ADHERENCE TO DATA PROTECTION GUIDELINES AMONG HEALTH RESEARCHERS AT KENYA MEDICAL RESEARCH INSTITUTE

KEBENEI ENOCK KIPCHIRCHIR (BSC. INFORMATION SCIENCES)

Q141/CTY/PT/27648/2014

DEPARTMENT OF HEALTH MANAGEMENT AND INFORMATICS

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF MASTER OF SCIENCE (HEALTH INFORMATION MANAGEMENT) IN THE SCHOOL OF PUBLIC HEALTH AND APPLIED HUMAN SCIENCES OF KENYATTA UNIVERSITY

FEBRUARY 2021

DECLARATION

This thesis is my original work and has not been presented for a degree in any other University.

Signature …….. .Date…….
Kebenei Enock Kipchirchir
Q141/CTY/PT/27648/2014

SUPERVISORS
This thesis has been submitted for review with our approval as University Supervisors.

Signature …… Date……..

Dr. George O. Otieno,
Department of Health Management and Health Informatics
Kenyatta University

Signature …… Date……

Dr. Kenneth Rucha
Department of Health Management & Informatics Department
Kenyatta University

DEDICATION

This research work is dedicated to my wife, Valentine and my sons, Brevyn and Bayne for their continued patience and support. I also extend dedication to my parents and siblings for their endless prayers.

ACKNOWLEDGEMENT

First and foremost, I take this opportunity to thank God for enabling me to go this far.
Secondly, my gratitude goes to all scientists and laboratory officers at the Kenya Medical Research Institute for their maximum co-operation during my data collection period. Special thanks to the Management of KEMRI for granting me permission to administer questionnaires to the employees.
Special thanks go to my research supervisors, Dr. George Otieno and Dr. Kenneth Rucha for their tireless effort throughout this work.
Finally, I would like to appreciate my 2014 September Msc. Health Management and Informatics classmates for their support and co-operation throughout the period of my study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS AND ACRONYMS

CBRD	Centre for Biotechnology Research and Development
CCR	Centre for Clinical Research
CGHR	Centre for Global Health Research
CGMR-C	Centre for Geographic Medicine Research-Coast
CIOMS	Council for International Organizations of Medical Sciences
CMR	Centre for Microbiology Research
CPHR	Centre for Public Health Research
CRDR	Centre for Respiratory Disease Research
CTMDR	Centre for Tradition Medicine and Drug Research
CVR	Centre for Virus Research
EHRs	Electronic Health Records
ERC	Ethics Review Committee
ESACIPAC	Eastern & Southern Africa Centre of International Parasite Control
EU	European Union
GCP	Good Clinical Practice
GDPR	General Data Protection Regulation
HRR	Health Research Regulations
KEMRI	Kenya Medical Research Institute
NACOSTI	National Commission for Science, Technology and Innovation
SERU	Scientific and Ethics Review Unit
WMA	World Medical Association

DEFINITION OF TERMS

Adherence: In this study, adherence refers to the degree or level of conformity to the national (NACOSTI) guidelines on data protection in human health research. A total of nine questions drawn from the NACOSTI guidelines were asked to determine if the respondents conform to or follow what the guideline recommends.

Data Protection is a process of putting controls to the personal information used by organizations, businesses, government, or individuals.

Health Research Data refers to the facts, figures, and processed information about the research participants.

Health Research/Medical Research/Clinical Research refers to the systematic investigation into the ways of treating and preventing diseases.

Research Participants refers to the people or subjects who take part in a clinical trial, or any type of research by being the target of observation.

Individual factors in this study are the personal issues or determinants that may influence adherence to data protection guidelines such as own attitudes, beliefs, skills, and features among others.

Organizational factors in this study are the institutional determinants that may influence adherence to data protection guidelines such as the predominant cultures, policies, and availability of equipment among others.

Health research data protection practices in this study are the ways which the researcher collects, stores, organizes, or disseminates data from their respective researches.

Research sponsor in this study is the funder of the project who may have own requirements.

ABSTRACT

Researchers are expected to keep participants' data in a highly confidential and private manner. A study conducted in Kenya in 2014 revealed that the research stakeholders face different challenges relating to the sharing of public health data. The exposure of data occurs through stigmatization, invasion of privacy, disrespecting autonomy and unfair competition either intentionally or unintentional 'misuse' of data. The general research objective of this study was to examine the adherence to the data protection guidelines in health research in KEMRI, Kenya. Cross-sectional study design was used and it employed quantitative methods of data collection and analysis. The sample size for this study was 128 research participants, however an extra 10% was added to cover for non-respondence. The study targeted the KEMRI's scientists who have participated in any research project. Stratified sampling method and the "Probability Proportional to Size" (PPS) was used to get the desired sample in each of the KEMRI center. The data analysis was done using SPSS Version 23. Descriptive statistics and chi-square test were done to determine significant association and results presented in tables, graphs and charts. A total of nine questions were asked to determine the adherence to the national guidelines. A respondent is considered to have adhered if he/she has agreed to all the 9 items. The neutral respondents were considered as non-adherence. The findings reveal that 18 (12.6%) of the respondents adhered to data protection guidelines in health research while the majority did not adhere 121 (87.4%). P-values <0.05 were considered significant. Results further showed that guidelines or policies on data protection within the institute are the organizational factor which highly influences adherence to data protection (p-value of 0.01). Restricting access to the authorized persons and use of codes to conceal participant's identity (p-value of 0.04) are the best ways of protecting health research data. In conclusion, most researchers do not comply with all aspects of national guidelines on data protection which may lead to the exposure and leakage of participant's data. In view of the findings, the researcher recommends the creation of awareness through workshops and trainings as well as the development of institutional guidelines as the best ways of adhering to data protection guidelines.

CHAPTER ONE – INTRODUCTION

1.1. Background to the study

Research data exist in different forms depending on type of research and the discipline. Data can be in the form of numbers, texts, audio, video, electronic or physical among others (Trewhella, 2014). In health, research data may be in the form of laboratory specimens, samples, field notebooks, electronic or manual databases, clinical records, questionnaires, laboratory results, photographs, manuscripts, artefacts, and audio-visual materials among others (Trewhella, 2014).

The Article 10 of the Declaration of Helsinki states, "It is the duty of the physician in medical research to protect the life, health, privacy, and dignity of the human subject" (WMA, 2006). Article 20 states that "every precaution should be taken to respect the privacy of the subject and the confidentiality of the patient's information" (Sheikh, 2008). The "Council for International Organizations of Medical Sciences" (CIOMS) published international guidelines for medical research involving human subjects. The guidelines states that "patients have the right to expect that their physicians and other health-care professionals will hold all information about them in strict confidence and disclose it only to those who need, or have a legal right to, the information, such as other attending physicians, nurses, or other health-care workers who perform tasks related to the diagnosis and treatment of patients" (CIOMS, 2002). The "Australian Code for the Responsible Conduct of Research" states, "Researchers given access to confidential information must maintain that confidentiality." The Irish Data Protection Acts of 1988 and 2003 provide legislations that safeguards the patient information collected for medical purposes through pseudonymisation, anonymisation, protection of explicit

consent, and development of Electronic Health Records (EHRs) among others (Hawkes, 2007).

The European Union (EU) developed and approved the "General Data Protection Regulation" (GDPR) on 18th March 2018. The objective of these guidelines was to guide member states on the safeguarding of personal information including the health data (Kirwan et al, 2020). The Health Research Regulations (HRRs) of Ireland found out that additional guidelines that govern human health research will have an impact on the different aspects of conducting research. one of the requirements put forward in the HRR was the need to have an 'explicit consent" from a participant before including any identifiable personal information (Kirwan et al, 2020).

The main concepts of data management/best practices that must be put into consideration in carrying out any form of research include "data ownership, gathering, storage, protection and, retention, analysis, sharing, reporting, and data destruction" (Hochstetler, 2009). Data breach refers to a situation where private and confidential data usually the ones that can identify an individual, gets to the wrong or unintended hands (Filkins and Radcliff, 2008). Data protection programs and models aim at safeguarding inadvertent data leaks.

In developing countries, there is a limited guidance to inform regulatory and ethical decisions on issues of privacy and data protection especially with the implementation of eHealth. This concept has posed ethical and regulatory challenges in the management of health research data (Leon et al, 2013).

The "Kenya Access to Information Act No. 31 of 2016" provides the principles of safeguarding personal data in different sectors. The Section 5 of the Act states, "every

person has a right to privacy with respect to their personal data relating to their private and family life." (Laws of Kenya, 2016). The Kenya "National Commission for Science, Technology and Innovation" (NACOSTI) provides the guidelines for conducting health research in Kenya. The guideline number 15 states that the "investigator has to put in place mechanisms to protect the safety and to respect the privacy of the research subjects, and to maintain the confidentiality of the data" (NACOSTI, 2004).The then KEMRI ERC and now Scientific and Ethics Review Unit (SERU) has specific standards and procedures for responsible conduct of human health research in relation to data management (KEMRI, 2004).

NACOSTI is the national body with the mandate to regulate all research activities carried out in Kenya. They developed "Guidelines for Ethical Conduct of Biomedical Research Involving Human Subjects in Kenya" in the year 2004. The guidelines in the NACOSTI document borrow from the regional and international policies/guidelines for carrying out biomedical studies.

The purpose of this research was to investigate the level of adherence to the data protection guidelines among the KEMRI researchers, Kenya.

1.2. Problem statement

In health research, human participants are enrolled and information on their health status is usually captured for research purposes. Researchers are expected to keep participants' data in a highly confidential and private manner. Nevertheless, there are cases where this private and confidential information leaks to the public against research ethics. Research data may potentially leak to the public or escape the researcher at any of the research process. In the year 2010, a doctor from Columbia University mistakenly exposed

confidential clinical information belonging to a Presbyterian patient from the New York City (Boulton, 2014). A study conducted in Kenya (involving KEMRI as one of the Research Stakeholders) revealed that research data exposure might occur through stigmatization, invasion of privacy, disrespecting autonomy and unfair competition. Such exposure exists through either intentional or unintentional 'misuse' of data leading to stigmatization and invasion of privacy of the enrolled participants (Kombe et al., 2015).

Research scientists are expected to observe ethical and scientific principles mentioned in the NACOSTI "guidelines for ethical conduct of biomedical research involving human subjects in Kenya."

In light of the above problem, the purpose of this study was to examine the adherence to the national data protection guidelines among the KEMRI researchers, Kenya.

1.3. Justification of the Study

The core business of KEMRI justifies the need to carry out this study in the specific site. The 1979 Science and Technology Act (Amended in 2013) established KEMRI as a National organization with the responsibility of conducting human health research in the country. The adherence to the data protection guidelines among the KEMRI researchers will help in the formulation of policies and guidelines relating to human health research in Kenya. Secondly, the research findings will be helpful in the attempts to protect the privacy, rights, interests, and welfare of the human participants enrolled in a study. The researcher chose KEMRI because it is one of the leading Institutions that carry out health research in Kenya. It is a state corporation mandated by law to carry out research in human health. The researcher was interested in studies involving human health, hence the justification to consider KEMRI as a study site. The topic was chosen KEMRI took part

in a study which revealed that research data exposure might occur either intentionally or unintentionally (Kombe et al., 2015).

1.4. Research Questions

The research objectives translate into the following research questions;

1. What are the organizational factors influencing adherence to the data protection guidelines among health researchers in KEMRI, Kenya?
2. What are the individual factors influencing adherence to health research data protection guidelines among KEMRI researchers in Kenya?
3. What are the practices of protecting health research data among KEMRI researchers in Kenya?

1.5. General hypothesis

Null hypothesis: There are no factors influencing adherence to the data protection guidelines among health researchers in KEMRI, Kenya.

1.6. Study Objectives

The general research objective of this study was to examine the adherence to the data protection guidelines among health researchers in KEMRI, Kenya

Specific Objectives;

i. To determine organizational factors influencing adherence to the data protection guidelines among health researchers in KEMRI, Kenya.
ii. To find out the individual factors influencing adherence to health research data protection guidelines among the KEMRI researchers in Kenya.
iii. To determine health research data protection practices among the KEMRI researchers in Kenya.

1.7. Significance of the Study

The findings of this study will benefit KEMRI particularly the Scientific and Ethics Review Unit (SERU) in the development of compliance programs to the data protection guidelines in health research. The results can be adopted in the Standard Operating Procedures on data protection. Other beneficiaries of this study include the researchers at KEMRI because privacy and confidentiality of participants' information is one of the very important tenets of any health research. Lastly, this study shall benefit the enrolled human subjects because data protection practices helps in safeguarding their private and confidential information. The adherence to the data protection guidelines among the health researchers in KEMRI will help in the formulation of policies and guidelines relating to human health research in Kenya.

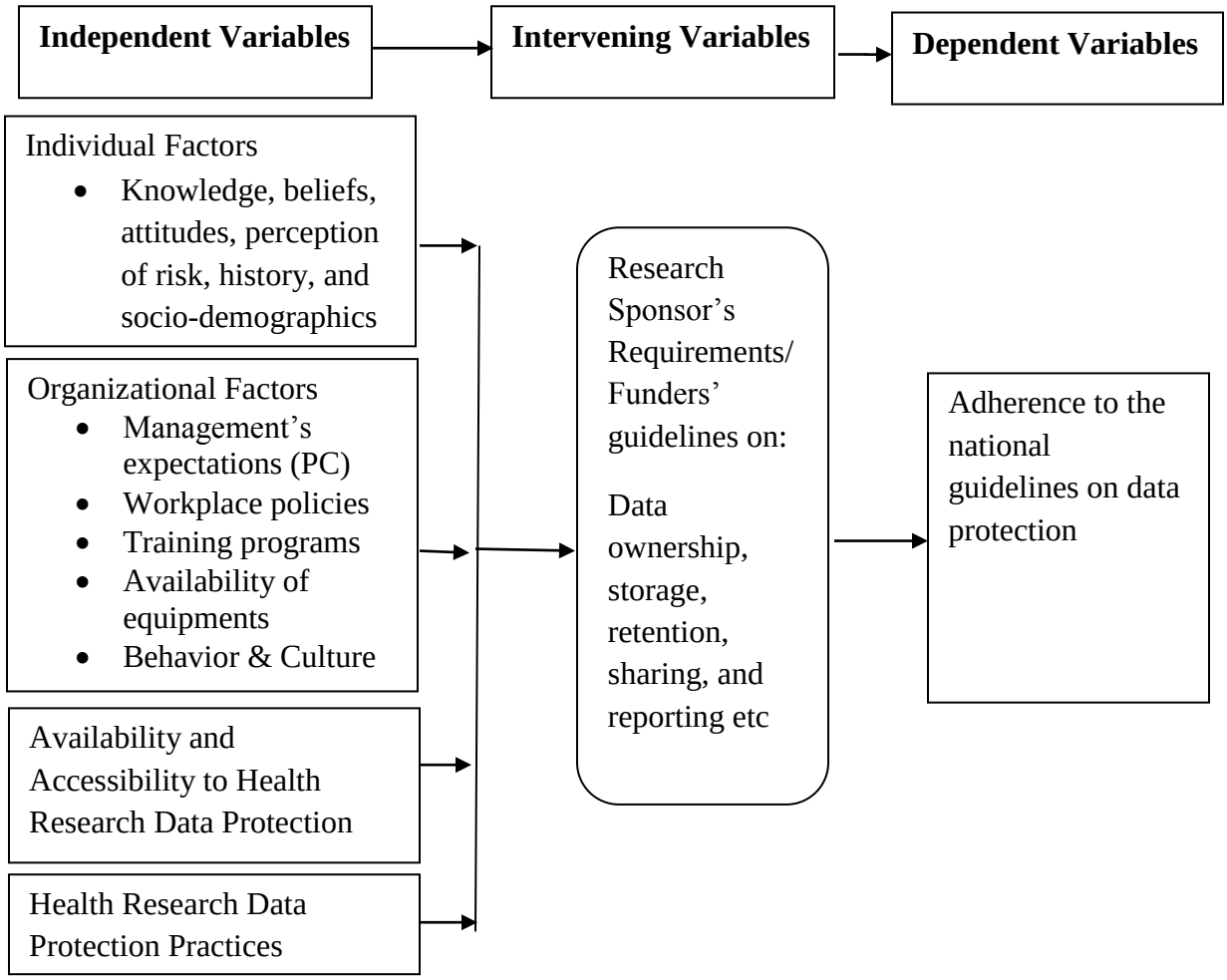
1.8. Delimitations and Limitations

Delimitations are the choices that the researcher makes and they define the boundaries of the study. Geographically, this study was limited to the researchers at the 11 KEMRI research centers in Nairobi, Kilifi, Busia, and Kisumu, Kenya. The researcher visited the target respondents in their respective offices in Nairobi and the satellite centers to collect data. The study did not assess all the processes involved in research, but rather it examined the element of data protection in health research because it is a research area of interest.

Limitations are the influences beyond the control of the researcher. The limitations in this study include the time and financial constraints. The researcher has recommended future studies that can be carried out in this area. Such studies will require more time and financial resources and the future researchers will consider and provide mitigation

measures for the identified constraints. The researcher utilized questionnaire (a combination of interviewer-administered and the self-administered) as the research instruments to collect data. The researcher collected the data from the selected sample of the researcher from the 11 KEMRI centers. This study would have benefited more from mixed methods involving qualitative approaches. The future studies are also recommended to utilize mixed methods approaches.

1.9. The Conceptual and Theoretical Framework



Source: Modified from Kombe et al, (2015)

Figure 1.1: Conceptual Framework

There are several theories that researchers have developed to form a base of the level of adherence to specific guidelines. One theory that is associated with the individual factors influencing adherence to guidelines is the Theory of Reasoned Action (TRA). Martin Fishbein and Icek Ajzen developed this theory in the year 1975. The theory establishes an association 'between behavior and attitudes in a human action'. The intention to perform an action is determined by the behavior (Azjen et al, 1986).

CHAPTER TWO - LITERATURE REVIEW

2.1. Introduction

Khamadi (1992) states that the main aim of carrying out literature review is to assist the researcher undertake his/her project appropriately because researches done before shows crucial facts in the problem area. Literature review increases the researcher's knowledge in the problem area and demonstrates to the rest that he/she is well versed and updated on the research topic.

The purpose of the literature review is to describe the previous publications related to a particular topic. This chapter entails the overview of literature related to the study undertaken by the researcher. The main areas focused are health research, health research data, national guidelines for data protection in health research, and the data protection laws and practices in health research.

2.2. Health Research and Health Research Data

According to the "Health Research Authority" (2015), the term health, clinical or medical research refers to the systematic investigation that aims at understanding the human health. The objective of such research can either be preventive or curative in nature. It is one of the significant ways of boosting people's care and treatment around the globe.

There exists various forms of health research such as clinical trials of medical devices and drugs, operational research, and qualitative studies among others. All types of research involving human participants present different level risks and benefits and risks, hence the various ethical issues that need consideration during the study approval process (Health Research Authority, 2015).

Health research can also be categorized basing on the designs such as the Randomized Controlled Trials (RCT), surveys, case control, and cohort studies among others. The research instruments used in survey studies include the questionnaires, interviews, observations, online surveys, case study, and other qualitative approaches.

The data and information from the research process are important in the improvement of human health. The overall aim of gathering and organizing data is to aid in planning, policy formulation, and decision-making in different health programmes. Ultimately, the health research data is useful in promoting global health outcomes as well as the equity.

According to the "Canadian Institutes of Health Research" abbreviated as CIHR (2016), both the researchers and the stakeholder communities acknowledge the fact that data is an essential output and input element into research as a process. Data is also used to make evidence-based decisions. There has been a global move towards the adoption of 'open' as well as the 'open' data where people around the world share data and other resources. In Canada, the CIHR is mandated to promote the effective and ethical use of health research data and information. They ensure that the available data advances knowledge, expand the available opportunities in research and improving the quality of life as well as the health products and services. CIHR has the mandate to control effective access, analysis, connection, integration, utilization, storage, dissemination, and preservation of health research data (CIHR, 2016).

2.3. National and International Guidelines for Protecting Health Research Data

According to the UNAIDS/WHO (2012) guidance point 18, "researchers have an ongoing obligation to participants to develop and implement procedures to maintain the confidentiality and security of information collected."

The Office of Research Integrity in the United States developed the "Guidelines for Responsible Data Management in Scientific Research." It is crucial for the principal investigators and the research team to understand and address data management challenges. The researcher should take into consideration the data management practices involved carrying out research. They include aspects such as data gathering, ownership, storage, safety, retention, analysis, interpretation, sharing, destruction, and data reporting among others (Coulehan, and Wells, 2014).

The "National Commission for Science, Technology, and Innovation" (NACOSTI) provides the "Guidelines for Ethical Conduct of Biomedical Research Involving Human Subjects in Kenya." Notably, the fifteenth guideline states that "the relevant ethical review committee should determine whether the investigator has put in place mechanisms to protect the safety and to respect the privacy of the research subjects, and to maintain the confidentiality of the data" (NACOSTI, 2004).

2.4. Health Research Data Protection Laws and Practices

The efforts and practices aiming at protection of privacy in the health sector should be taken in a positive and collaborative manner. It should be perceived as an opportunity to advance knowledge sharing and not a hindrance or threat to research. Many rules and regulations governing the protection of health research data promote public trust, hence the willingness to continue providing personal and private information for the research purpose. The laws and policies on data protection are not hindrances to human health research, as some people contend. The rules, regulations, and guidelines provide mechanisms of information access and use for different purposes such as research (Willison, 2007).

In the year 2000, the Canadian Government established a law that governs the use of personal information called the “Protection and Electronic Documents Act” (PIPEDA). This law provided autonomy for the individual citizens to control all aspects of their personal information such as the collection, utilization, and disclosure. However, many researchers in Canada argue that this law hinders them in accessing data for health research (Tu, et al., 2004).

2.5. Factors Influencing Adherence to Data Protection Guidelines in the Health Sector

In Germany, hospitals’ employees are expected to adhere to the laws and regulations on data protection in their daily work. Foth (2016) carried out a survey among hospital workers in Germany to determine the most significant factors that influence the intention of the employees to comply with the data protection guidelines as well as the variance in intention between male and female. The findings showed that the psychological factors such as employee’s attitude, subjective norms and their perceived behaviour control influence the adherence level significantly. It was also evident that the intention to adhere to the data protection policies is significantly different from one gender to another (Foth, 2016).

In the year 2010, a doctor from Columbia University mistakenly exposed confidential clinical information belonging to a Presbyterian patient from the New York City. The data leaked to the Internet through the shared network when the physician attempted to deactivate one of the computer servers in the facility. The “Office for Civil Rights” investigated and reported this unfortunate incident entailing a breach of patients’ privacy and confidentiality. The “New York-Presbyterian Hospital” and the “Columbia

University" share the data about the patients for academic research purposes (Boulton, 2014).

A study conducted by Chua et al. (2017) showed that there are various factors that influence adherence to data protection guidelines in health research. Some of the guidelines reviewed in this study included the 'OECD Protection of Privacy and Transborder Flows of Personal Data of 2013' and the 'General Data Protection Regulations (GDPR)' of the European Union. The first factor that influence adherence to data protection guidelines is the inadequate understanding of the participant's expectations. The study by Chua et al, (2017) stated that Universities need to to understand the privacy expectations of students so that they enact and implement relevant laws that will be adhered to (Chua et al, 2017).

Another study by Da Veiga et al (2019) found out that some Universities in South Africa and Zimbabwe were gathering personal information from students, but they often use it for purposes that was not originally intended. This will lead to the breach of privacy, explicit consenting, and confidentiality requirements set out in the GDPR. The authors of this study revealed that Universities had not instilled an information protection culture that will help in adhering to the data protection guidelines in place.

Another study conducted by Szymkowiak, (2019) revealed that individual factors influence adherence to data protection policies especially in digital marketing. The researcher found out that users may not read the privacy policies because they are available. Users of some websites find there is no problem relating to data protection in sites that have data protection policies in place. People tend to accept the terms and conditions of privacy without even reading them out (Szymkowiak, 2019).

Other studies have found out that the time taken to read and understand privacy policies influence adherence to the set out principles. Studies have revealed that an average policy should be read and understood between 8 and 12 minutes. A reader should not exceed 20 minutes reading privacy policies with 5,000 words (McDonald & Cranor, 2008).

A qualitative study was conducted in Kenya between January and June 2014 to investigate the views of the research stakeholders on the challenges for "Public Health Research Data Sharing in Kenya." The exposure of data occurs through stigmatization, invasion of privacy, disrespecting autonomy and unfair competition. Such exposure exists through either intentional or unintentional 'misuse' of data (Kombe et al., 2015).

2.6. Summary of Literature Review

Previous studies conducted in the area of data protection in health research reveal the challenges in data sharing, ownership, disposal, and storage. The study by Kombe et al, (2015) showed that human participants data may leak to unintended places intentionally or unintentionally. Various factors influencing adherence to data protection guidelines and policies in a global and local scope were pointed out. For instance, the structure and content of the guidelines may influence the adherence because the reader needs clear and understandable texts. However, there is an inadequate understanding on the factors influencing adherence to the national data protection guidelines in Kenya. This study addressed the gaps relating to the adherence to the national guidelines in place. The researcher assessed the organizational and individual factors that influence data protection practices and adherence to the guidelines.

CHAPTER THREE - MATERIALS AND METHODS

3.1. Research Design

This study adopted a descriptive research design to aid in the achievement of the set objectives. The researcher chose this design due to its flexibility and expansiveness in alleviating the potential issues that may arise in the field as the questionnaires are being administered and interpreted. It uses quantitative techniques to collect, analyze and summarize data in this research (Williams, 2007).

In this study, cross-sectional study design was used and it employed quantitative methods of data collection. The data was not collected retrospectively or prospectively, but rather the participants were contacted once and data collected at that specific point in time. The researcher chose cross-sectional design due to its associated low cost, minimal time, and the ability to "capture a specific point in time."

3.2. Variables

Independent variables are the presumed causes (Laura et al, 2014). In this study, the independent variables include the organizational and individual factors influencing adherence to the data protection in health research in KEMRI. Another independent variable is the availability and accessibility to the data protection guidelines in health research.

Dependent variables are the presumed effects of the independent variables in the study (Laura et al, 2014). In this study, the dependent variable is the adherence to the national/NACOSTI guidelines of data protection in health research.

Intervening variables are other variables that influence the effect of independent variable. In this study, the specifications of the research sponsor are the intervening variables.

3.3. Location of the Study

This research was carried out at the "Kenya Medical Research Institute" (KEMRI) in Nairobi County, Kenya. Nairobi centers selected included "Eastern & Southern Africa Centre of International Parasite Control" (ESACIPAC), "Centre for Tradition Medicine and Drug Research" (CTMDR), "Centre for Biotechnology Research and Development" (CBRD), "Centre for Virus Research" (CVR), "Centre for Respiratory Disease Research" (CRDR), "Centre for Microbiology Research" (CMR), "Centre for Public Health Research" (CPHR) and "Centre for Clinical Research" (CCR).

Other centers include the "Center for Global Health Research" (CGHR) in Kisumu, "Centre of Geographical Medicine Research Coast" (CGMRC) in Kilifi, and "Centre for Infectious and Parasitic Diseases Control Research" (CIPDCR) in Busia.

The researcher chose KEMRI centers because it is one of the leading Institutions that carry out health research in Kenya. It is a state corporation mandated by law to carry out research in human health. The researcher was interested in studies involving human health, hence the justification to consider KEMRI as a study site. There is a need to determine the adherence to the data protection guidelines in health research. The results will help in the formulation of data management guidelines as well as the protection of rights, welfare, interests, and the privacy of human participants enrolled in research.

3.4. Study Population

The study population for this study included the scientists at the 11 research centers of the Kenya Medical Research Institute (KEMRI) located in Nairobi, Busia, Kisumu, and Kilifi Counties, Kenya. This specific population was chosen because they carry out research in human health and are required to protect the rights and safety of participants enrolled in research.

Inclusion criteria

1. KEMRI scientists who are employed on permanent and pensionable terms.
2. The study targeted the scientists who have participated in any research project involving human subjects.

Exclusion criteria

1. Researchers who seek approval from KEMRI but work outside the KEMRI centers or not affiliated to the Institute.
2. Participants who did not consent to participate in this study.

3.5. Sampling Techniques for the Research Scientists

The researcher adopted the "stratified sampling method" to identify the sample. In this "probability sampling technique", the entire/whole population were the KEMRI scientists while the centers they belong form the subgroups/strata. The random selection of the final subjects was carried out in a proportionate manner from the various subgroups. The first stage was the "Probability Proportional to Size" (PPS) (Skinner, 2016).

Table 3.1. shows the distribution of the research scientists in KEMRI across the eleven centers (KEMRI Human Resource Department, 2017)

Sample Size Determination

The Fishers' et al, 1998, formula was used to calculate the required sample size;

$$n = \frac{Z2pq}{d2}$$

Where;

n=minimum required sample size

Z=standard normal deviation at 95% CI (1.96)

p=Proportion in the target population approximated as 50% (50% was adopted because the level adherence to data protection guidelines among health researchers is unknown)

d=absolute precision, (0.05)

n=384

Since the number of KEMRI scientists is a finite population and less than 10,000, the overall sample size was determined using the "finite population correction factor" (Yamane, 1967).

$$NF = \frac{n}{1+(\frac{n}{N})}$$

Where n is the sample size per the fisher's et al formula above:

N is the population size, 192

Therefore;

$$NF = \frac{384}{1+(\frac{384}{192})}$$

n= 128

The figure 128 is the baseline number that the researcher intended to reach in collecting the data. However, the researcher considered an extra 10% of the sample size, which was 13 more respondents to cater for the unreturned questionnaires making a total of 141 respondents.

Table 3.1: Distribution of the KEMRI research scientists across the 11 centers

Cluster No.	**Center**	**Total No of Researchers (x)**	**Formula (x/X)*n**	**Number**
1.	ESACIPAC	6	(6/192)*128	4
2.	CCR	28	(28/192)*128	19
3.	CBRD	21	(21/192)*128	14
4.	CVR	27	(27/192)*128	18
5.	CTMDR	21	(21/192)*128	14
6.	CRDR	11	(11/192)*128	7
7.	CPHR	30	(30/192)*128	20
8.	CMR	21	(21/192)*128	14
9.	CGHR	17	(17/192)*128	12
10.	CIPDCR	4	(4/192)*128	3
11.	CGMR-C	6	(6/192)*128	4
Total		**192**		**128**

The final participants were selected randomly from each of the 11 KEMRI centers also called strata in this study.

3.6. Construction of the Research Instruments

Both the self-administered and the interviewer-administered questionnaires were utilized to gather the required facts and figures from the KEMRI scientists. The questionnaire contained five main sections basing on the research questions. The first section was for official purposes and the instructions for filing the questionnaire. The second section contained the general and socio-demographic questions. The third section had the questions relating to the individual and the organizational factors influencing adherence to data protection in health research. The fourth section contained the questions relating to the data protection practices among the KEMRI Researchers in Kenya. The last section contained the questions that focus on the outcome of this study (adherence to data protection guidelines in health research). (*Appendix 2)*

In this section, nine variables were generated and named *Q1, Q2, Q3, Q4, Q5, Q6, Q7, Q8, Q9;* where for each variable; Strongly Agree and Agree were grouped together and coded 1, Strongly Disagree, Disagree and Neutral also grouped together and coded 0, creating series of binary variables. All the nine variables were drawn from the NACOSTI guidelines on data protection.

From the scores variable generated and depicted in Table 4.2, a new variable is generated called Adherence scores where; score of 9/9 is coded as Agree and scores of 8,7,6,5,4,3,2,1,0 is combined and coded as Disagree.

Pre-Testing

Pre-testing of the questionnaires for the researchers was done at the Center for Microbiology Research (CMR). This involved researchers who meet the inclusion criteria of the study.

Validity

Validity of an instrument refers to the extent to which a particular tool truly measures what is supposed to measure (Houser, 2008). Validity of the questionnaires were established through designing questions that ask and answer the attributes that the researcher wanted to investigate (relevant questions). The goals and objectives of research were defined and operationalized clearly.

Reliability

Reliability of an instrument refers to the consistency or the ability of the instrument to produce stable and consistent results (Houser, 2008). For the questionnaires, they should yield the same results when used from time to time by different respondents. To ensure reliability, the researcher was objective in designing the questions. In addition, the questionnaires were pre-tested with a few selected individuals from the Center for Microbiology Research (CMR), KEMRI. The questionnaires were distributed and comments from the respondents were noted for action.

The Cronbach's alpha for the adherence to data protection guidelines in health research is **0.78** which suggests that the adherence scale has relatively high internal consistency. Cronbach's alpha method was used to measure the scale reliability or the internal consistency. Determination of scale reliability was critical since the questionnaire had multiple Likert questions.

3.7. Data Collection Techniques

Cross-sectional data from each of the 11 KEMRI centers was collected at one specific point in time. Data was collected from a representative subset of the population of scientists at each of the KEMRI research centers. A mixture of "open-ended and closed-ended questions" were adopted in formulating questionnaires administered to the selected KEMRI scientists. The researcher visited the scientists in their respective offices within the Nairobi centers and other satellite centers. The researcher adopted both the interviewer and the self-administered questionnaires to the investigators within Nairobi Centers. This technique suited the study since it collected the data to address the specific objectives and provide an association between the identified independent and dependent variables.

3.8. Data Analysis

The data analysis was done using SPSS Version 23. The research combined various descriptive statistics such as the frequency distribution, percentages, mean/averages, and standard deviation. Inferential statistics such as regression and correlation were used to determine the degree of relationship between the independent and dependent variables.

Chi-square test statistic was used in this study. The justification for choosing chi-square was to test the relationship between the different categorical variables. Since this study was driven by a null hypothesis, it was necessary to have a chi-square test statistic to determine if there was a relationship between the categorical variables.

Chi- Square test of independence was used to determine if there is a significant relationship between the organizational factors, individual factors and health research

data protection practices against the adherence to the national data protection guidelines in health research.

3.9. Logistical and Ethical Considerations

The authority to carry out this research was obtained from the Kenyatta University Graduate School (*Appendix 3*) and KU Ethics Review Committee (*Appendix 4*). The proposal was then submitted to the "National Council for Science and Technology and Innovation" (NACOSTI) to get the permit to carry out the study (*Appendix 5 and 6*). The researcher sought the authority to collect data from the Kenya Medial Research Institute (KEMRI) management (*Appendix 7*). Lastly, the researcher sought the consent from the respondents before distributing questionnaires (*Appendix 1*).

Informed consent forms in English were distributed to the potential study participants. The principles of participant's voluntariness and autonomy to make independent decision were upheld.

Participants had a right and freedom to join or withdraw from the research participation. They were not subjected to any cost for their decision to either participate or withdraw from the study.

The researcher ensured the privacy and confidentiality of the enrolled participants using the unique codes. All the information gathered in the course of this research were utilized for the intended research purposes only. The computers, laptops and other storage devices containing the confidential information were secured using strong passwords. The student kept the filled questionnaires in a safe custody until they are destroyed at end of the study.

Community considerations; the community of interest in this research includes the enrolled participants, investigators, and the affiliated organizations. The researcher involved the community in the design of this study. This research is relevant to the participants who have been enrolled in the various studies conducted at KEMRI since their personal information will be kept private and confidential. The research is also relevant to KEMRI and NACOSTI who formulate policies on health research data management. The community of selected researchers will also be consulted in the course of conducting this research. This research will also contribute to the capacity building since it will inform the training needs on health research data management within KEMRI and other research organizations. The research results will be availed to the KEMRI scientists in different forums such as publications, workshops, conferences, and seminars.

CHAPTER FOUR: RESULTS

4.1. Introduction

This chapter provides the study findings or the results. It is organized in terms of the socio-demographic characteristics of the respondents, individual and organizational factors influencing adherence to the data protection guidelines in health research, health research data protection practices, and the adherence scores. Respondents involved in this study are the researchers from KEMRI, Kenya.

4.2. Characteristics of respondents

From the *Table 4.1*, the larger proportion 85(59.4%) of the respondents are males followed by 58(40.6%) females. A total of 63 (44.3%) respondents are specialized/professional graduate while a smaller proportion 3(2.1%) being the University/college diploma. Majority of the respondents 47(32.9%) lie in age group 31-40 followed closely by age group 41-50 at 42(29.4%). A larger proportion 133(93.0%) of the respondents are research officers followed by lab technologist 5(3.5%).

Table 4.1: Socio- demographic characteristics of the participants

Characteristic	**Category**	**Frequency (n=139)**	**Percentage (%)**
Gender	Male	85	59.4
	Female	58	40.6
Highest Level of education	Diploma	3	2.1
	Bachelor's degree	34	23.9
	Specialized/professional graduate or post-graduate diploma	63	44.3
	PhD	42	29.6
Age group in years	< 30	29	20.3
	31 – 40	47	32.9
	41-50	42	29.4
	>50	25	17.5
Job title	Research officer	133	93
	Data manager	1	0.7
	Data analyst	1	0.7
	Student	3	2.1
	Lab technologist	5	3.5

4.3. Adherence levels to data protection guidelines

The respondents with adherence scores of 9/9 means that they agreed or strongly agreed to all the nine questions 18 (12.6%) indicating high adherence to national data protection guidelines as shown in *Table 4.2*. The rest 121 (87.4%) disagreed or strongly disagreed to the nine questions. All the participants who agreed or strongly disagreed to all the nine questions were considered to be fully adhering to the national guidelines on data protection. Those who disagreed, strongly disagreed or responded as neutral to at least one question was considered to have not adhered to the national guidelines on data protection.

Table 4.2: Adherence scores out of the possible 9/9

Adherence scores	**N**	**%**
0/9	2	1.4
1/9	0	0
2/9	3	2.1
3/9	15	10.5
4/9	14	9.8
5/9	23	16.1
6/9	17	11.9
7/9	24	16.8
8/9	27	18.9
9/9	18	12.6

The findings showed that 18 (12.6%) of the respondents adhered to data protection guidelines in health research while the majority showed non-adherence 121 (87.4%) as shown in *Figure 4.1*.

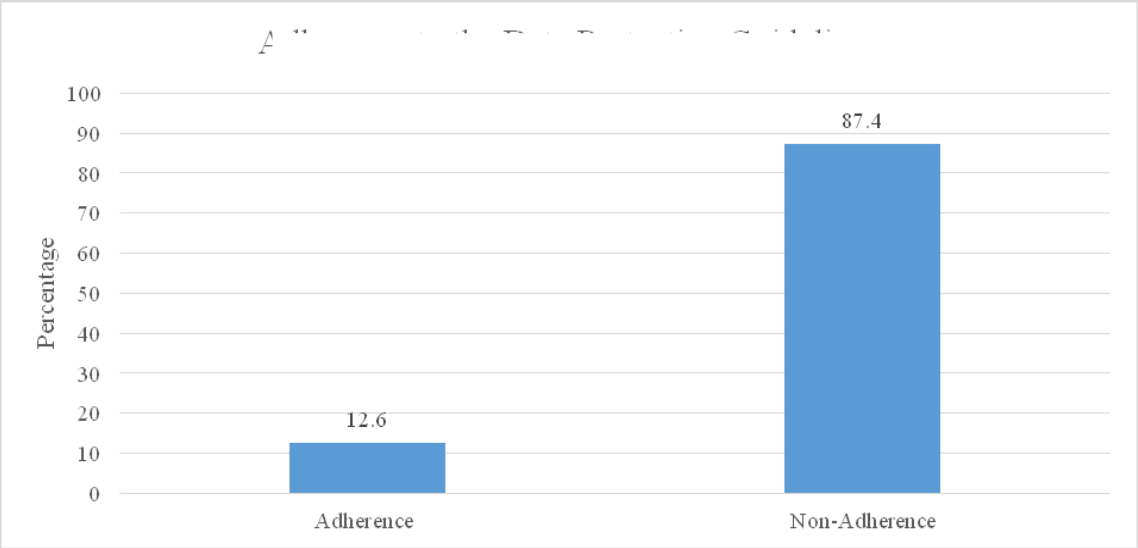


Figure 4.1: Adherence to data protection guidelines

4.4. Individual factors influencing adherence to data protection guidelines

The results showed that the main form in which data may be potentially leaked to the unauthorized individuals is through the use of clinical data/records followed closely with computer readable formats. Use of photographs and test responses does not in any way result to potential data leakage (*Table 4.3*)

Restricting access to the authorized persons and use of codes to conceal participant's identity are the best ways of protecting health research data. Also, data is potentially leaked to the unintended persons at the data sharing stage.

The significant individual factors that influences adherence to the national data protection guidelines among KEMRI researchers are common forms in which data may leak to unintended persons/places (p-value of 0.04) and research stages (p-value of 0.03). All the other individual factors/variables were insignificant. This means that other characteristics such as having done research ethics course and the researcher's thoughts about the major threats to health data were not found to be statistically significant. Any characteristic that reported a p-value of >0.05 was considered statistically insignificant.

Table 4.3: Individual factors influencing adherence to the data protection guidelines

			High Adherence Level		Low Adherence Level				
Characteristic	**Category**	**N**	**N**	**%**	**N**	**%**	**d.f.**	X^2	**p-value**
Research ethics & data course	Yes	98	10	10.2	88	89.8	1	1.40	0.24
	No	40	7	17.5	33	82.5			
The major threats to health research data	Theft	50	7	14	43	86	4	0.13	0.11
	Water spillage	36	0*	0*	36	100			
	Insects	15	2	13.3	13	86.7			
	Excessive heat	27	2	7.4	25	92.6			
	Unauthorized access	12	16	12.6	111	87.4			
Mechanisms for data protection	Restricting access	54	10	18.5	44	81.5	4	1.52	0.06
	Lockable cabinets	57	4	7	53	93			
	Codes	72	10	13.9	62	86.1			
	Strong passwords	67	9	13.4	58	86.6			
	Physical security	35	4	11.4	31	88.6			
The common forms in which data may leak to unintended persons/places	Field notebooks	60	9	15	51	85	7	2.87	0.04
	Filled questionnaires	76	11	14.5	65	85.5			
	Clinical data/records	60	10	16.7	50	83.3			
	Computable readable formats	69	13	18.8	56	81.2			
	Photographs	51	6	11.8	45	88.2			
	Audio visual recordings	46	7	15.2	39	84.8			
	Test responses	41	4	9.8	37	90.2			
	Slides, samples & specimens	28	4	14.3	24	85.7			
Research stages in which data may leak	Data collection	95	7	7.4	88	92.6	4	0.59	0.03
	Data analysis	41	6	14.6	35	85.4			
	Dissemination	28	4	14.3	24	85.7			
	Data sharing	83	12	14.5	71	85.5			
	Data destruction	45	3	6.7	42	93.3			

* Acknowledged as a weakness in chi-square assumptions

4.5. Organizational factors influencing adherence to the data protection guidelines in health research

The *Table 4.4* shows that availability of guidelines or policies on data protection within the institute is the organizational factor which highly influences adherence to data protection with a p-value of 0.01 (this shows that it is highly significant). Respondents were also asked if they had attended any workshop on data management that had been organized by the Institute. The responses indicated that this characteristic is not significant with a p-value of 0.27. Asked whether the Institute avails the necessary equipment and materials to aid in data protection, the respondents indicated that there was no sufficient equipment. This characteristic was not significant with a p-value of 0.20. Institutional Ethics Review Boards (IRB) and Data Safety & Monitoring Boards (DSMBs) clearly do not play a critical role in data protection in health research with a p-value of 0.77(this shows that it is highly insignificant).

Table 4.4: Organizational factors influencing adherence to the data protection guidelines in health research

			High Adherence Level		Low Adherence Level				
Characteristic	**Category**	**N**	**N**	**%**	**N**	**%**	**d.f**	**X^2**	**p-value**
Attendance of workshops on data management.	Yes	105	10	9.5	95	90.5	1	0.98	0.06
	No	37	8	21.6	29	78.4			
The Institute organized the workshop	Yes	84	9	10.7	75	89.3	1	1.23	0.27
	No	27	1	3.7	26	96.3			
Institute avails equipment to aid in data protection	Yes	101	9	8.9	92	91.1	1	1.63	0.20
	No	29	5	17.2	24	82.8			
IRBs and DSMBs play a role in data protection	Agree	73	9	12.3	64	87.7	1	1.13	0.77
	Disagree	16	9	8.3	15	91.7			
There are institutional policies for data protection	Yes	105	7	6.7	98	93.3	1	1.99	0.01
	No	23	6	26.1	17	73.9			

4.6. Health research data protection practices

Table 4.5. shows the different characteristics relating to the data protection practices among the health researchers at KEMRI. The respondents were asked about the different data protection methods that they have employed in their previous research including lockable cabinets, computer memories, and open shelves. This characteristic was not significant as it reported a p-value of 0.07. The second characteristics in the table is the level of data access. They were asked about the people who have access to the data including the researcher, sponsors, and other study team members. This characteristic was considered insignificant as it reported a p-value of 0.06. The third characteristic was the methods that the researchers use to destroy the data including paper shredding, incineration, permanent deletion, or keeping the data for future use. This characteristic

was also insignificant with a p-value of 0.23. The participants were also asked if they anonymize the data or not. A p-value of 0.34 showed that this characteristic was insignificant. The findings showed that the preferred data management practices including electronic and paper-based formats influence adherence to the data protection guidelines (p-value of 0.02) (highly insignificant). Other characteristics such as the use of a data management tool, data analysis softwares, influence of sponsor's guidelines, and whether researchers have reported any case of non-compliance to the data protection guidelines had p-values of 0.33, 0.09, 0.59, 0.33, and 0.06 respectively. All these characteristics were considered insignificant in influencing adherence to the data protection guidelines.

Table 4.5: Health research data protection practices among the KEMRI researchers in Kenya

			High		Low				
Characteristic	**Category**	**N**	**N**	**%**	**N**	**%**	**d.f.**	**X^2**	**p-value**
Data protection methods	lockable cabinets	80	10	12.5	70	87.5	4	2.12	0.07
	password protected hard drives	92	11	12	81	88			
	personal digital memories	31	2	6.5	29	93.5			
	computer memories	50	5	10	45	90			
	open shelves	30	5	16.7	25	83.3			
Levels of data access	Researcher	133	15	11.3	118	88.7	2	1.89	0.06
	Study coordinators	77	9	11.7	68	88.3			
	Sponsors	42	3	7.1	39	92.9			
Common methods of data destruction	paper shredding	60	6	10	54	90	3	2.09	0.23
	Incineration	44	6	13.6	38	86.4			
	keep them for future	47	7	14.9	40	85.1			
	permanent deletion	55	3	5.5	52	94.5			
Anonymization of data at KEMRI	Yes	107	16	15	91	85.1	2	2.15	0.34
	No	12	1	8.3	11	91.7			
	not sure	23	1	4.4	22	95.7			
Most preferred formats of data management	electronic(soft copies)	42	10	23.8	32	76.2	2	1.67	0.02
	paper-based(hard copies)	13	0*	0*	13	100			
	Both	86	8	9.3	78	90.7			
Use of data management tool	Yes	131	18	13.7	113	86.3	1	0.95	0.33
	No	6	0*	0*	6	100			
Commonly used softwares	MS Excel	50	7	14	43	86	3	1.09	0.09
	STATA	39	8	20.5	31	79.5			
	SPSS	44	5	11.4	39	88.6			
	R-Studio	47	2	4.3	45	95.7			
Sponsor provides data protection guidelines	Yes	92	11	12	81	88	1	0.29	0.59
	No	46	7	15	39	84.8			
Sponsor's requirements in line with NACOSTI's	Yes	43	7	16.3	36	83.7	2	2.22	0.33
	No	8	1	12.5	7	87.5			
	Not sure	47	3	6.4	44	93.6			
Failure to adhere to data protection guidelines	Yes	37	1	2.7	36	97.3	2	2.07	0.06
	No	62	12	19.4	50	80.7			
	Don't know	41	5	12.2	36	87.8			

* Acknowledged as a weakness in chi-square assumptions

CHAPTER FIVE: DISCUSSION, CONCLUSIONS AND RECOMMENDATIONS

5.1. Introduction

This chapter provides a detailed discussion of the results in comparison with other similar studies. It also covers the conclusions drawn from the findings generated by the study instruments used. Implications of the study results have also been presented in terms of the recommendations. The areas for further research which this study did not address are also suggested at the end of the chapter.

5.2. Discussion

5.2.1. Adherence to the data protection practices

The findings of this study reported in the previous chapter mean that less than a quarter of the respondents follow the national guidelines to the letter while majority of them adhere partially. Compliance with all the components of guidelines is a way of assuring members of the public that their data is protected and that the research data are credible. Another implication of these key findings relates to the rights, privileges, and welfare of participants enrolled in human health research. The full adherence to the national guidelines on data protection implies that the welfare and rights of human participants are upheld. However, not all types of research may fit to all the principles set out in the national guidelines for conducting biomedical research.

These findings coincide with Henley et al, (2017) who reported that it is difficult for all types of research to adhere to all the requirements and principles set out in the Good Clinical Practice (GCP) guidelines.

These findings also coincide with those of (Chua et al, 2017) that most employees in an organization are not aware of the existence or the specific contents of data protection policies.

5.2.2. Organizational factors influencing adherence to health research data protection guidelines

This study revealed that the availability of the guidelines or policies on data protection within the institute has a high influence on the adherence to data protection (p-value=0.01). It also revealed that the workshops and trainings on data management are highly significant in adherence to data protection guidelines. These findings mean that the organization play a critical role in ensuring that the employees adhere to the available guidelines for data protection. This study also revealed that Institutional Review Boards (IRB) and Data Safety & Monitoring Boards (DSMBs) are highly insignificant in terms of the roles they play in data and human participant's protection in health research (p-value=0.77).

These findings disagree with (Fabiana, 2015) who found out that all the scales of the organizational factors has a low average score. He reported that the availability of materials and equipment at the hospitals, training opportunities, and the management commitment to the safety standards would not influence the adherence to the safety procedures among the nurses.

These findings also disagree with (Neal and Sarwate, 2016) which suggest that onus is on the IRBs to safeguard data breaches. The authors went further to indicate that the researcher must take the full responsibility for the use and management of participant's data.

5.2.3. Individual factors influencing adherence to health research data protection guidelines

Majority of the respondents believe that patient's information leaks in the form of clinical data/records (p-value of 0.02) followed closely by the computer readable formats (p-

value of 0.04). The findings showed that the opinions, perceptions, and attitudes of the researchers in respect to potential data leakage have an influence on the adherence to data protection guidelines. These findings were supported by Neal and Sarwate, (2016) results which revealed that 8% of data breaches in the US hospitals are as a result of the use of Electronic Health Records.

These findings also agree with the survey targeting the hospital employees in Germany which revealed that the intention to comply with data protection standards are influenced by psychological factors such as attitude, subjective norms and perceived behavior control (Foth, 2016).

Other respondents were of the opinion that restricting access to the authorized persons (p-value of 0.04) and use of codes to conceal participant's identity (p-value of 0.04) are the best ways of protecting health research data. These findings agree with (Thomson, 2011) which revealed that several workers of the Berkeley Heart Lab, California, US accessed patient's data without prior authorization and taken the critical information to the competitor in November 2011.

Majority of the respondents believed that data is potentially leaked to the unintended persons at the data sharing stage (p-value of 0.06). These findings agree with (Rathi et al., 2012) which suggests that data sharing presents ethical and social challenges in terms of the protection of the rights and dignity of participants.

The findings also disagree with (Obar and Oeldorf-Hirsch, 2016) who found out that 74% of people who read privacy policies on the website skip them because of the wordy and complicated contents on the site.

5.2.4. Health research data protection practices

The main group of respondents who accesses the stored data is the researcher (p-value of 0.06) followed by the study coordinators (p-value of 0.08). These results mean that the researcher has the overall responsibility in ensuring that the data is protected against unintended use. These findings agree with (Parker and Bull, 2015) who argued that all the research stakeholders including the researchers and sponsors should ensure that interests and rights of participants are respected at all times.

STATA is widely used by the researchers for data analysis and data management (p-value of 0.007). These findings coincide with (Dembe, Partridge, and Geist, 2011) which found out that STATA and SAS were overwhelmingly the most commonly used software applications (in 46.0% and 42.6% of articles respectively) in the US journals.

5.3. Conclusion

In conclusion, there are factors influencing adherence to the data protection guidelines among health researchers in KEMRI, Kenya. The individual factors, organizational influences, and the health research data protection practices has an effect on the adherence to the national guidelines on data protection as discussed in the following sections.

The study findings showed that the organization has a critical role to play in ensuring adherence to the data protection guidelines. The development of institutional guidelines, organization of workshops, and availing the necessary equipment will promote adherence to data protection.

The individual factors such as the views, opinions, and beliefs of the researchers influence the adherence to the data protection guidelines. These individual factors influence the practices and culture of the researchers. The information relating to the data

protection guidelines that the researchers possess is attributed to the ethics courses attended and the experience in conducting research.

The practices of health researchers are driven by the organizational and individual factors. Most researchers keep the participant's data in the open shelves (p-value of 0.06). This practice shows that they are not sensitive to the potential data loss and breach of confidentiality. It was also revealed that most researchers keep the participant's data for future use (p-value of 0.02). This practice may lead to the further loss of participant's privacy and confidentiality.

Lastly, the study revealed that 12.6% of the respondents showed adherence to data protection guidelines in health research while the majority showed non-adherence (87.4%). These findings reveal that most researchers do not comply with the national guidelines on data protection. The non-adherence to the data protection standards may lead to the exposure and leakage of participant's data. It poses danger to the safety, interests, and rights of the participants enrolled in human health research.

5.4. Recommendations

In order to adhere to the national guidelines on data protection, the individual researchers, research organizations, and regulatory bodies should work together at all times. The following suggestions/recommendations promotes adherence to the data protection guidelines;

1. The Institute should invest in the materials and equipment for health research data protection such as lockable cabinets and highly secured softwares.
2. The management should organize workshops and trainings on health research data protection to create awareness among the researchers.

3. The Institute needs to develop and create awareness on the policies and guidelines for data protection.

5.5. Further Research

The following gaps were left out by this study, and they require further investigation through research:

1) A study to compare the institutional, national, and international guidelines on data protection in health research.
2) A research on the adverse effects of data leakage on the privacy and confidentiality of enrolled research participants.
3) A study on the uptake and awareness of the national guidelines on data protection among the researchers in Kenya.

REFERENCES

Azjen, Icek; Madden, Thomas (1986). Prediction of goal-directed behavior: Attitudes, intentions, and perceived behavioral control. *Journal of Experimental Social Psychology*. 22 (5): 453–474.

Boulton, C. (2014). Patient Data Leak Leads to Largest Health Privacy Law Settlement. The Wall Street Journal. The CIO Journal Report. Web. Accessed [10th May 2017]. http://blogs.wsj.com/cio/2014/05/09/patient-data-leak-leads-to-largest-health-privacylaw-settlement/

Canadian Institutes of Health Research (CIHR). (2016). Health Research Data-Advancing Data-Intensive Research. Web. Accessed [17th May 2017]. Available at http://www.cihr-irsc.gc.ca/e/49938.html

Chua, H. N. et al. (2017). Impact of Employees' Demographic Characteristics on the Awareness And Compliance of Information Security Policy in Organizations. *Telematics and Informatics*. 35(6).

Council for International Organizations of Medical Sciences (CIOMS). (2002). International Ethical Guidelines for Biomedical Research Involving Human Subjects. Geneva, Switzerland.

Coulehan, M. B., and Wells, J. F. (2014). Guidelines for Responsible Data Management in Scientific Research. The Office of Research Integrity: US Department of Health and Human Services.

Da Veiga, A. et al. (2019). A conceptual framework for a student personal information privacy culture at universities in Zimbabwe. *Conference: Proceedings of 4th International Conference on the Internet, Cyber Security and Information Systems*, October 2019.

Dembe, A. E, Partridge, J. S, and Geist, L. C. (2011). Statistical software applications used in health services research: analysis of published studies in the U.S. *BMC Health Services Research*. 2011; 11: 252.

Fabiana, O. (2015). Factors influencing adherence to standard precautions among nursing Professionals in psychiatric hospitals. USP, vol.49 no.3, São Paulo, June 2015.

Filkins, B., and Radcliff, D. (2008). Data leakage landscape: Where data leaks and how to apply next generation tools. SANS Institute.

Foth, M. (2016). Factors Influencing the Intention to Comply with Data Protection Regulations in hospitals: Based on Gender Differences In Behaviour and Deterrence. *The European Journal of Information Systems*, Volume 25, 2016 - Issue 2.

Hawkes, B. (2007). Data Protection Guidelines on research in the Health Sector. Irish Data Commisioner's Office.

Health Research Authority (HRA). (2015). What is health research? London.

Henley, P. et al. (2017). Development of the Good Health Research Practice course: ensuring quality across all health research in humans. *Health Research Policy and Systems*. Volume 15, Article number: 28 (2017).

Hochstetler, K. (2009). Research Ethics: Data Management. University of Alaska Fairbanks. Web. Accessed [9th May 2017]. Available at: http://www.uaf.edu/ori/responsible-conduct/data-management/

Hochstetler, K. (2009). Research Ethics: Data Management. University of Alaska Fairbanks. Web. Accessed [12th May 2017]. Available at: http://www.uaf.edu/ori/responsible-conduct/data-management/

Houser, J. (2008). *Scientific Inquiry: Precision, Reliability, and Validity: Essential Elements of Measurement in Nursing Research*. Wiley Periodicals, Inc.

Khamadi, S.I.D. (1992). Using the Library and Writing Research Proposals, Reports and Papers. Eldoret: Moi University Press.

Kenya Medical Research Institute (KEMRI). (2004). Guidelines and Standard Operating Procedures (SOPs).

Kirwan, M et al. (2020). What GDPR and the Health Research Regulations (HRRs) mean for Ireland: "explicit consent"—a legal analysis. *Irish Journal of Medical Sciences*. 2020 Jul 30 : 1–7.

Kombe, F. et al. (2015). Research Stakeholders' Views on Benefits and Challenges for Public Health Research Data Sharing in Kenya: The Importance of Trust and Social Relations. PLoS One. 2015 Sep 2;10(9).

Laura, T. et al, (2014). Independent, Dependent, and Other Variables in Healthcare and Chaplaincy Research. *The Journal of Healthcare Chaplaincy*. Volume 20, Issue 4, 2014.

Laws of Kenya. (2016). Access to Information Act No. 31 of 2016. The National Council For Law Reporting.

Leon, N. et al. (2013). Applying a framework for assessing the health system challenges to scaling Up mhealth in South Africa. *BMC Medical Informatics and decision making at 123*. (12).

McDonald, A. M. and Cranor, L. F. (2008). The Cost of Reading Privacy Policies. *I/S: A Journal of Law and Policy for the Information Society.*

National Commission for Science, Technology and Innovation (NACOSTI). (2004). Guidelines For Ethical Conduct of Biomedical Research, Involving Human Subjects in Kenya. Nairobi, Kenya.

Neal, D. G. and Sarwate, A. D. (2016). Privacy, security, and the public health researcher in the Era of electronic health record research. *The Online Journal of Public Health Informatics*. 2016; 8(3): e207.

Obar, J. A. and Oeldorf-Hirsch, A. (2016). The Biggest Lie on the Internet: Ignoring the Privacy Policies and Terms of Service Policies of Social Networking Services. *Information, Communication & Society*, pp. 1-20, 2016.

Parker, M. and Bull, S. (2015). Sharing Public Health Research Data: Toward the Development of Ethical Data-Sharing Practice in Low- and Middle-Income Settings. *Journal of Empirical Research on Human Research Ethics*, 2015, Vol. 10(3) 217–224.

Rathi V., Dzara K., Gross C. P., Hrynaszkiewicz I., Joffe S., Krumholz H. M., . . Ross J. S. (2012). Sharing of clinical trial data among trialists: A cross sectional survey. British Medical Journal, 345.

Szymkowiak, A. (2019). Managing Data in Digital Marketing. Poznan University of Economics. Dec 2019

Sheikh, A. A. (2008).The Data Protection Acts 1988 and 2003: Some Implications for Public Health and Medical Research. Health Research Board (HRB). Dublin Ireland.

Skinner, C. J. 2016. *Probability Proportional to Size (PPS) Sampling*. Wiley StatsRef: Statistics Reference Online. 1–5.

Thomson, L. (2012). Data Breach and Encryption, Chapter 15: Health Care Data Breaches & Information Security Addressing Threats and Risks to Patient Data. American Bar Association

Trewhella, J. (2014). Research data management policy 2014, Trobe University.

Tu J., Willison D., Silver F., Fang J., et al. (2004). Impracticability of Informed Consent in the Registry of the Canadian Stroke Network. New England Journal of Medicine. 2004; 350(14):1414–22.

UNAIDS/WHO Guidance Document. (2012). Ethical Considerations in Biomedical HIV Prevention Trials.

Williams, C. (2007). Research Methods. *The Journal of Business & Economic Research*. March 2007 Volume 5, Number 3.

Willison, D. J. (2007). Data Protection and the Promotion of Health Research: If the Laws Are Not the Problem, Then What Is? HHealthCare Policy. 2007 Feb; 2(3): 39–43.

World Medical Association (WMA). 2006. Declaration of Helsinki.

Yamane, T. 1967. Statistics, an Introductory Analysis, (2nd Ed.), New York: Harper and Row.

APPENDIX 1: INFORMED CONSENT DOCUMENT

Study Title: Assessing Adherence to Data Protection Guidelines among Health Researchers in KEMRI, Kenya

Investigator	Institution	Study role
Mr. Kebenei Enock Kipchirchir	Kenyatta University	Principal Investigator (Student)
Dr. George O. Otieno	Kenyatta University	Supervisor
Dr. Kenneth Rucha	Kenyatta University	Supervisor

Introduction

I am Mr. Kebenei Enock Kipchirchir, a student at the Kenyatta University pursuing MSc. Health Information Management at the School of Public Health. I am collecting the information outlined in the questionnaire for research purposes only.

Why is the study being done?

The study is being carried out to examine the adherence to data protection guidelines as outlined in the NACOSTI/national document. You are being asked to participate in this study because you are research scientist at the Kenya Medical Research Institute.

What are you required to do?

You will fill a questionnaire that focuses on the individual and organizational factors that influence adherence to data protection guidelines in health research. The whole questionnaire will take approximately 30 minutes to be completed.

Are you under any possible risks by participating in the study?

There is a minimal risk of breach of confidentiality from participating in this study. However, you will not be required to provide your name and will be referred to by your initials only. The filled questionnaire will be kept securely in a lockable cabinet accessed by the Principal Investigator.

What do you stand to gain from the study?

There are no direct benefits to you; however, the knowledge gained will help the KEMRI as an institute as well as the SERU in developing training needs and data protection guidelines. It will also help the research participants in ensuring privacy and confidentiality of their data.

Is there any cost to your participation?

There is no cost to you for participating.

Is there any compensation for participating?

There is no financial compensation for your participation in this research.

Who else will know about your participation?

All information and records relating to your participation in the study will remain confidential. Your name will not be used in any report resulting from this study. The information will not be used for any other purpose other than for this study and will be accessible only to the study investigator. The filled questionnaires will be destroyed at the end of the study. The filled questionnaire will be stored in a lockable cabinet at the researcher's workplace accessed only by the principal investigator.

"Your identity in this study will be treated as confidential. However, any records or data obtained as a result of your participation in this study may be inspected by the by KU ERC and the supervisors.

Are you under any obligation to participate in this study?

You are not under obligation to participate in this study, nor are you obliged to answer any question you do not want to. You may leave the study at any time without any further recourse. If you have understood the information in this consent form, you will be asked to sign before filling in the questionnaire. If there is any portion of this consent document that you do not understand, please ask me before signing.

AVAILABLE SOURCES OF INFORMATION

Any further questions you have about this study will be answered by the Principal Investigator:

Name: Mr. Kebenei Enock Kipchirchir
Phone Number: +254 728 059 048
Any questions you may have about your rights as a research subject will be answered by:
Name: The Chairman, Kenyatta University, Ethics Review Committee
P.O. Box 43884-00100

Nairobi, Kenya.

Consent Form

I have read and understood the information in the consent document and willingly consent to participate in this study. I have been given the opportunity to ask questions concerning the study and all questions have been answered to satisfaction. I understand that I am under no obligation to participate and that I can choose to leave or not answer any question I do not want to.

Participant Name (Printed or Typed):

Date : _____

Participant Signature :

Principal Investigator's Name (Printed or Typed) :

Date: _____

APPENDIX 2: QUESTIONNAIRE FOR THE KEMRI SCIENTISTS

Introduction

I am Mr. Kebenei Enock Kipchirchir, a student at the Kenyatta University pursuing MSc. Health Information Management at the School of Public Health. The general objective of this study is to examine the adherence to the data protection guidelines among health researchers in KEMRI, Kenya.

FOR OFFICIAL PURPOSES ONLY	
Data Collector:	**Questionnaire Code**:
Date Collected: _____	_____
Center: _____	

SECTION A: INSTRUCTIONS FOR FILING IN THE QUESTIONNAIRE AND GENERAL INFORMATION

- ✓ All questionnaires are completed anonymously
- ✓ Please answer all the questions as honestly as possible.
- ✓ Please place a check mark (√) in the box that best answers the question
- ✓ Kindly make only one selection unless otherwise instructed

SECTION B: GENERAL AND SOCIO-DEMOGRAPHIC QUESTIONS

1. What gender are you?

   ☐ Male

   ☐ Female

2. Please indicate your age in the following categories

   ☐ Under 30 years

   ☐ 31 – 40 years

☐ 41 – 50 years

☐ Over 50 years

3. What is the highest level of education you have completed?

☐ University/College Diploma

☐ Bachelor's Degree

☐ Specialized/Professional Graduate or Post-graduate Degree (e.g. MSc, MPH, PhD)

Other (Please specify) _____

4. What is your current job title?

_____

SECTION C: ORGANIZATIONAL FACTORS INFLUENCING ADHERENCE TO THE DATA PROTECTION GUIDELINES IN HEALTH RESEARCH

5. Have you attended any workshop/conference/training on data management in health research?

☐ Yes

☐ No

If yes, please provide a brief description of the training_____

_____

6. If the answer to question (5) is yes, did your Institute organized for the workshop/conference/training?

☐ Yes

☐ No

7. Does the Institute avail the necessary equipment and materials to protect health research data?

☐ Yes

☐ No

If yes, please state the equipment/materials _____

_____

9. Institutional Ethics Review Boards (IRBs) and Data Safety & Monitoring Boards (DSMBs) play a critical role in data and human participant's protection in health research.

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

10. Are there any guidelines or policies on data protection within the Institute?

☐ Yes

☐ No

If yes, please state_____

SECTION D: INDIVIDUAL FACTORS INFLUENCING ADHERENCE TO THE DATA PROTECTION GUIDELINES IN HEALTH RESEARCH

11. Have you done any course on health research ethics that covers data management units

☐ Yes

☐ No

If yes, please provide a brief description of the training_____

_____

12. In your view, what are the major threats to the health research data

☐ Theft

☐ Water spillage

☐ Insects

☐ Excessive heat

☐ Unauthorized access

☐ Others, (Please specify) _____

13. In your opinion, what is the best way of protecting health research data

☐ Restricting access to the authorized persons

☐ Use of lockable cabinets

☐ Use of codes to conceal participant's identity

☐ Use of strong passwords and usernames

☐ Physical security of storage places such as the security guard

☐ Others, (please specify) _____

14. In your opinion, data may potentially leak to the unauthorized individuals in which forms?

☐ Field notebooks

☐ Filled questionnaires

☐ Clinical data/records

☐ Computer readable formats

☐ Photographs

☐ Audio-visual recordings

☐ Test responses

☐ Slides, samples and specimens

☐ Others, (please specify) _____

15. Basing on your experience, data may potentially leak to the unintended persons at what stage of the research process?

☐ Data collection

☐ Data analysis

☐ Data dissemination

☐ Data sharing

☐ Data destruction/archiving

☐ Others, (please specify) _____

SECTION E: HEALTH RESEARCH DATA PROTECTION PRACTICES AMONG THE KEMRI RESEARCHERS IN KENYA

16. Which of the following methods of data protection did you use in your last project work? (You can check more than one box)

☐ Lockable cabinets

☐ Password protected hard drives

☐ Personal Digital Assistance' memories

☐ Computer memories

☐ Open Shelves

☐ Others (Please Specify) _____

17. Who accesses the stored data?

☐ Researcher

☐ Study coordinators

☐ Sponsors

☐ Others (Please Specify) _____

18. Which of the following methods of data destruction do you always apply?

☐ Paper shredding

☐ Incineration

☐ I always keep them for future use

☐ Others (Please Specify) _____

19. I always de-identify/anonymize data and information relating to my research participants.

☐ Yes

☐ No

☐ Not Sure

20. In your opinion, which of the following forms of health research data do you prefer?

☐ Electronic (soft copies)

☐ Paper-based (hard copies)

☐ Both

21. In your previous project, did you use a database management tool that incorporate how data will be stored before and after analysis?

☐ Yes

☐ No

If you used computerized method in data analysis, which software(s) did you use?

☐ MS Excel

☐ STATA

☐ R-Studio

☐ Others (Please Specify) _____

22. Does the research sponsor/donor provide requirements/guidelines for data storage?

☐ Yes

☐ No

If the above answer is yes, do the requirements comply with the NACOSTI guidelines?

☐ Yes

☐ No

☐ Not Sure

SECTION F: ADHERENCE TO THE DATA PROTECTION GUIDELINES IN HEALTH RESEARCH

No.	**Please tell me your level of agreement to the following statements**	**Strongly Agree**	**Agree**	**Neutral**	**Disagree**	**Strongly Disagree**
1.	I always put in place mechanisms to protect the safety of the human participant.					
2.	I always ensure that the privacy of the research participants is respected.					
3.	I always maintain the confidentiality of the human participants' data.					

4.	I always obtain an explicit consent from the enrolled research participants.					
5.	I always ensure that the informed consent contain the plan for maintaining privacy and confidentiality of participants' data.					
6.	I always ensure that there is a medically qualified person responsible for the safety of the enrolled research participants.					
7.	I always ensure that human subjects are protected against social, psychological, or physical harm.					
8.	I have ever experienced a case of participant's data loss or breach of confidentiality.					
9.	I always report any case of data loss or breach of confidentiality to the responsible IRB, Study Sponsor or DSMB.					

THANK FOR TAKING TIME TO COMPLETE THIS QUESTIONNAIRE

APPENDIX 3: APPROVAL FROM KENYATTA UNIVERSITY GRADUATE SCHOOL

KENYATTA UNIVERSITY
GRADUATE SCHOOL

E-mail: dean-graduate@ku.ac.ke
Website: www.ku.ac.ke

P.O. Box 43844, 00100
NAIROBI, KENYA
Tel. 020-8704150

Internal Memo

FROM: Dean, Graduate School **DATE:** 14th September, 2017

TO: **Mr. Kebenei Enock Kipchirchir**
C/o Department of Health Management & Informatics

REF: E55/CE/24760/12

SUBJECT: APPROVAL OF RESEARCH PROJECT PROPOSAL

We acknowledge receipt of your Research Project Proposal after fulfilling recommendations raised by the Graduate School Board of **12th July, 2017.**

You may now proceed with your Data collection, subject to clearance with the Director General, National Commission for Science, Technology & Innovation.

As you embark on your data collection, please note that you will be required to submit to Graduate School completed Supervision Tracking Forms per semester. The form has been developed to replace the Progress Report Forms. The Supervision Tracking Forms are available at the University's Website under Graduate School webpage downloads.

Thank you.

JACKSON LUVUSI
FOR: DEAN, GRADUATE SCHOOL

CC. Chairman, Department of Health Management & Informatics

Supervisors:

1. Dr. George O. Otieno
   C/o Department of Health Management & Informatics
   Kenyatta University

2. Dr. Kenneth Rucha
   C/o Department of Health Management & Informatics
   Kenyatta University

JL/cww

APPENDIX 4: ETHICAL CLEARANCE LETTER FROM KUERC

KENYATTA UNIVERSITY
ETHICS REVIEW COMMITTEE

Fax: 8711242/8711575
Email: kuerc.chairman@ku.ac.ke
kuerc.secretary@ku.ac.ke
Website: www.ku.ac.ke

P. O. Box 43844,
Nairobi, 00100
Tel: 8710901/12

Our Ref: **KU/ERC/ APPROVAL/VOL.1** (143) Date: 14th th June, 2018

Kebenei Enock Kipchirchir
P.O Box 43844-0100
Nairobi

Dear Kebenei Enock,

APPLICATION NUMBER: PKU/807/I873 "ASSESSING ADHERENCE TO DATA PROTECTION GUIDELINES AMONG HEALTH RESEARCHES IN KEMRI, KENYA"

1. IDENTIFICATION OF PROTOCOL

The application before the committee is with a research to **"Assessing Adherence to Data Protection Guidelines Among Health Researches in Kemri, Kenya "** received on 7th February, 2018 and discussed on 12th June, 2018.

2. APPLICANT

Kebenei Enock

3.SITE

Kemri, Kenya

4. DECISION

The committee has considered the research protocol in accordance with the Kenyatta University Research Policy (section 7.2.1.3) and the Kenyatta University Ethics Review Committee Guidelines and **APPROVED that the research may proceed for a period of ONE year from 20th June , 2018.**

5. **ADVICE/CONDITIONS**

i. Progress reports are submitted to the KU-ERC every six months and a full report is submitted at the end of the study.
ii. Serious and unexpected adverse events related to the conduct of the study are reported to this committee immediately they occur.
iii. Notify the Kenyatta University Ethics Committee of any amendments to the protocol.
iv. Submit an electronic copy of the protocol to KUERC.

When replying, kindly quote the application number above.
If you accept the decision reached and advice and conditions given please sign in the space provided below and return to KU-ERC a copy of the letter.

KENYATTA UNIVERSITY OFFICE OF THE CHAIRMAN 03 JUL 2018 ETHICS REVIEW COMMITTEE P.O. BOX 43844 - 00100 NAIROBI

DR. TITUS KAHIGA
CHAIRMAN ETHICS REVIEW COMMITTEE

I KEBENEI ENOCK K. .....accept the advice given and will fulfill the conditions therein.

Signature..... Dated this day of...3RD JULY..... 2018.

cc.

DVC-Research Innovation and Outreach

APPENDIX 5: RESEARCH AUTHORIZATION FROM NACOSTI

NATIONAL COMMISSION FOR SCIENCE, TECHNOLOGY AND INNOVATION

Telephone:+254-20-2213471, 2241349,3310571,2219420
Fax:+254-20-318245,318249
Email: dg@nacosti.go.ke
Website : www.nacosti.go.ke
When replying please quote

NACOSTI, Upper Kabete
Off Waiyaki Way
P.O. Box 30623-00100
NAIROBI-KENYA

Ref: No. **NACOSTI/P/18/87904/23863**

Date: **15th September, 2018**

Enock Kipchirchir Kebenei
Kenyatta University
P.O. Box 43844-00100
NAIROBI.

RE: RESEARCH AUTHORIZATION

Following your application for authority to carry out research on ***"Assessing adherence to data protection guidelines among health researchers in KEMRI, Kenya"*** I am pleased to inform you that you have been authorized to undertake research in **Busia, Kilifi and Nairobi Counties** for the period ending **14th September, 2019.**

You are advised to report to **the Director, Kenya Medical Research Institute, the County Commissioners, the County Directors of Education and the County Directors of Health Services of the selected Counties** before embarking on the research project.

Kindly note that, as an applicant who has been licensed under the Science, Technology and Innovation Act, 2013 to conduct research in Kenya, you shall deposit **a copy** of the final research report to the Commission within **one year** of completion. The soft copy of the same should be submitted through the Online Research Information System.

BONIFACE WANYAMA
FOR: DIRECTOR-GENERAL/CEO

Copy to:

The Director
Kenya Medical Research Institute

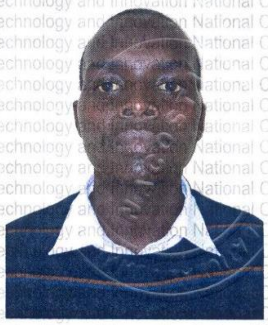
APPENDIX 6: RESEARCH PERMIT FROM NACOSTI

THIS IS TO CERTIFY THAT:
MR. ENOCK KIPCHIRCHIR KEBENEI
of KENYATTA UNIVERSITY, 54840-200
NAIROBI,has been permitted to conduct
research in *Busia , Kilifi , Nairobi*
Counties

on the topic: *ASSESSING ADHERENCE TO DATA PROTECTION GUIDELINES AMONG HEALTH RESEARCHERS IN KEMRI, KENYA*

for the period ending:
14th September,2019

Permit No : NACOSTI/P/18/87904/23863
Date Of Issue : 15th September,2018
Fee Recieved :Ksh 1000



.....
Applicant's
Signature

.....
Director General
National Commission for Science,
Technology & Innovation

APPENDIX 7: RESEARCH PERMIT FROM KEMRI

KENYA MEDICAL RESEARCH INSTITUTE

P.O. Box 54840-00200, NAIROBI, Kenya
Tel: (254) (020) 2722541, 2713349, 0722-205901, 0733-400003, Fax: (254) (020) 2720030
E-mail: director@kemri.org, info@kemri.org, Website. www.kemri.org

KEMRI/RES/7/68/ **10th September, 2018**

Kebenei Enock Kipchirchir
P.O.Box 43844-0100
NAIROBI

Dear Kebenei Enock,

RE: RESEARCH STUDY

Reference is made to your letter dated 4th July, 2018 on the above mentioned Subject.

The Institute has granted you permission to carry out research study on **"Assessing Adherence to Data Protection Among Health Researches in Kemri, Kenya"**

It is noted that you wish to collect data through distributing questionnaires to the KEMRI Scientists.

We would be grateful if you would send us a copy of your findings.

Yours Sincerely,

Rowland Muyeshi
For: DIRECTOR
KENYA MEDICAL RESEARCH INSTITUTE