerned.

Please tick appropriate option in the box provided and the researcher will highly appreciate your responses.

Thank you in advance

Yours faithfully

Milton Baryamureeba

SECTION A: DEMOGRAPHIC CHARACTERISTICS

1. Gender   Male   ☐   Female   ☐

2. Age 18-25 ☐   26-35 ☐   36-45 ☐   40+ ☐

3. Period spent on the job/years

<1   ☐   1 – 5 ☐   6 – 10 ☐   11+ ☐

4. Period of clinic's existence

<1 ☐   1 – 5 ☐   6 – 10 ☐   11+ ☐

5. Level of education

Diploma   ☐   Degree   ☐   Postgraduate ☐

6. How do you rate data quality in your clinic?

Very Good ☐   Good ☐   Poor ☐   Very Poor ☐   Not sure ☐

7. How do you compile clinic data?

 Electronic system ☐ Tools/Registers ☐   Both   ☐

8. Do you take part in the compilation of reports?

Yes ☐   No ☐

9.If Yes, Have you ever received any formal training in Data Management?

Yes ☐   No ☐

SECTION B: DATA QUALITY DMIENSIONS IN UHMG SUPPORTED CLINICS

10.In the table below, indicate your level of agreement on the following items regarding data quality in UHMG supported private clinics in Kampala using the following dimensions.

Data quality dimensions	Response				
Item	Strongly agree	Agree	Disagree	Strongly disagree	Not sure
Our Clinic data is accurate					
Our Clinic data is complete					
Our Clinic data is valid					
Our Clinic data is timely					
Our Clinic data is consistent					
Our Clinic data is current					
Our Clinic data is relevant					

11. In your own view, what other dimensions show that the data collected by your clinic is of good quality?

……………………………………………………………………………………………………

……………………………………………………………………………………………………

……………………………………………………………………………………………………

SECTION C: INTERNAL FACTORS AFFECTING DATA QUALITY

12. In the table below, indicate your level of agreement on the following items regarding the internal factors affecting data quality in UHMG supported private clinics in Kampala using the following dimensions.

Scale: 1= Strongly Agree,   2= Agree,   3=Disagree,   4= Strongly Disagree 5=Not Sure

Internal factors	Response				
Item	Strongly agree	Agree	Disagree	Strongly disagree	Not sure
Quality healthcare data depends on the availability of data itself in the clinics.					
Poor documentation within clinics have adverse effects on the final data set to be compiled					
Most clinics compile inaccurate data which automatically distorts data quality in a long run					
There is insufficient data communication which result in errors and adverse incidents in final data sets					
Cases of poor data collection, sharing, and reporting impacts reduces data quality in private clinics					
Most data lack appropriate content in a usable and accessible form which negatively affect data quality					
Accurate data leads to quality information that is required for quality decision making and patient care.					
Most clinics don't have consistent data models that would ensure the integrity and quality of the data					
Completeness of information					

entered into the clinic record is not dependable					

SECTION D: EXTERNAL FACTORS AFFECTING DATA QUALITY

13. In the table below, indicate your level of agreement on the following items regarding the external factors affecting data quality in UHMG supported private clinics in Kampala using the following dimensions

External factors	Response				
Item	Strongly agree	Agree	Disagree	Strongly disagree	Not sure
Maintaining quality data provided by clinics offers a challenge ensuring the integrity of the healthcare data					
Documentation and data content within clinics is not universally understood by data users, thus affecting its quality					
Monitoring and keeping track of data over time and reporting variations in the data affects quality of data.					
Data compiled by private clinics is not relevant with what users health needs requirements					
Private clinics compile insufficient data compared to what is required by the data users					
Once data sets are hard to interpret, it would become very hard for users to pick any meaningful information out of it, hence compromising data quality					

SECTION E: IMPROVING DATA QUALITY

Data quality dimensions	Response				
Item	Strongly agree	Agree	Disagree	Strongly disagree	Not sure
Uncover the data source/genuine source					
Ensure effective cleaning of clinic					

data.					
Ensuring vigilance about missing data.					
Perform regular reviews of your data to uncover anomalies.					

14. In your own views, how best can we improve on data quality in private clinics in Uganda?

………

………

………

End   Thank you for your cooperation

APPENDIX II: INTERVIEW GUIDE FOR CLINIC INCHARGES

Thank you for receiving me My name is **Baryamureeba Milton** a student of Uganda

Technology And Management University. I'm collecting data for a study on the factors affecting

the quality of data in private clinics supported by UHMG in Kampala district. I have learnt that

you are one of the key stakeholders in Data management from all health facilities be private or

public. For this reason, am interested in getting information from you and I would like you to

share your experiences, views, knowledge, and opinions with me in an open and honest manner.

If you find the information sensitive to discuss you reserve the right not to answer the question or

to quit the discussion at any time. Never the less I want to assure you that the information

obtained from you will be important for this study. The interview will take approximately 45

minutes. Are you willing to proceed?

Thank you

Basic information

a) Date of interview………………………………………

b) Organization of the respondent ………………….. ….

c) Title/ occupation of the respondent………. ………….. Sex……………..

d) How do you collect data in this clinic ……………. …………..

e)How do you use the data collected in this clinic?

f)what challenges do you face in your efforts to produce quality data at this clinic?

g) What can be done improve the quality of data in this clinic?

APPENDIX III: INTERVIEW GUIDE FOR MINISTRY OF HEALTH OFFICIALS
Introduction

Good Morning/Good evening sir.

My name is **Milton Baryamureeba** a student of Uganda Technology and Management University . I'm collecting data for a study on the factors affecting the quality of data in private clinics supported by UHMG in Kampala District.. I have learnt that you are one of the key stakeholders in data management. For this reason, am interested in getting information from you and I would like you to share your experiences, views, knowledge, and opinions with me in an open and honest manner. If you find the information sensitive to discuss you reserve the right not to answer the question or to quit the discussion at any time. Never the less I want to assure you that the information obtained from you will be important for this study.

The interview will take approximately 45 minutes. Are you willing to proceed?

Date of interview……………………………………..

Organization of the Respondent ………………………………..

Title/Occupation of the respondent………………………………….

Age………………………..

Sex……………………………………

- Do you monitor health data collected by clinics?
- If yes, how often do you monitor?
- What do you do to the data provided?
- What kind of support do you offer to private clinics?
- How do you rate the quality of data provided by private clinics?
- Do you have any challenges with the data provided by clinics
- If yes, what are some of the challenges with their data?
- Do clinics meet deadlines to submit their health data?
- What do you recommend to be done in order to improve data compilation in clinics?

Thanks for your cooperation

End

APPENDIX IV: TIME TABLE FOR THE RESEARCH PROJECT

Time frame	Activity
Jan-30th March 2015	Working on the research proposal and its final submission.
1st April-30th - April 2015	Development of Tools
1st May-30th 2015	Data collection
1st June-30th– June 2015	Data entry
1stJuly-30th July 2015	Data analysis
1st August-30th Aug 2015	Reporting writing
1st Sept-30th Oct 2015	Review and Submission of the dissertation
1st Nov-30th Nov 2015	Viva

APPENDIX V: GANTT CHART

ACTIVITY	JAN '2015	FEB '2015	MAR'2015	APR '2015	MAY'2015	JUN '2015	JUL '2015	AUG'2015	SEP '2015	OCT'2015	NOV'2015	DEC'2015
Working on the research proposal and its final submission.	■	■	■									
Development of Tools				■								
Data collection					■							
Data entry						■						
Data analysis							■					
Reporting writing								■				
Review and Submission of the dissertation									■	■		
VIVA											■	

APPENDI VI: SAMPLE SIZE TABLE

N	S	N	S	N	S
10	10	220	140	1200	291
15	14	230	144	1300	297
20	19	240	148	1400	302
25	24	250	152	1500	306
30	28	260	155	1600	310
35	32	270	159	1700	313
40	36	280	162	1800	317
45	40	290	165	1900	320
50	44	300	169	2000	322
55	48	320	175	2200	327
60	52	340	181	2400	331
65	56	360	186	2600	335
70	59	380	191	2800	338
75	63	400	196	3000	341
80	66	420	201	3500	346
85	70	440	205	4000	351
90	73	460	210	4500	354
95	76	480	214	5000	357
100	80	500	217	6000	361
110	86	550	226	7000	364
120	92	600	234	8000	367
130	97	650	242	9000	368
140	103	700	248	10000	370
150	108	750	254	15000	375
160	113	800	260	20000	377
170	118	850	265	30000	379
180	123	900	269	40000	380
190	127	950	274	50000	381
200	132	1000	278	75000	382
210	136	1100	285	1000000	384

Note.—N is population size. S is sample size.

Source: Krejcie & Morgan, 1970