

**FACTORS AFFECTING DATA QUALITY IN PRIVATE CLINICS IN UGANDA: THE
CASE OF SELECTED UGANDA HEALTH MARKETING GROUP-SUPPORTED
CLINICS IN KAMPALA**

BY

MILTON BARYAMUREEBA

JAN15/PM&E/0366U

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

SEPTEMBER, 2015

DECLARATION

I hereby declare that this dissertation is my own original piece of work and has not been submitted for any degree or examination in another University or any other Institution of Higher Learning. In all cases where other people's ideas have been used, they have been duly acknowledged by complete references.

Full name of Student: **MILTON BARYAMUREEBA**

Signature..... Date.....

APPROVAL

This dissertation has been submitted with my approval.

Supervisor: **DR.WILLY NGAKA**

Signature..... Date.....

DEDICATION

This dissertation is dedicated to my family for the tireless efforts they put in to support me. I also dedicate this piece of work to my supervisors and friends.

ACKNOWLEDGEMENTS

Special thanks go to the Almighty God who has given me the life and strength to accomplish this academic work. My sincere gratitude goes to my supervisor, Dr. Willy Ngaka, who has tirelessly corrected me and to ensure that I deliver quality work. Special thanks also go to the staff and management of Uganda Technology and Management University (UTAMU) -- more especially for their e-learning mode of delivery that helped me to accomplish my Masters degree on time.

To all those who stood by me and supported me during my studies and compilation of this report: may God bless you.

TABLE OF CONTENTS

DECLARATION	i
APPROVAL	ii
DEDICATION	iii
ACKNOWLEDGEMENTS	iv
LIST OF FIGURES	ix
LIST OF ABBREVIATIONS	x
ABSTRACT.....	xi
CHAPTER ONE	1
INTRODUCTION	1
1.1 Introduction.....	1
1.2 Background of the study	1
1.2.1 Historical background.....	1
1.2.2 Theoretical background	3
1.2.3 Conceptual background	5
1.2.4 Contextual background	6
1.3 Problem statement.....	9
1.4 Purpose of the study.....	9
1.5 Specific objectives	10
1.6 Research questions.....	10
1.7 Hypothesis.....	10
1.8 Conceptual framework.....	11
1.9 Scope of the study	12
1.9.1 Content scope.....	12
1.9.2 Geographical scope	12
1.9.3 Time scope	12
1.10 Justification of the Study	12
1.11 Significance of the Study	13
1.12 Operational Definitions.....	14
CHAPTER TWO	15
LITERATURE REVIEW	15
2.1 Introduction.....	15

2.2 Theoretical Review	15
2.3 Internal factors and Data Quality	17
2.4 External factors and Data quality	21
2.5 Conclusion	24
CHAPTER THREE	25
METHODOLOGY	25
3.1 Introduction.....	25
3.2 Research design	25
3.3 Study Population.....	26
3.4 Sample size and selection	26
3.4 Sampling techniques and procedure	27
3.4.1 Purposive sampling.....	27
3.4.2 Simple random sampling.....	27
3.5 Data Collection Methods	28
3.5.1 Questionnaire survey	28
3.5.2 Interviews.....	29
3.5.3 Documentary review	29
3.6 Data collection instruments.....	29
3.6.1 Questionnaires.....	29
3.6.2 Interviews Guide.....	30
3.6.3 Documentary Review Guide	30
3.7 Validity and Reliability of Instruments.....	30
3.7.1 Validity	30
3.7.2 Reliability	31
3.8 Data analysis	32
3.8.1 Quantitative data analysis	32
3.8.2 Qualitative data analysis.....	33
3.9 Measurement of Variables	33
3.10 Procedure for Data collection	34
3.11 Ethical Considerations	34
CHAPTER FOUR.....	35

PRESENTATION, ANALYSIS AND INTERPRETATION OF RESULTS	35
4.0 Introduction.....	35
4.1 Response Rate:.....	35
4.2 Demographic characteristics	36
4.3 Data quality in UHMG-supported clinics	41
4.4 Internal factors affecting data quality in UHMG Supported private clinics	44
4.5 External factors affecting data quality in UHMG-supported private clinics	49
4.6 Multiple Regression analysis and Model development	52
CHAPTER FIVE	55
SUMMARY, DISCUSSION, CONCLUSIONS AND RECOMMENDATIONS	55
5.1 Introduction.....	55
5.2 Summary of findings.....	55
5.2.1 Internal factors affecting data quality in UHMG-supported private clinics	55
5.2.2 External factors affecting data quality in UHMG-supported private clinics.....	56
5.3 Discussion of findings.....	57
5.3.1 Internal factors affecting data quality in UHMG-supported private clinics	57
5.3.2 External factors affecting data quality in UHMG-supported private clinics.....	60
5.4 Conclusion	62
5.4.1 Internal factors affecting data quality in private clinics	62
5.4.2 External factors affecting data quality in private clinics	62
5.5 Recommendations.....	62
5.6 Contribution of the study	63
5.7 Areas for future research.....	64
APPENDIX I: QUESTIONNAIRE FOR CLINIC STAFF AND IN-CHARGE.....	69
APPENDIX II: INTERVIEW GUIDE FOR CLINIC INCHARGES.....	74
APPENDIX III: INTERVIEW GUIDE FOR MINISTRY OF HEALTH OFFICIALS	75
APPENDIX IV: TIME TABLE FOR THE RESEARCH PROJECT.....	77
APPENDIX V: GANTT CHART	78
APPENDIX VI: SAMPLE SIZE TABLE	79
APPENDIX V: LETTER OF INTRODUCTION.....	80
APPENDIX VI: LANGUAGE EDITOR CERTIFICATE	81

LIST OF TABLES

Table 1: Category of respondents	27
Table 2: Overall Reliability Test.....	32
Table 3: Response Rate.....	35
Table 4: Gender of respondents	36
Table 5: Age of respondents	36
Table 6: Period of clinic existence.....	38
Table 7: Ways of compiling data	39
Table 8: Receiving formal training in data management.....	40
Table 9: Rating data quality in UHMG-supported private clinics	41
Table 10: Data quality dimensions in UHMG-supported clinics.....	42
Table 11: Internal factors affecting data quality in UHMG-supported private clinics	44
Table 12: Correlation between Internal factors and data quality in private clinics	48
Table 13: External factors affecting data quality in UHMG-supported private clinics	49
Table 14: Correlation between External factors and data quality in private clinics	52
Table 15: Regression Analysis between Factors and Data quality in UHMG-supported clinics .	53

LIST OF FIGURES

Figure 1:Period spent on the Job.....	37
Figure 2:Level of Education	38

LIST OF ABBREVIATIONS

PMTCT	Prevention of Mother to Child Transmission
DHIS	District Health Information System
HIV	Human Immunodeficiency Syndrome
UHMG	Uganda Health Marketing Group
MOH	Ministry of Health
ICT	Information and Communications Technology
UTAMU	Uganda Technology and Management University
SPSS	Statistical Package for Social Sciences
CME	Continuous Medical Education
DDIU	Data Demand and Information Use
M&E	Monitoring and Evaluation
IAIDQ	International Association for Information and Data Quality
QA/QI	Quality Assurance and Improvement
IV	Independent Variable
DV	Dependent Variable

ABSTRACT

This study sought to establish the factors that affect data quality in UHMG-supported private clinics in Kampala District. The objectives of the study were: to examine the effect of internal factors on data quality in UHMG-supported private clinics; to find out how external factors affect data quality in UHMG-supported clinics and to find out how data quality can be improved in UHMG-supported private clinics. Specific emphasis was put on the effect of internal factors as well as external factors on data quality in UHMG-supported clinics in Kampala District. The study used a case study research design on a population which entailed the staff of respective clinics composed of administrators, in-charge and data entry staffs, among others. A total of 111 respondents were selected for the study. The researcher used both random and non-random sampling techniques in selecting the samples. The study was guided by a quantitative paradigm, but with substantial complementary qualitative methods. Questionnaires were self-administered which provided sufficient data from the sample selected, and interviews were used in order to get detailed data to complement and triangulate data which was collected using questionnaires. Data from the questionnaires was analyzed quantitatively using Statistical Package for Social Scientists (SPSS) where correlation was used to establish the relationship between the factors and data quality. Data from questionnaires was presented in form of frequency tables and bar graphs. The study findings confirmed that internal and external factors negatively affect data quality in UHMG-supported private clinics. The study recommended that private clinics should purposely invest in data departments that can be in charge of the monitoring and evaluation function, conducting of formal trainings for all staff in data management and joint supervision in quality assurance and improvement (QA/QI) to promote sustainability in private clinics.

CHAPTER ONE

INTRODUCTION

1.1 Introduction

This study investigated factors affecting data quality in private clinics supported by the Uganda Health Marketing Group (UHMG). It considered 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 (NCOA) registry. This technology saved large companies millions of dollars in comparison to manual 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 the 1960s, Zero Defects (or ZD) was a data management-led programme 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 programme 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 the 1990s, most of the companies all over the world began 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 do not only arise from incorrect data; inconsistent data is a problem as well. This has necessitated the elimination of data shadow systems; and centralization of 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, 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 dwelt 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 non-living, 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 the 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 (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 that the systems theory did not fully capture the complex dynamics that occur within social systems (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 not only be incorporated into a single sustained programme that enforces a consistent enterprise understanding of data, but that also provides the information necessary to support day-to-day operations.

The theory maintains 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

shortcoming 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, modelling 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 independent 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 are those 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 Prevention of Mother to Child Transmission (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 District Health Information System (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 the 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 do not 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 raw 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. Hence, 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 they have not fully represented the time period in which they are supposed to be compiled. This therefore leaves a lot to be desired, given the fact that the data sets are urgently 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, 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 the public domain (Wang and Storey, 1996). It is on this note that UHMG supports private clinics through training, mentoring and provision of data gathering tools so that they can collect, analyze, and report to the 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 so as to improve programme performance (UHMG Data Quality Assessment Report, 2015)

The above statement therefore depicts that the essentials of data management, especially the clinical coding procedure, are often neglected issues in health clinics databases and very often, health-related data are used uncritically without consideration of the errors they contain within, which can lead to erroneous results, misleading information, unwise decisions and increased costs. The study therefore intended 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 was to establish the factors affecting Data Quality in private clinics in Uganda with special emphasis on Uganda Health Marketing Group (UHMG)-supported private clinics in Kampala District.

1.5 Specific objectives

This study was 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.

1.6 Research questions

This study sought 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?

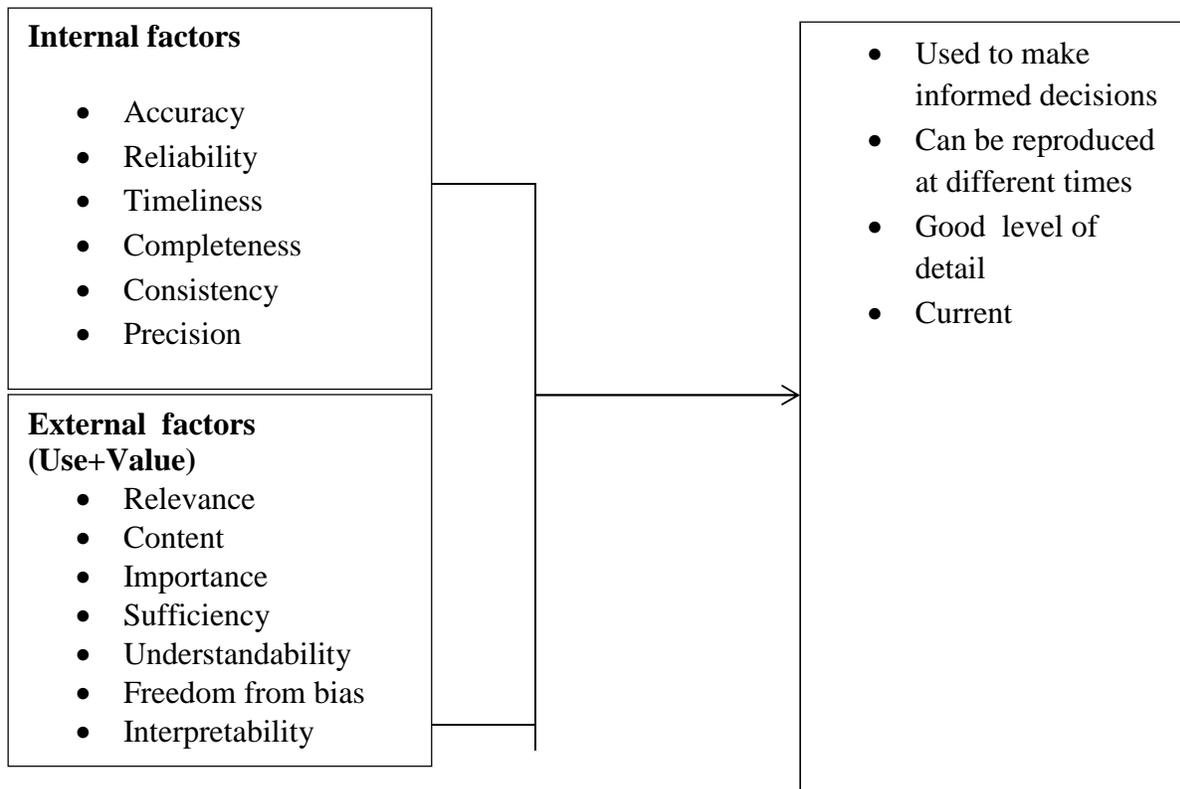
1.7 Hypothesis

- 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)



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 focused on internal and external factors viz-a-viz data quality in privately-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 was 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 focused on a time frame from 2012 to 2014. This was a reference period for the study. This was a guidance period which gave 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 in training, mentoring and provision of data gathering tools so that they can collect, analyze, 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 it these reports is a struggle. This data therefore makes it hard for the stakeholders to use it make informed decisions in order to 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. 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 it.

1.11 Significance of the Study

The findings may benefit the Ministry of Health as they 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 may 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 studies. The study findings should stimulate further research in data quality issues.

The study helped the researcher to be awarded a degree since it is part of the requirements for the award of a Masters degree in Project Monitoring and Evaluation of Uganda Technology and Management University.

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). From another perspective, 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 were also reviewed and arranged according to the study objectives. The gaps identified in the literature review are also indicated.

2.2 Theoretical Review

This study dwelt 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 non-living, 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 argues that systems theory does 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 programme that enforces a consistent enterprise understanding of data, but that also provides the information necessary to support day-to-day operations.

The theory puts it 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 shortcoming with the General Theory of Data Quality is that too often quality failures are not traced to their real cause in the process.

2.3 Internal factors and Data Quality

Clinical coding procedure is used to ensure data quality in health records. Whether undertaking simple coding for primary healthcare 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 per cent of notes contained at least one documentation error, with an average of 7.8 documentation mistakes per patient and this affected its overall quality (Weir, 2003). It should, however, be noted that this particular study was done over 15 years back, which leaves a time gap. Hence this study aimed to fill this gap.

Processes that manipulate the data inside the healthcare 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). The views of Arkady may be correct but no study has been done in UHMG-supported private clinics to ascertain his findings -- a gap this study endeavoured to fill.

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, hence compromising data quality.

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 labelled and organized clearly, data entry repetitions are eliminated, and data is not required when it is not yet available or is already forgotten. The views of Jones seem to be correct but no study has been done in UHMG-supported private clinics to ascertain his findings -- a gap this study endeavoured to fill.

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 programmes 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.

Furthermore, a more subtle problem is when processing is accidentally done at the wrong time. Then the correct programme may yield wrong results because the data is not in the state it is supposed to be. A simple example is running the programme that calculates weekly compensation before the numbers from the hours tracking system are 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 has 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, which made the study at UHMG-supported clinics necessary.

A 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. This study was, however, done in Zimbabwe and not in Uganda, which makes a study in UHMG supported clinics necessary.

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. The views of Hotchkiss may seem convincing but maybe his data sources may have not been so accurate and reliable enough, a gap a study in UHMG intends to fill.

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). Mate and his

colleague may have had good views but these may not apply to Uganda's geographical environment, a gap a study in UHMG endeavoured to fill.

2.4 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 a 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). It is very clear that this study was done over 20 years back and many changes including technological changes have happened ever since. A study that entails the current technological advancement was therefore necessary to close the gap that exists in between these two studies.

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. Shaw's views may seem correct, but these may not apply to Uganda and private clinics in particular -- a gap this study in UHMG endeavoured to fill.

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 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 of Berendsohn (2000) may be correct but no study has been done in Uganda to this effect, which makes this study necessary.

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 titled, 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 capacity 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.

2.5 Conclusion

The studies reviewed above have 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, while this study is specific to Uganda. More so, these studies were not done in Uganda Health Marketing Group which also makes this particular study an important step to close this gap.

CHAPTER THREE

METHODOLOGY

3.1 Introduction

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

3.2 Research design

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

Cross-sectional study was selected in this study because it emphasizes detailed contextual analysis of a limited number of events or conditions and their relationships. The researcher employed both quantitative and qualitative research approaches because they complement one another. Using both helped 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 were used for the study. The clinics are: Zaaam Clinic & Maternity Centre, Suubi Medical Clinic, Vine Medicare, St. Stephen's Dispensary & Maternity Center, YCS Dispensary, St Joseph's Clinic, Span Medicare, Royal Health Care, Rift Valley Medical Services, Mulago II Medical Centre, Mengo Doctors Clinics, Rhona Medical Center LTD, Rick Medical Centre, Martyrs' Family and Maternity Home Ggaba, Medik Medical Centre, Kyanja Community Health Centre, Liberty Medical Centre, Martyrs Clinic & Maternity Centre, Good Will Polyclinic, Equator Medical Services, JB Clinic, Kampala West Medical Clinic, Dembe Medical Centre, Miline Medical Centre, Keen Medical Centre, Mukwaya Hospital, St.Catherine, Alpha Medical Centre - Banda, Church Road Clinic and St.James Domiciliary. The population included 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 were chosen because they were 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). A total sample size of 111 was selected by use of Krejcie and Morgan sampling determination table.

Table 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	Simple Random 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 targeted a specific group of respondents in the selected group of respondents at the clinics supported by UHMG because they were believed to be reliable and knowledgeable about the topic and so they were in position to give dependable and detailed information about the topic of investigation (Sekaran, 2000). This was used on the management of the respective clinics supported by UHMG.

3.4.2 Simple random sampling

Simple random sampling was used to select the staff of the respective clinics supported by UHMG who participated in the study. Simple random sampling was done by using pieces of

paper with numbers from 1 to 300 written on them and respondents randomly picked the papers. Whoever picked an odd number was considered part of study until the sample size was attained. This method allowed 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 was derived from both primary and secondary sources. To investigate the variables of the study exhaustively, the researcher used a combination of data collection methods by way of methodological triangulation. This was done to enable the various methods to complement one another, thereby making up for the weakness in each method. As a result, the researcher was able to capture a more comprehensive variety of information, to reveal more discrepancies in the data collected and to eliminate more bias than would have been possible if the researcher had a single method (Mugenda and Mugenda, 2003). Primary data was obtained using two methods: the questionnaire survey method and interviews. Secondary data was obtained by means of documentary review.

3.5.1 Questionnaire survey

The selection of the questionnaire survey method was guided by the nature of data to be collected, the time available and the objectives of the study (Touliatos and Compton, 1988). This method was used on all respondents who were selected to participate in this study and whose particulars appear in Table 3.1 above. One of the reasons why this method was preferred is because the study involved 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 was used on some few respondents in order to supplement the data obtained from the questionnaires. The sixteen were administrators/management of UHMG supported clinics. The reason why the interview method was preferred for these respondents because the researcher intended to capture in-depth, accurate and sensitive information which could not be obtained using the questionnaire method (Mugenda and Mugenda, 2003).

3.5.3 Documentary review

Document analysis was used in studying the already existing literature and documents in order to either find gaps that could be filled by the study or evidence that could support or contradict the quantitative and quantitative findings. More so, document review was mostly done in reviewing literature. To exhaustively investigate the study, the researcher used triangulation to capture a variety of information, and reveal discrepancies that a single technique might not reveal (Mugenda and Mugenda, 2003).

3.6 Data collection instruments

The study used two research instruments to collect primary data: a questionnaire and an interview guide. The selection of these instruments was 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 were effective and popularly used tools for collecting data in research surveys (Kothari, 2003).

3.6.1 Questionnaires

Self-administered questionnaire with structured questions were used in data collection. The instruments were adopted because they were time-saving since they enabled respondents to

freely tick their opinions from predetermined ideas and many respondents could fill a questionnaire at the same time and in the absence of the researcher. Moreover, because the study required some confidentiality, respondents freely participated without fear because they completed the questionnaires in total privacy or with minimum guidance. As a result, this method generated 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 were conducted with some few people, especially the management and officials from Ministry of Health. An interview guide as a checklist to guide the interviewer was used in the interview process to ensure uniformity and consistency of the information that would be provided. The interview guide allowed probing for questions in addition to pre-determined questions so as to elicit detailed and precise data. This helped in digging deep into the issues under investigation (Kothari, 2003).

3.6.3 Documentary Review Guide

This was a guiding document on what the reader should review, how to select what to review and the key elements to look out for and document while reviewing and how to analyze them.

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 was ascertained by, first of all, discussing the questionnaire and interview schedule

drafts with the supervisor. The content validity of the instrument was found worth executing for the pilot run and thus the study. After constructing the questionnaire, the researcher contacted the supervisor 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 was used to test validity index.

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

$$\text{CVI} = 34/41$$

$$\text{CVI} = 0.829$$

According to Gay (1996), any CVI that is greater than 0.5 means that instrument is valid to be used for any study. Therefore, since CVI was 0.829, it was evident enough that the instrument was valid for executing the study.

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 was established basing on the preliminary results derived from the pilot study based on Cronbach's Alpha Coefficient. The study instruments were 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 were discussed with the supervisor and the content reliability of the instruments found worth using for data collection. Specifically, 10 per cent of the questionnaires were

selected randomly and pre-tested to a few respondents, in order to evaluate data collected, and then any possible amendments were done accordingly.

Table 2: Overall Reliability Test

		N	%
Cases	Valid	12	100.0
	Excluded ^a	0	.0
	Total	12	100.0

a. Listwise deletion based on all variables in the procedure.

Cronbach's Alpha	N of Items
.737	41

The overall reliability of the questionnaire was 0.737 which is above 0.5, meaning that the instrument can provide reliable information once used, as argued by Amin (2001), and that is why it was used in the study.

3.8 Data analysis

3.8.1 Quantitative data analysis

Data from the questionnaires was arranged, coded, edited for consistency and easiness and later entered using Statistical Package for Social Scientists (SPSS). The entered data was later analyzed and the relationship between the factors that affect data quality in UHMG-supported clinics using Pearson's correlation coefficients was established. 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.

Furthermore, regression analysis using SPSS was also used to analyze how (the extent to which) these factors under investigation affect data quality. Data from questionnaires was later presented in form of frequency tables, pie charts and bar graphs for ease of interpretation.

3.8.2 Qualitative data analysis

Regarding qualitative data, the different answers from the respective respondents were categorized into common responses. Qualitative data was descriptive and obtained from interviews and open-ended questions. This data was presented in accordance with the objectives of the study and helped to substantiate findings from quantitative data. Some themes and appropriate response from the interview were 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 used 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 were assigned only for purposes of identification but not for comparison of variables. The ordinal measurement was 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 was used to collect opinion data and this was 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 obtained an introductory letter from UTAMU after having developed a research proposal under the guidance of the supervisors that introduced him to the relevant respondents at the respective UHMG-supported clinics. The researcher sought permission from the relevant clinic authorities who allowed him to conduct a study. The researcher constructed questionnaires for data collection and obtained an accompanying letter to assure the respondents that any information they would give would be kept confidential and it was for academic purposes. The researcher then proceeded to administer the questionnaires, conduct interviews with the target population. The questionnaires were administered by the researcher himself, filled by the respondents and returned to the researcher there and then. While for the interviews, the researcher took some time and interviewed respondents face-to-face and filled in the interview guides.

3.11 Ethical Considerations

Permission to do the study was sought from UTAMU. The researcher sought the consent of the respondents to conduct the study in UHMG-supported private clinics. Strict confidentiality was observed. Pseudo-names of study participants were recorded on questionnaires and interview guides. Filled questionnaires were kept under lock and key and only the researcher had access to the keys.

CHAPTER FOUR

PRESENTATION, ANALYSIS AND INTERPRETATION OF RESULTS

4.0 Introduction

This study sought to establish the factors affecting data quality in private clinics in Uganda, with particular reference to UHMG-supported clinics in Kampala. This chapter presents data analysis and interpretation of study findings on data quality as well as the factors that affect data quality in UHMG-supported clinics in Kampala. The chapter is arranged according to the study findings.

4.1 Response Rate:

The study ensured 100% response rate. This was possible through re-visits to the clinics under study since the personnel in data management positions scheduled appointments to collect data, particularly interviews, whereas others retained the tools, especially questionnaires, and filled them at their convenience. This flexibility enabled the researcher to have ample space to make necessary visits to get the required number of respondents per clinic, that is, four respondents. This response rate indicates that data was collected from a reasonable number of respondents compared to the target population; hence study findings can be relied on according to Mugenda and Mugenda, (1999).

Table 3: Response Rate

Category of Respondents	Target population	Sample Size	Responses	Response rate
Management of the clinics (In Charge)	30	28	28	100%
Clinic Staff	90	73	73	100%
MOH officials	10	10	10	100%
Total	130	111	111	100%

4.2 Demographic characteristics

The demographic information of the employees in terms of age, sex, level of education, period spent on the job, period of clinic's existence, and formal training in data management was obtained. The results are summarized in the tables below:

Table 4: Gender of respondents

Gender	Frequency	Percentage
Male	61	55
Female	50	45
Total	111	100%

Source: Primary data

Findings in the table above show that the majority of the respondents were males (55%) and 45% were females. The findings mean that both males and females in UHMG-supported clinics participated in the study, but more males than the females were in data management positions.

Table 5: Age of respondents

Age	Frequency	Percentage
18-25	6	5.4
26-35	34	30.6
36-45	57	51.4
46+	14	12.6
Total	111	100%

Source: Primary data

Respondents were also asked to mention their age and the majority (51.4%) were aged 36-45, followed by 30.6% who were aged 26-35 while 12.6% and 5.4% were aged 40 and 18-25 respectively. The findings mean that majority of the respondents were mature and were in

position to provide reliable and dependable information about data quality in UHMG-supported clinics.

Period spent on the Job

Respondents were asked to mention the years they had spent on the job as shown in the figure 1 below.

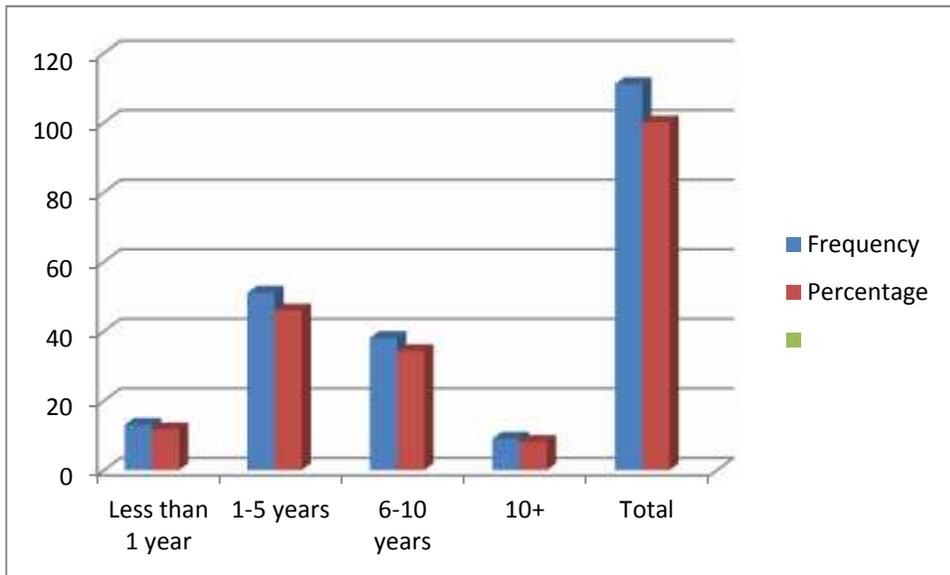


Figure 1: Period spent on the Job

Source: Primary data

Regarding period spent on the job, 45.9% had spent 1-5 years on their jobs whereas 34.2% had spent a period between 6-10 years on their jobs but 11.7% and 8.1% of the respondents had spent less than one year and more than 10 years on their current jobs respectively. The findings mean that the most of the respondents had spent a considerable time on their jobs and were in better position to offer the required information regarding data management in UHMG-supported clinics.

Table 6: Period of clinic existence

Period	Frequency	Percentage
1-5	13	11.7
6-10	31	27.9
10+	67	60.4
Total	111	100%

Source: Primary data

Respondents were also asked to mention the period of their clinic’s existence and 60.4% revealed that their clinics had spent more than 10 years of existence, whereas 27.9% said that their clinics had spent between 6-10 years of existence and only 1.7% noted that their clinics had only been in existence for a period of 1-5 years. The findings meant that the clinics had operated for a considerably long period and therefore had much information regarding data management as well as factors which affect data quality in UHMG-supported clinics.

Level of education

The study asked respondents to reveal their respective levels of education and Figure 2 shows the details.

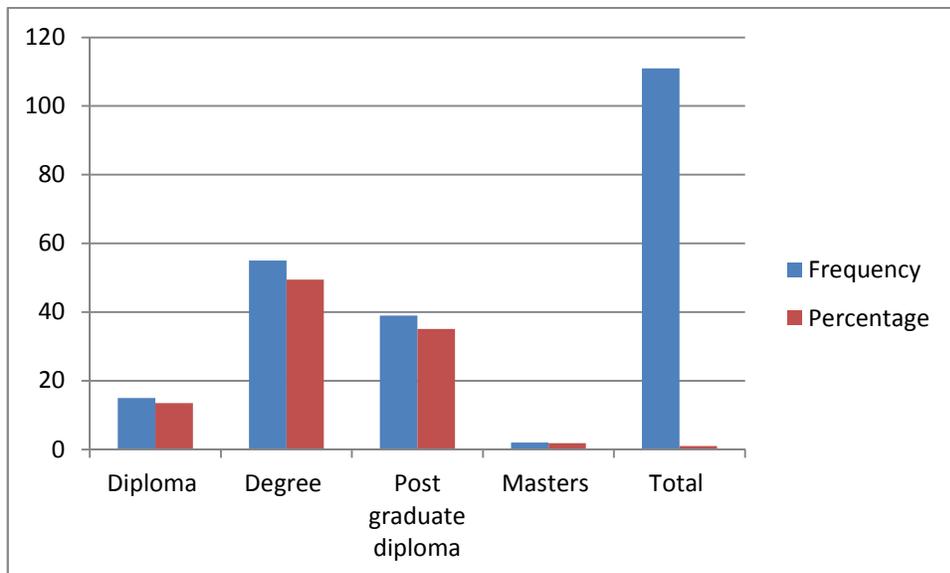


Figure 2: Level of Education

Source: Primary data

On the level of education of the respondents, majority (49.5%) noted that they had attained degrees, followed by 35.1% who revealed they had post-graduate diplomas while 13.5% and 1.8% had attained diplomas and masters degrees respectively. The findings mean that the respondents had the minimum education level which was vital in the interpretation of the data collection tools, especially the questionnaire.

Table 7: Ways of compiling data

Level	Frequency	Percentage
Electronic system	12	10.8
Registers /tools	67	60.4
Both	32	28.8
Total	111	100%

Source: Primary data

Respondents were also asked to mention how they compiled data in their respective clinics and 60.4% noted that they used registers; whereas 28.8% said that they used both registers and electronic systems to compile data; and only 10.8% noted that they used electronic systems to compile their data. The findings therefore meant that most private clinics still used paper registers to compile patient data, a factor which compromised the safety of the data in most private clinics. The findings further implied that most private clinics have not yet adopted the use of electronic or computerized data compilation which is far more efficient than manual registers. The findings are in line with Fischer (2014) who noted that most private clinics have resorted to manual clinic data entry as opposed to electronic data management which employs various means to verify the data right from entry and analysis.

Table 8: Receiving formal training in data management

Response	Frequency	Percentage
Yes	45	40.5
No	66	59.2
Total	111	100%

Source: Primary data

On whether employees in data management positions and responsibilities have ever received formal training, 59.2% noted that they did not actually have any formal training; whereas 40.5% said that they had ever received formal training in data management. The findings meant that most employees entrusted with data management positions like entry and analysis in most private clinics did not actually have any formal training to do the job, which may also have serious implications on data quality in most private clinics whether supported by UHMG or not.

In an interview with one In-charge, she was quoted saying:

“Here, we choose a number of employees from our staff members who receive a small on job and hands on training on how to enter data but we don’t send them to institutions to acquire extra training or certificates in data management”.

The above response therefore confirms that employees entrusted with data management in UHMG-supported private clinics do not get formal training in formal schools to handle data tasks, a factor that may also prove detrimental to the overall data quality in these clinics.

4.3 Data quality in UHMG-supported clinics

Table 9: Rating data quality in UHMG-supported private clinics

Level	Frequency	Percentage
Very good	5	4.5
Good	18	16.2
Poor	63	56.8
Very poor	24	21.6
Not sure	1	0.9

Source: Primary data

Regarding data quality, respondents were asked to rate data quality in their respective clinics and 78.4% revealed that their data quality was poor, 20.7% noted that their data was good, and 0.9% were not sure on the idea whether data quality in their respective clinics was poor or good. The findings therefore confirm that data quality in private clinics was poor, an indication that there were reasons why data quality was poor -- a case which made this study necessary.

In an interview with one of the clinic administrators, he was quoted saying:

Data quality issues present challenges in both identifying where data problems exist and in quantifying the extent of the problems existing within the clinic data. In this clinic, the amount of missing information is alarming and what bothers me is that very few people here see this as a problem.

The findings above confirm that data quality in UHMG-supported private clinics is poor and therefore poses a challenge to the clinical staff that would love to use this data for patient management, planning, budgeting and forecasting procurement of drugs as well as UHMG and MOH that rely on this data to inform strategic decision making.

Table 10: Data quality dimensions in UHMG-supported clinics

Data quality dimensions	Response				
	Strongly agree	Agree	Disagree	Strongly disagree	Not sure
Our Clinic data is accurate	13(11.7%)	18(16.2%)	51(45.6%)	20(18%)	9(8.1%)
Our Clinic data is complete	6(5.4%)	25(22.5%)	36(32.4%)	36(32.4%)	8(7.2%)
Our Clinic data is valid	7(6.3%)	34(30.6%)	43(38.7%)	17(15.3%)	10(9%)
Our Clinic data is timely	7(6.3%)	29(26.1%)	50(45%)	20(18%)	5(4.5%)
Our Clinic data is consistent	12(10.8%)	34(30.6%)	44(39.6%)	16(14.4%)	5(4.5%)
Our Clinic data is current	3(2.7%)	25(22.5%)	42(37.8%)	28(25.2%)	13(11.7%)
Our Clinic data is relevant	15(13.5%)	27(24.3%)	45(40.5%)	19(17.1%)	5(4.5%)

Source: Primary data

Respondents were also asked to respond to the data quality dimensions on accuracy, completeness, validity, timeliness, consistency, currency and relevance. The majority of the respondents (63.6%) noted that their clinic data was not accurate, with 27.9% saying that their clinic data was accurate, while 8.1% were not sure about the idea. The findings are complemented by an interview with one In-charge quoted below:

“The challenge for the clinic is to make data accuracy a requirement because most data entered is found lacking some important information like place of residence and last date when patient visited a health facility and sexually related vital information among others and reduces the accuracy of our clinic data. This means that inaccuracy later transforms into incomplete data and finally compromising data quality.”

More so, 64.8% revealed that their clinic data was incomplete, whereas 27.9% said that their clinic data was complete, and only 7.2% were not certain about the completeness of their clinic data.

In an interview with an In-charge, he was quoted saying:

“...Data errors in clinic database can be costly, even dangerous but we have tried with the meager resources to invest in data management but there are several ways bad data can get into our databases which leads to inaccuracy and inconsistent data quality in a long run.”

Regarding validity of clinic data, 54% noted that their clinic data was invalid, whereas 36.9% acknowledged that their clinic data was valid and 9% revealed that they were not sure about validity of their clinic data. This was supported by an interview from one MOH official who was quoted saying:

“Most private clinic data is characterized with vast validity problems that are severe enough to limit its usefulness in examining clinic’s and patients’ progress. On several occasions, clinics have not sufficiently defined categories that use codes to profile sicknesses, they cannot accurately profile the patients details, there is lack of sufficient recordkeeping capabilities and health personnel in data management are led to make judgments and estimations of data that may not be representative of information collected which proves detrimental to the overall process of data quality in the clinics.”

Regarding timeliness of clinic data, majority of the respondents (63%) noted that their clinic data was not timely, whereas 32.4% agreed that their clinic data was accurate, and 4.5% were not sure about the idea. The findings meant that much of the clinic data collected from patients was not immediately recorded into consistent data management tools for consistency and reference on time which means that the content may be distorted in a long run.

4.4 Internal factors affecting data quality in UHMG Supported private clinics

Table 11: Internal factors affecting data quality in UHMG-supported private clinics

Internal factors Item	Response				
	Strongly agree	Agree	Disagree	Strongly disagree	Not sure
Quality healthcare data depends on the availability of data itself in the clinics.	22(19.8%)	42(37.8%)	20(18%)	19(17.1%)	8(7.2%)
Poor documentation within clinics have adverse effects on the final data set to be compiled	40(36%)	36(32.4%)	6(5.4%)	21(18.9%)	8(7.2%)
Most clinics compile inaccurate data which automatically distorts data quality in a long run	35(31.5%)	43(38.7%)	2(1.8%)	25(22.5%)	6(5.4%)
There is insufficient data communication which result in errors and adverse incidents in final data sets	31(27.9%)	32(28.8%)	21(18.9%)	24(21.6%)	3(2.7%)
Cases of poor data collection, sharing, and reporting impacts reduces data quality in private clinics	15(13.5%)	45(40.5%)	19(17.1%)	19(17.1%)	13(11.7%)
Most data lack appropriate content in a usable and accessible form which negatively affect data quality	22(19.8%)	38(34.2%)	39(35.1%)	5(4.5%)	7(6.3%)
Accurate data leads to quality information that is required for quality decision making and patient care.	33(29.7%)	44(39.6%)	13(11.7%)	11(9.9%)	10(9%)
Most clinics don't have consistent data models that would ensure the integrity and quality of the data	24(21.6%)	58(52.3%)	19(17.1%)	5(4.5%)	5(4.5%)
Completeness of information entered into the clinic record is not dependable	17(15.3%)	52(46.8%)	26(23.4%)	10(9%)	6(5.4%)

Source: Primary data

Respondents were asked to respond to statements on how internal factors affect data quality in UHMG-supported clinics and 57.6% agreed to the idea that quality healthcare data depends on the availability of data itself in the clinics, but 35.1% disagreed, and 7.2%) were no decide about the idea. An interview with one MOH revealed:

“Data is impacted by numerous processes, most of which affect its quality to a certain degree but for data to be worked on like entering it, processing and analyzing it, it must be available. This has been one of the major challenges private clinics face. They don’t readily get the required information from patients to enter into their respective compilation tools. This means that data quality is broken at the start of the data management process.”

The above findings therefore implied that databases rarely begin their life empty, hence, must be available which means that the starting point in the lifecycle is a data conversion from some previously existing data source, which in this case is the patient and if it is a bad beginning, it affects the overall data quality.

Furthermore, 68.4% agreed to the idea that poor documentation within clinics has adverse effects on the final data set to be compiled, whereas 24.3% disagreed, and 7.2% were not sure about it.

In an interview with an administrator, he was quoted saying:

:For now we have tried to tackle the internal issues affecting our clinic data on many fronts, we have involved many people, tools and techniques to help us put good clinic data into our systems and then constantly work at preventing it going bad but there are many cases like that of completeness and consistency, which affects data quality.”

Majority of the respondents (70.2%) revealed that most clinics compiled inaccurate data which automatically distorted data quality in the long run; but 24.3% disagreed with the idea; and only 2.7% were not sure about it. An MOH official was quoted saying:

“The processes that bring data into the clinic database from outside either manually or through various interfaces and data integration techniques are inaccurate where 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 exportation and these data traffic dramatically magnify these problems.”

There is insufficient data communication which results in errors and adverse incidents in final data sets as shown by 56.7% response; whereas 40.5% disagreed; and 2.7% were not sure about the idea. These findings mean that data is rarely exchanged between the systems through real-time interfaces, implying that as soon as the data enters one database, it does not trigger procedures necessary to send information to other downstream clinic databases and this does not give room for immediate propagation of data to all the relevant clinic databases thus negatively impacting data quality.

In an interview, one of the respondents was quoted saying:

“Our common clinic error source is data entry and it is much a bigger cause of poor data communication right from entry to other clinic departments. Over time, a common data entry problem is handling missing values in clinic data where users may assign the same blank value to various types of missing values which distorts all the data set in a long run.”

In addition to the above, there are cases of poor data collection, sharing, and reporting impacts reduces data quality in private clinics as shown by 54% response, whereas 34.2% disagreed with the idea, and 11.7% were not sure about the statement. This is supported by an interview with one respondent quoted as saying:

“Data compilation using a form or register also leads to an increase in the number of errors. Clinic staffs in data management often prefer to find and use the easiest way to complete patient forms, which means they adopt a data entry format which is easier but with adverse effects on data quality. This also means that data quality is compromised by data entrants de through making deliberate data management mistakes internally.”

Most private clinic data lack appropriate content in a usable and accessible form which negatively affects data quality as agreed by 54% respondents, whereas 39.6% of the respondents disagreed with the idea, and 6.3% were not sure about it. More so, 69.3% agreed to the idea that accurate data leads to quality information that is required for quality decision making and patient care, but 21.6% disagreed with the idea.

Majority of the respondents (73.9%) revealed that most clinics did not have consistent data models that would ensure the integrity and quality of the data; but 21.6% noted that their respective clinics had models or procedures they followed in data management which were meant to ensure integrity and quality of clinic data; though 4.5% were not sure about the idea. Completeness of information entered into the clinic records is not dependable as revealed by 62.1% of respondents; whereas 32.4% noted that their data is dependable and useful in actual use for decision making by both the clinic and the relevant authorities like government and other health-related institutions.

In an interview with one clinic administrator, he said:

“A big challenge among the internal set up of clinics is that data quality specifications often do not reflect actual clinic data requirements. As a result, data may be brought in compliance with some theoretical model but remain undependable and incorrect for actual use.”

The above findings implied that without data dependability, quality of data would not process because of undependable data is useless for decision making. Therefore, the fact that most clinic data was found not to be dependable implies that data quality is low because dependability is a big factor that can affect data quality in terms of clinic business and long-run survival.

Table 12: Correlation between Internal factors and data quality in private clinics

Correlations			
		Data quality in Private clinic	Internal Factors in UHMG clinics
Data quality in Private clinic	Pearson Correlation	1	-.282**
	Sig. (2-tailed)		.003
	N	111	111
Internal Factors in UHMG clinics	Pearson Correlation	-.282**	1
	Sig. (2-tailed)	.003	
	N	111	111

** . Correlation is significant at the 0.01 level (2-tailed).

The analysis of data revealed that internal factors in private clinics significantly ($P=0.01$) and negatively ($r= -0.282$) affect data quality in UHMG-supported private clinics in Kampala. Hence this implies that data quality in UHMG-supported private clinics in Kampala is negatively affected by internal factors like low levels of accuracy, completeness and validity among others.

Hence we accept the hypothesis that, “internal factors in affect data quality in UHMG-supported private clinics in Kampala”.

4.5 External factors affecting data quality in UHMG-supported private clinics

Table 13: External factors affecting data quality in UHMG-supported private clinics

External factors Item	Response				
	Strongly agree	Agree	Disagree	Strongly disagree	Not sure
Maintaining quality data provided by clinics offers a challenge ensuring the integrity of the healthcare data	18(16.2%)	36(32.4%)	31(27.9%)	20(18%)	6(5.4%)
Documentation and data content within clinics is not universally understood by data users, thus affecting its quality	13(11.7%)	57(51.4%)	30(27%)	8(7.2%)	3(2.7%)
Monitoring and keeping track of data over time and reporting variations in the data affects quality of data.	22(19.8%)	48(43.2%)	27(24.3%)	8(7.2%)	6(5.4%)
Data compiled by private clinics is not relevant with what users health needs requirements	14(12.6%)	48(43.2%)	22(19.8%)	16(14.4%)	11(9.9%)
Private clinics compile insufficient data compared to what is required by the data users	19(17.1%)	61(55%)	13(11.7%)	8(7.2%)	10(9%)
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	41(36.9%)	32(28.8%)	15(13.5%)	17(15.3%)	6(5.4%)

Source: Primary data

Majority of the respondents (48.6%) noted that maintaining quality data provided by clinics offers a challenge ensuring the integrity of the healthcare data, whereas 45.9% disagreed saying that maintaining quality data provided by clinics does not offer a challenge ensuring the integrity of the healthcare data, with 5.4% not sure about the idea.

More so, 63.1% agreed to the idea that documentation and data content within clinics is not universally understood by data users, thus affecting its quality, whereas 34.2% disagreed with the idea, and 2.7% were not sure it. In an interview with one health personnel in charge of data management, she was quoted saying;

“We experience a challenge in data quality because the information provided by the patients who come to the clinic offer confusing information which is not understandable by the data entrants and the rest of the users. This implies that poor data is captured as is provided by the patient and entered, which is therefore depended on in clinic databases hence affecting data quality in private clinic data sets.”

Another interview response from an In-charge noted as follows:

“Quality data in healthcare is cooperation between the patient and the healthcare provider but if a patient starts the data collection process by providing incomplete information, data quality is negatively affected. Personal factors of the provider and the patient, and factors pertaining to the clinic, healthcare system, and the broader environment affect clinic data quality.”

Furthermore, monitoring and keeping track of data over time and reporting variations in the data affects quality of data as revealed by 63% response; but 28.5% disagreed saying that it was not tracking and monitoring which affects data quality in private clinics. In an interview with a clinic administrator, he was quoted saying;

“Data quality issues are rarely one. Regular monitoring of key data quality metrics, with common examples such as free of error, completeness and consistency, ensures that reports are accurate and make it to their destination in a timely manner but this is a myth to most clinics because they rarely monitor their datasets.”

The above findings mean that the lack of data monitoring affects data quality since trustworthy data motivates users to harness the information in new ways, giving rise to fresh ideas and helps them when eliminating inaccuracies and duplication from information systems. It is about creating new opportunities by harmonizing the data from disparate systems and providing stakeholders with quality data.

Data compiled by private clinics is not relevant, with user health needs requirements as shown by 55.8% response; whereas 34.2% disagreed with the idea, with 9.9% not sure about the idea. More so, 72.1% agreed to the idea that private clinics compile insufficient data compared to what is required by the data users but 18.9% disagreed and 9% were not sure.

The findings above mean that improperly selected data can strongly influence the conclusions of a health assessment, if such data are not of high quality or of solid relevance, it may hinder private clinics from making sense out of them for the health of the users and Ugandans at large.

More so, 65.7% agreed to the idea that 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; though 28.8% disagreed with the idea, and 5.4% were not sure. The findings therefore confirm that data quality should not be compromised right from the start up to interpretation.

Table 14: Correlation between External factors and data quality in private clinics

Correlations			
		Data quality in Private clinic	External Factors in UHMG clinics
Data quality in Private clinic	Pearson Correlation	1	-.131
	Sig. (2-tailed)		.170
	N	111	111
External Factors in UHMG clinics	Pearson Correlation	-.131	1
	Sig. (2-tailed)	.170	
	N	111	111

There was a negative ($r = -0.131$) and but not very significant ($P=0.01$) relationship between external factors and data quality in UHG private clinics in Kampala. This implied that data quality in UHMG supported private clinics in Kampala was negatively affected ($r = -0.131$) by external factors like unfriendly data users and incomplete information provided by patients during data collection and entry. Hence we accept the hypothesis that, “there is a relationship (negative relationship) between external factors and data quality in UHMG supported private clinics in Kampala”.

4.6 Multiple Regression analysis and Model development

Correlations alone could not help the researcher find a conclusive position as to what extent the independent variable impacts (affects) the dependent variable. So to ascertain the extent of the relationship between the independent variables (factors) and the dependent variable (data quality), multiple regression analysis was carried out to establish the extent to which each independent variable impacts on data quality. The results are shown in the table below.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.363 ^a	.132	.116	.72809

a. Predictors: (Constant), External Factors in UHMG clinics, Internal Factors in UHMG clinics

Table 15: Regression Analysis between Factors and Data quality in UHMG clinics

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	4.113	.306		13.445	.000
	Internal Factors in UHMG clinics	-.206	.054	-.356	-3.780	.000
	External Factors in UHMG clinics	-.175	.069	-.241	-2.560	.012

a. Dependent Variable: Data quality in UHMG Private clinics

From the regression analysis table above, Data quality was treated as the dependent variable influenced by Internal factors and External factors. The hypothetical regression model that guided this study was in the multiple regression equation form of:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_n X_n$$

Where: Y is the dependent variable (Data quality), “ ” is a regression constant; β_1 , β_2 , β_3 and β_n are the beta coefficients; and X_1 , X_2 , X_3 , and X_n are the independent (predicator) variables. Therefore, standardized beta coefficients were substituted in the hypothesized regression equation. This revealed that Data quality can be predicated as:

$$Y = 4.113 - .356 X_1 - .241 X_2$$

Where: Y is Data quality; X_1 , is Internal Factors in UHMG clinics X_2 , is and External Factors in UHMG clinics.

Interpretation of the Multiple Regression model findings

From the regression analysis and model above, Internal and External Factors in UHMG clinics were found to have negative and significant impact on Data quality in UHMG-supported clinics in Kampala. For instance, internal factors with ($\beta_1 = -0.356$) and external factors with ($\beta_2 = -0.241$) imply that they negatively impact on data quality. We also observe that about 11.6% of the variation (adjusted $R^2 = 0.116$) in data quality in Kampala, is explained by a combination of internal and external factors. Hence, the other factors that this study may not have examined could be taken to explain the remaining 88.4% of the variation in data quality in UHMG-supported private clinics.

CHAPTER FIVE

SUMMARY, DISCUSSION, CONCLUSIONS AND RECOMMENDATIONS

5.1 Introduction

This chapter presents the summary of findings, discussion of findings, conclusions and recommendations regarding the study findings. The chapter is also arranged according to the study objectives.

5.2 Summary of findings

5.2.1 Internal factors affecting data quality in UHMG-supported private clinics

Study findings revealed that that quality healthcare data depends on the availability of data itself in the clinics which implied that databases rarely begin their life empty, hence, must be available which means that the starting point in the lifecycle is a data conversion from some previously existing data source, which in this case is the patient and if it is a bad beginning, it affects the overall data quality.

Furthermore, poor documentation within clinics has adverse effects on the final data set to be compiled and most clinics compiled inaccurate data which automatically distorts data quality in the long run. The findings also showed that there is insufficient data communication which results in errors and adverse incidents in final data sets which means that data is rarely exchanged between the systems through real-time interfaces.

In addition to the above, findings confirmed that there are cases of poor data collection, sharing, and reporting impacts that reduce data quality in private clinics and most private clinics' data lack appropriate content in a usable and accessible form which negatively affects data quality.

Majority of the respondents also revealed that most clinics do not have consistent data models that would ensure the integrity and quality of the data and completeness of information entered into the clinic records is not dependable. The findings implied that without data dependability, quality of data would not process because undependable data is useless for decision making.

The findings revealed that internal factors in private clinics significantly and negatively affect data quality in UHMG-supported private clinics in Kampala which implies that internal factors like low levels of accuracy, completeness and validity among others.

5.2.2 External factors affecting data quality in UHMG-supported private clinics

The study findings revealed that maintaining quality data provided by clinics offers a challenge ensuring the integrity of the healthcare data and documentation and data content within clinics is not universally understood by data users, thus affecting its quality.

Furthermore, poor monitoring and keeping track of data over time and reporting variations in the data affects quality of data, which means that such data is not of any use because it may be distorted once it is not followed up very strictly by the relevant users especially the Ministry of Health officials and other authorities.

The findings mean that the lack of data monitoring affects data quality since trustworthy data motivates users to harness the information in new ways, giving rise to fresh ideas and helps them when eliminating inaccuracies and duplication from information systems. It is about creating new opportunities by harmonizing the data from disparate systems and providing stakeholders with quality data.

More so, the study found out that data compiled by private clinics is not relevant with what users of health data need. The study also shows that private clinics compile insufficient data compared to what is required by the data users. The findings above mean that improperly selected data can strongly influence the conclusions of a health assessment, if such data are not of high quality or solid relevance, hence the users need relevant data from private clinics to help them make sense out of them for the health of the users and Ugandans at large.

More so, the study findings confirmed that clinic data sets are hard to interpret by external users and it therefore becomes very hard for users to pick any meaningful information out of it, hence compromising data quality. The findings therefore confirm that data quality should not be compromised right from the start up to interpretation.

The findings revealed that there was a negative but not very significant relationship between external factors and data quality in UHG private clinics in Kampala, hence, external factors like unfriendly data users and incomplete information provided by patients during data collection and entry.

5.3 Discussion of findings

5.3.1 Internal factors affecting data quality in UHMG-supported private clinics

Findings revealed that that quality healthcare data depends on the availability of data itself in the clinics which implied that databases rarely begin their life empty, hence, must be available which means that the starting point in the lifecycle is a data conversion from some previously existing data source, which is this case the patient and if it is a bad beginning, it affects the overall data quality. The findings are in line with Weir (2003) who noted that data availability from patients

determines how the clinic databases will use the rest of the information received from the patients.

Furthermore, poor documentation within clinics has adverse effects on the final data set to be compiled and most clinics compile inaccurate data which automatically distorts data quality in the long run. The findings are complemented by Arkady (2007) who also argued that processes that manipulate the data inside the health care databases like documentation affect the data quality.

He further noted that clinics poorly document patients data and 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 and therefore not surprising that internal data processing often leads to numerous data problems which reduce data quality.

The findings also showed that there is insufficient data communication which results in errors and adverse incidents in final data sets which means that data is rarely exchanged between the systems through real-time interfaces. These are supported by Abdelhak, Grostick&Hankin, 2001) who also found out that there are aspects in health data management that cause accurate data to become inaccurate over time, without any physical changes made to it, a case in point being insufficient data communication among the data handlers and users.

In addition to the above, findings confirmed that there are cases where poor data collection, sharing, and reporting impacts reduces data quality in private clinics and most private clinic data lack appropriate content in a usable and accessible form which negatively affects data quality. This is similar to a study done by Jones (2003) who argued that poor data entry in health forms and instructions somewhat increase or mitigate data entry problems internally.

Majority of the respondents also revealed that most clinics do not have consistent data models that would ensure the integrity and quality of the data and completeness of information entered into the clinic records is not dependable. The findings implied that without data dependability, quality of data would not process because undependable data is useless for decision making. This however contradicts with Arkady (2007) who acknowledged that health facilities try to have consistent data sets that are used as a basis for monitoring and evaluation of the patients' progress and trends.

The study findings also found out that there were cases of low reliability levels of the data collected by private clinics which later negatively affects data quality. The findings are complemented by a study done by Moyo (2005) in Zimbabwe which 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, otherwise, it reduces data quality once the data values are not within specified bounds.

The analysis of data revealed that internal factors in private clinics significantly and negatively affect data quality in UHMG-supported private clinics in Kampala. The findings are similar to the one done by Burger (2007) who argued that timeliness affects data quality and many more internally motivated factors which reduce data quality on an everyday basis.

The findings are however contradicted with AbouZahr (2005), who noted that it was only a wrong precision within data sets that have negatively affected data quality in many organizations. He further argued that this is worsened by poor data quality specifications which often do not reflect actual data requirements.

5.3.2 External factors affecting data quality in UHMG-supported private clinics

The study findings revealed that maintaining quality data provided by clinics offers a challenge ensuring the integrity of the healthcare data and documentation and data content within clinics is not universally understood by data users, thus affecting its quality. These findings are similar to the one done by Arkady (2007) who noted that processes that bring data into the database from outside either manually or through various interfaces and health data integration techniques affect health data quality. Jones (2004) also completed Arkady's argument saying that some of the incoming data are incorrect in the first place and simply migrate from one place to another which means they come from the outside source with poorly framed information which later distorts the latter's quality.

Furthermore, poor monitoring and keeping track of data over time and reporting variations in the data affects quality of data which means that such data is not of any important use because it may be distorted once it is not followed up very strictly by the relevant users especially the Ministry of Health officials and other authorities.

The findings mean that the lack of data monitoring affects data quality since trustworthy data motivates users to harness the information in new ways, giving rise to fresh ideas and helps them when eliminating inaccuracies and duplication from information systems. It is about creating new opportunities by harmonizing the data from disparate systems and providing stakeholders with quality data.

More so, the study found out that data compiled by private clinics is not relevant to what users of health data need which is supported by Jones (2004) who also argued that clinics in urban centres compile irrelevant data which later becomes useless for the final users. The findings above mean

that improperly selected data can strongly influence the conclusions of a health assessment, if such data are not of high quality or solid relevance hence the users need relevant data from private clinics to help them make sense out of them for the health of the users and Ugandans at large.

The study also shows that private clinics compile insufficient data compared to what is required by the data users. This is complemented by Brouwer (2006) who noted that 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.

More so, the study findings confirmed that clinic data sets are hard for external users to interpret and it therefore becomes very hard for users to pick any meaningful information out of it, hence compromising data quality. The findings therefore confirm that data quality should not be compromised right from the start up to interpretation. These findings are however in line with Redman (2001) who noted that understandable information leads to data quality reporting in many organizations and vice versa.

The study confirmed a negative relationship between external factors and data quality in UHMG private clinics in Kampala which therefore implies that external factors like unfriendly data users and incomplete information provided by patients during data collection and entry. The findings are in agreement with Berendsohn (2000) who argued that there are many data quality principles that apply when dealing with species data and especially with the spatial aspects of those data and if these principles are not involved at all stages of the data management process, it negatively affects data quality.

5.4 Conclusion

5.4.1 Internal factors affecting data quality in private clinics

The study confirmed that internal factors have a negative relationship with data quality implying that if these internally driven factors are not improved or rectified, data quality in private clinics will continuously be poor and vice versa.

5.4.2 External factors affecting data quality in private clinics

The study confirmed that external factors have a negative relationship with data quality implying that if these externally-driven factors are not improved or rectified, data quality in private clinics will continuously be poor and vice versa.

5.5 Recommendations

The study recommends the following:

Private clinics need to invest in Monitoring and Evaluation departments. Although data in the clinics is collected by all health workers that interact with patients, clinics must purposely set up M&E departments to oversee the whole reporting function. This oversight department would make sure that data collected meets the required standards and that submission of reports is done timely. It would also be responsible for mentoring other sections in data management.

Private clinics need to motivate their employees to reduce on the high levels of staff turnover that create knowledge gaps where all the trained and mentored staff in data management keep leaving the clinics for greener pastures in Government facilities. Paying of staff timely, giving job contracts for job security and having clear times of references and employing the right numbers of staff to avoid overworking employees would greatly reduce staff turnover.

UHMG should invest in electronic systems in these private clinics if they are to collect accurate and reliable data. The use of papers and registers is very laborious and prone to many errors. Data cleaning and extraction using manual systems is an uphill task that produces unreliable results. Purchasing computers and installing software for data capture and retrieval would be of great help

The study also recommends that UHMG in partnership with the Ministry of Health should organize formal trainings for all clinic staff that handle patients' data to equip them with skills that can enable them collect clean and reliable data. Empowering the clinics with data management skills and having M&E topics included in the clinics' Continuous Medical Education (CME) sessions would translate into the culture of data demand and information use (DDIU) which is the ultimate reason for compiling data.

Lastly, the study recommends that the Ministry of Health should start supervising private clinics not only for licenses but also for quality assurance across all departments. This would ensure that sustainable M&E departments are created in the private health sector since the donor-supported implementing agencies that support these clinics like UHMG are time-bound and likely to stop their support when there is no more funding from the donors.

5.6 Contribution of the study

The findings and recommendations of this study will be very useful to different stakeholders as seen below:

The findings will benefit the Ministry of Health as they 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.

UHMG will use the findings of this study to come up with strategies to bridge the gaps that were identified to improve 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 will add new concepts and knowledge to the existing body of knowledge of data quality. The study findings will provide up-to-date literature to the academicians who may wish to carry out similar or related study. The study findings should stimulate further research in data quality issues.

5.7 Areas for future research

The researcher recommends the following areas for further research:

Why is Data Demand and Information Use Culture very low in Uganda's Health sector despite the fact that they collect big volumes of data on a daily basis?

References

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

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). Basingstoke, UK: Macmillan Publishers.

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.

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

Barifaijo, K.M., Basheka, B. & Oonyu, J. (2010). *How to Write a Good Dissertation* (1st Edition), Kampala: New Vision Printing and Publishing Company Ltd.

Berendsohn, W.G. (2000). *A taxonomic information model for botanical databases* (4th Edition). Essex, England: Pearson Education Ltd.

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

Brouwer, H. (2006). *Data quality improvement in general practice*. Lasi: Poliron Publishing House.

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

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>

Communication.in V.H. Heywood, (ed.) *Global Biodiversity Assessment*. Cambridge University Press, London, UK. pp. 607–670.

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), New York, USA, McGraw-Hill Publishers.

Educational and Psychological Measurement, Ed Deusto, Harvard University Press, Bilbao Publishers.

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

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

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

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

Ivanov, K. (1972). Quality-control of information: On the concept of accuracy of information in data banks and in management information Systems. (ed), Stockholm University Press, Sweden.

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.

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

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*, New York, USA, McGraw-Hill Publishers.

Kothari, C.R. (2003). *Research Methodology: Methods and Techniques* (2nd Edition). New Age International (P) Ltd Publishers, New Delhi.

Krejcie, R.V. and Morgan, D.W. (1990). Determining sample size for research activities, Liepins, G. E. &Uppuluri, V. R. R. *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.

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 Publishers.

Mugenda, O.M. and Mugenda, A.G. (2003). *Research Methods: Quantitative and Qualitative Approaches* (2nd Edition). African Centre of Technology Studies: Nairobi, Kenya.

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

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.

Olivieri, S., Harrison, J. and Busby, J.R. (1995). *Data and Information Management and*

Olson, J.E. (2003). *Data Quality. The Accuracy Dimension.* (2nd Edition), San Francisco, Morgan Kaufmann Publishers.

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

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

Sekaran, U. (2000). *Research Methods for Business,* (edn). USA: Hermitage Publishing Services Publication Company.

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

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 Publishers.

Wang, R.Y., Storey, V.C., Firth, C.P. (1995). A framework for analysis of data quality research, *IEEE Transactions on Knowledge and Data Engineering* (Ed). UK: Cambridge University Press.

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

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

APPENDIX I: QUESTIONNAIRE FOR CLINIC STAFF AND IN-CHARGE
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 DIMENSIONS 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					

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 meMy 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 Baryamurebaa** 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-30 th March 2015	Working on the research proposal and its final submission.
1 st April-30 th April 2015	Development of Tools
1 st May-30 th 2015	Data collection
1 st June-30 th June 2015	Data entry
1 st July-30 th July 2015	Data analysis
1 st August-30 th Aug 2015	Reporting writing
1 st Sept-30 th Oct 2015	Review and Submission of the dissertation
1 st Nov-30 th 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

<i>N</i>	<i>S</i>	<i>N</i>	<i>S</i>	<i>N</i>	<i>S</i>
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	100000	384

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

Source: Krejcie & Morgan, 1970

APPENDIX V: LETTER OF INTRODUCTION



UGANDA TECHNOLOGY AND MANAGEMENT UNIVERSITY

UTAMU

1st June 2015

TO WHOM IT MAY CONCERN

RE: BARYAMUREEBA MILTON-REG. NO. JAN15/PM&E/0366U

This is to introduce Baryamureeba Milton who is a student in the School of Business and Management doing a Masters in Monitoring and Evaluation of Uganda Technology And Management University (UTAMU).

As part of the course, he wants to undertake a research study on **"Factors affecting Data Quality in Uganda Health Marketing Group supported private Clinics in Kampala District"**.

Any assistance rendered to him will highly be appreciated. In case you need any further information, do not hesitate to contact the undersigned.

Sincerely,

Professor Benon C. Basheka

Dean, School of Business and Management

Cc. Deputy Vice Chancellor, UTAMU
Director, Academic Affairs, UTAMU

APPENDIX VII: LANGUAGE EDITOR CERTIFICATE

MUKOTANI RUGYENDO

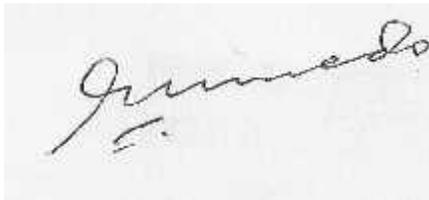
P.O. BOX 31178

KAMPALA

TEL: 0701707093

CERTIFICATE OF PROOF THAT DISSERTATION HAS BEEN EDITED

This is to certify that the Master's Degree dissertation entitled, 'FACTORS AFFECTING DATA QUALITY IN PRIVATE CLINICS IN UGANDA: THE CASE OF SELECTED UGANDA HEALTH MARKETING GROUP-SUPPORTED CINICS IN KAMPALA' by MILTON BARYAMUREEBA has been reviewed and corrected in order to ensure clarity of expression and consistency regarding key style aspects like content layout, sentence construction, logical flow, spelling, word use, punctuation, citation and referencing.

A handwritten signature in black ink, appearing to read 'Mukotani Rugyendo', is written on a light-colored background.

Mukotani Rugyendo

Professional Editor