A CROSS SECTIONAL ANALYSIS OF THE FUSION AND INTEGRATION OF DATA SCIENCE AND HEALTH IN ZIMBABWE

AFRICA UNIVERSITY

A UNITED METHODIST RELATED UNIVERSITY

2025

AFRICA UNIVERSITY

(A United Methodist-Related Institution)

THE FUSION AND INTEGRATION OF DATA SCIENCE AND HEALTHCARE

BY

CHIDO KAMBA

A DISSERTATION SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF BACHELOR OF SCIENCE HONOURS IN COMPUTER SCIENCE IN THE FACULTY OF ENGINEERING AND APPLIED SCIENCES

2025

ABSTRACT

This research focused on identifying and highlighting the significance of integrating data science techniques into the healthcare industry. The research involved exploring applications of data science in healthcare, addressing technical barriers, and developing strategies for successful integration. The collected data was analysed and compared with existing literature to validate the results. Findings showed that while the healthcare industry has had a brief introduction to data science techniques, there is need for improvement on implementing those techniques. Lack of adequate training, awareness, skill and exposure have affected the integration of data science techniques in health. This study is concluded with recommendations on how to effectively integrate and fuse data science and health.

DECLARATION PAGE

I declare that this dissertation is my original work except where sources have been cited and acknowledged. The work has never been submitted, nor will it ever be submitted to another university for the award of a degree.

………………CKamba……………   …………28/03/2025…………

      CHIDO KAMBA   DATE

        (Student)

…………………………………….   ……………………………..

      MR MAKAMBWA   DATE

       (Supervisor)

COPYRIGHT

No part of the dissertation may be reproduced, stored in any retrieval system or transmitted in any form or by any means for scholarly purposes without prior written permission of the author or of Africa University on behalf of the author.

ACKNOWLEDGEMENTS

Firstly, I would like to thank the Almighy God for the gift of life and for His grace and guidance that helped me until the completion of this dissertation.

Secondly, I would like to thank my supervisor, Mr Makambwa for his patience and guidance throughout the process of theis dissertation and his insights that helped me complete this dissertation. Thank you.

Lastly, I would like to thank my parents whose unwavering support kept me going throughout this dissertation and helping me when I needed it the most. I would also like to acknowledge and appreciate their prayers and sacrifices through this journey. Thank you so much.

LIST OF ACRONYMS AND ABBREVIATIONS

RBV – Resource-Based View

Table of Contents

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE: INTRODUCTION

1.1 INTRODUCTION

Data science is driving a transformative revolution in healthcare, ushering in a data-centric approach to patient care. The wealth of information produced by electronic health records, mobile health data, and wearable technology presents an opportunity to extract valuable insights using advanced analytics. This data-driven methodology offers a more unbiased alternative to traditional experience-based approaches in medicine. Its potential impact includes improved diagnoses, optimized resource allocation in hospitals, and potential cost reductions. Notably, clinical decision support systems and predictive analytics serve as prime examples of how data science is already making a difference.

However, integrating data science into healthcare does carry its own challenges. Securing data privacy and ensuring robust security measures are incredibly important. The split nature of healthcare data, stored in various formats across systems, requires seamless integration for comprehensive analysis. Standardized procedures for data collection are vital for accurate analysis and the generation of generalizable findings. Moreover, it is crucial to address potential biases within datasets to avoid preserving health inequalities. Overcoming these challenges is essential to fully unlock the transformative potential of data science in healthcare, leading to a future characterized by personalized and preventative care.

1.2 BACKGROUND TO THE PROBLEM

The healthcare industry faces complex challenges such as increasing costs, the frequency of chronic diseases, and the need for more personalized patient care. Data science offers an opportunity to revolutionize healthcare by making use of advanced analytics and machine learning algorithms to analyze the vast amount of health data available. These techniques can reveal valuable insights buried within massive datasets, enabling healthcare providers to develop predictive models for outcomes like hospital re-admissions and disease progression. By integrating data science into healthcare, providers can proactively intercede and provide individualized treatment plans tailored to each patient's unique needs. This shift to a more data-driven approach holds tremendous promise for improving patient outcomes and optimizing resource allocation.

Furthermore, data science plays a critical role in precision medicine initiatives and clinical decision systems. Technologies like genome sequencing and analysis allow researchers to identify genetic variations that influence an individual's response to treatment, enabling healthcare providers to prescribe more effective treatments with fewer side effects. Clinical decision systems using patient data and advanced algorithms can help physicians make more informed, evidence-based decisions, leading to increased diagnostic accuracy and treatment effectiveness. Data science also contributes to the efficiency of healthcare organizations by

analyzing supply chain management systems, optimizing planning processes, and leveraging insights from large-scale public health data. However, the integration of data science into healthcare also brings challenges that must be addressed, such as ensuring the privacy and security of sensitive patient information, integrating split health information systems, justifying biased data models, and overcoming cultural resistance to a more data-driven and evidence-based decision-making approach.

1.3 <u>STATEMENT OF THE PROBLEM</u>

The healthcare industry is constantly challenged to balance the provision of quality patient care with the need for efficient and cost-effective operations. Traditionally, health decision-making has relied heavily on the expertise and intuition of medical professionals, and the valuable insights that could be gathered from the wealth of information produced in the health ecosystem are often overlooked.

In recent years, the production of electronic health records, mobile devices and remote monitoring systems has led to an explosion in the amount, variety, and speed of health data. However, the healthcare industry has been slow to adopt advanced data approaches, and the disconnect between data availability and analytical execution has created untapped opportunities for improvement. This study aims to address the fundamental question of how healthcare organizations can effectively harness the power of data science to improve patient outcomes, optimize operational efficiency and drive meaningful industry change.

This research explores the following key questions:
1. Identify key applications of data science in health care, such as clinical decision support, predictive analytics, precision medicine, and population health management and research. their potential impact on improving patient care and organizational effectiveness.

2. Explore the challenges and limitations of integrating data science into the health ecosystem, including issues related to data protection, system interoperability, and algorithmic bias, and develop strategies to overcome these barriers.

3. Provides healthcare organizations with a comprehensive roadmap for effective use of data science. It presents best practices, critical success factors, and practical implementation guidelines that can be tailored to the unique needs and constraints of the healthcare industry.

4. We advance the growing body of knowledge at the intersection of data science and healthcare, providing valuable insights and practical recommendations for healthcare professionals, policymakers, and data science practitioners to drive data innovation and change in the field.

Addressing these key questions, this study aims to bridge the gap between the abundance of health data and the underutilization of advanced analytical techniques, enabling health

organizations to harness the full potential of data-driven decision-making and revolutionize the way patients are helped.

1.4 RESEARCH OBJECTIVES

This research study aims to achieve the following:
- Identify the key applications of data science in healthcare and assess their impact
- Examine the challenges and limitations of integrating data science in healthcare
- Develop a comprehensive roadmap for data-driven transformation in healthcare
- Contribute to the academic and industry knowledge on the intersection of data science and healthcare

1.5 RESEARCH QUESTIONS

This research aims to answer the following questions

i) What are the key applications of data science techniques and what is their potential impact on improving patient outcomes, reducing health costs and enhancing operational efficiency?
ii) What are the technical barriers that hinder the effective adoption and integration of data science methodologies in healthcare and how can these challenges be addressed?
iii) What strategies can be used to build data science capabilities and foster effective cross-functional collaboration between the two fields to drive data-driven transformation in the industry?
iv) What key insights and findings from this study can contribute to the academic and industry knowledge on the intersection of data science and healthcare?

1.6 SIGNIFICANCE OF STUDY

This research is important for the following reasons:

1. Improve patient outcomes and quality of care: The healthcare industry is under increasing pressure to improve patient outcomes, reduce medical errors, and improve the overall quality of care. By using data science techniques, healthcare organizations can make more informed, data-driven decisions that improve patient health, reduce hospitalizations, and better manage chronic diseases.

2. Improve operational efficiency and optimize costs: Healthcare systems are constantly struggling with rising costs and the need to optimize their operations. Data science applications can help identify areas of inefficiency, optimize resource allocation, and make management processes more efficient, leading to cost savings and greater financial sustainability.

3. Promotion of innovation and personalized medicine. Integrating data science into healthcare can pave the way for innovative and personalized approaches to patient care. Using insights from health databases, providers can develop personalized treatment plans, predict disease risk, and provide more targeted and effective interventions.

4. Addressing health issues and inequalities. Data science can play a key role in identifying and addressing health disparities, particularly among underserved populations. By utilizing predictive modelling and population health management techniques, organizations can better understand and address the social determinants of health, improve access to care, and work to reduce health disparities.

5. Informing Policy and Regulatory Decisions: The findings and recommendations of this study can provide valuable data to policymakers, regulators, and health managers to make more informed decisions about the implementation and management of data-driven initiatives. in the health sector.

6. Advancing Academic and Industry Knowledge: This research contributes to the existing knowledge base at the intersection of data science and healthcare by providing a comprehensive understanding of the opportunities, challenges and best practices associated with the integration of these fields. The insights and practical guidelines generated can inform future academic research and guide healthcare organizations and data science practitioners in their efforts to manage data changes.

Addressing these important aspects in this research on the integration of data science in health care can significantly improve patient care, operational efficiency, and overall health system performance, ultimately benefiting both health care providers and the communities they serve.

1.7 <u>ASSUMPTIONS</u>

This study on the integration of data science in the healthcare industry is based on several basic assumptions.

First, healthcare organizations are expected to have access to sufficient quality, structured and unstructured data, including patient records, clinical data, administrative data, and other relevant data sets. The study also assumes that these various data sources can be effectively integrated and curated to enable the application of data science techniques.

Another important assumption is that healthcare organizations have the technical capabilities and infrastructure to implement data science solutions, such as advanced analytics platforms, machine learning algorithms, and data visualization tools. The study also assumes a certain organizational readiness and commitment to data-driven decision-making, including the

availability of skilled data science and analytics professionals and the willingness and ability to adapt to new ways of working.

Furthermore, the study assumes that health data science applications can be developed and deployed within the existing regulatory framework, including data protection, security, and ethical guidelines, without significant legal or enforcement barriers. It also assumes that healthcare professionals, data science practitioners, and other key stakeholders such as patients and decision makers are willing to actively collaborate and participate in the integration of data science in healthcare.

Furthermore, the study anticipates that the insights, best practices, and recommendations of this study can be broadly applicable and transferable to different health settings, organizations, and contexts with appropriate adaptations and consideration of local factors. Finally, the study predicts that the field of data science and its applications in healthcare will continue to evolve with the continuous development of algorithms, computing power, and data management capabilities, increasing the potential for data-driven healthcare transformation.

Knowing and considering these assumptions during research increases the robustness and reliability of research findings and recommendations and ensures that the insights generated are based on realistic and well-considered assumptions.

1.8 HYPOTHESIS

The four key hypotheses of this study on the integration of data science in the healthcare industry are:

Hypothesis 1: The effective integration and application of data science techniques in the healthcare sector can lead to significant improvements in patient outcomes and care quality.

Hypothesis 2: The application of data science in operational and administrative processes can help healthcare organizations optimize their efficiency and cost management, improving the financial sustainability of the healthcare system.

Hypothesis 3: The insights and discoveries generated through data science can foster innovation and the development of more targeted, personalized interventions and personalized medicine in the healthcare industry.

Hypothesis 4: Leveraging data science techniques to analyse population-level data and social determinants of health can help healthcare organizations better understand and address the root causes of health disparities, leading to more equitable access to care and improved health outcomes for underserved communities.

1.9 <u>DELIMITATIONS</u>

Here are six main limitations of data integration in healthcare:

1. Scope: The research focuses specifically on healthcare and does not examine the integration of data science in other sectors.

2. Timeline: The analysis and conclusions of the study are based on data available up to a certain date since the assistant database was last updated at that time.

3. Geographic focus: The study does not make clear comparisons or generalizations between different geographic areas or health systems. In particular, the integration of data science into the context of the researcher's own country or region is considered.

4. Data availability: The study is limited by the availability and availability of data related to healthcare and data science, and restrictions or privacy concerns related to the use of sensitive patient data.

5. Expert Limitations: Although the research uses healthcare and data science experts, the researcher's knowledge is limited to the information available at a particular time and may not reflect the latest advances or perspectives in the field.

6. Outcome Measures: The study focuses primarily on evaluating the impact of data science integration on patient outcomes, operational efficiency, and health equity, and may not cover all possible outcome measures or side effects.

1.10   <u>LIMITATIONS</u>

Integrating data science with healthcare has six limitations of research:

1. Data quality and completeness: The quality, completeness and reliability of data related to healthcare and data science can limit research. available for analysis. Inadequacies or inconsistencies in the data may affect the accuracy and generalizability of the findings.

2. Technical Limitations: The analysis and conclusions of the study may be affected by the current state of data science technologies, tools, and techniques, which can evolve rapidly over time. More recent advances not accounted for in the study may limit the applicability of the results.

3. Organizational and Cultural Factors: Various organizational and cultural factors such as leadership support, employee acceptance, and data literacy can influence the

successful integration of data science into healthcare. Studies may not fully capture the nuances of these contextual factors in different healthcare settings.

4. Generalizability: Given the unique characteristics of health systems and organizations, the generalizability of study findings to other geographic areas or health services with different regulatory frameworks, resource constraints, or demographics may be limited.

5. Ethical and Privacy Considerations. The analysis and recommendations of the study must address the complex ethical and privacy issues associated with the use of patient data and the application of data-driven technologies in health care. These factors may place some limitations on the scope and depth of research.

6. Sustainability and scalability: Research may not fully address the long-term sustainability and scalability challenges associated with integrating data science into healthcare, such as the need for ongoing investment, talent acquisition, and change of management.

CHAPTER TWO: REVIEW OF RELATED LITERATURE

2.1 INTRODUCTION

An extensive amount of research has focused on the integration of data science in healthcare, making it a critical area of innovation with the potential to transform patient care, operational efficiency, and the healthcare ecosystem. This literature review aims to provide a comprehensive overview of the knowledge and research on the application of data science techniques within the healthcare sector.

The review establishes the background and relevance of data science in healthcare. It highlights the increasing volume, variety, and velocity of healthcare data driven by advancements in electronic medical records, diagnostic technologies, and connected devices. This data explosion presents both opportunities and challenges, necessitating the development of advanced analytical capabilities to extract meaningful insights and drive informed decision-making. Furthermore, the review delves into the specific applications of data science in healthcare. It explores how predictive analytics, machine learning, and other data-driven techniques can enhance patient outcomes, improve disease prevention and management, and optimize resource allocation and operational efficiency. The review also discusses the potential of data science to foster healthcare innovation and personalized medicine, addressing issues such as targeted therapies, biomarker identification, and interventions customized to individual patient needs.

By synthesizing existing research and empirical evidence, this literature review provides a solid foundation for understanding the current state of data science integration in healthcare, the realized and potential benefits, and the key considerations for successful implementation. The insights gained from this review will inform the design and execution of the proposed study, ensuring that it builds upon the existing knowledge and contributes meaningful insights to the evolving field of healthcare data science.

2.2 THEORETICAL FRAMEWORK

1. Systems Theory: The systems theory is essential to understand the integration of data science and the healthcare industry, as it recognizes the intrinsic complexity and the interconnectedness of the healthcare ecosystem. This theory emphasizes that healthcare organizations do not operate in isolation, but are part of a larger interdependent system. This system includes patients, providers, payers, regulators, and other stakeholders.

   One needs to understand how the different components of the healthcare ecosystem interact and influence each other in order to successfully integrate data science. This involves examining the flow of information, the alignment of incentives, the coordination of workflows, and the feedback mechanisms that form the overall performance of the system.

This systems theory, when applied to this proposed study, will explore how the fusion of data science can optimise the relationships and interactions within the healthcare system, leading to improved patient outcomes, enhanced operational efficiency, and better resource allocation. This study will also consider how external factors such as technological advancements, can impact the integration and effectiveness of data science initiatives.

2. Resource-Based View (RBV): The resource-based view of the firm provides a valuable lens for understanding how healthcare organizations can leverage their data and analytical capabilities as strategic resources to gain a sustainable competitive advantage.

   According to the RBV, an organization's ability to generate superior performance and outcomes is dependent on its possession of valuable, rare, inimitable, and non-substitutable resources and capabilities. In the context of healthcare, the effective integration of data science can be considered a strategic resource that enables organizations to generate unique insights, improve decision-making, and deliver better patient care.

   The proposed study will examine how healthcare organizations develop and deploy their data science capabilities, including the necessary people, processes, and technologies, to create value and outperform their competitors. The study will also investigate the organizational factors, such as leadership, culture, and knowledge management, that contribute to the effective utilization of data science as a strategic resource.

   By incorporating the RBV, the study will provide insights into the specific data science-related capabilities and resources that healthcare organizations need to develop and sustain a competitive edge in an increasingly data-driven industry. This understanding will inform the design of data science integration strategies and the allocation of resources to maximize the impact on patient outcomes and organizational performance.

   The integration of these two key theoretical frameworks – systems theory and the resource-based view – will provide a comprehensive and robust foundation for the proposed study, enabling a nuanced exploration of the complex dynamics involved in the successful integration of data science within the healthcare industry.

2.3 RELEVANCE OF THE THEORETICAL FRAMEWORK TO THE STUDY

   i. Systems Theory:

The systems theory is highly relevant to the proposed study as it provides a holistic perspective on the integration of data science within the healthcare industry.

This theory recognizes that healthcare organizations do not operate in isolation, but are part of a larger, interconnected system that includes patients, providers, payers, regulators, and other stakeholders.

By applying systems theory, the study will be able to examine how the integration of data science influences and is influenced by the various components of the healthcare system, such as clinical workflows, administrative processes, and patient-provider interactions.

This theoretical lens will enable the researchers to understand the feedback loops, information flows, and unintended consequences that may arise from the integration of data science, ensuring a comprehensive analysis of the phenomenon.

The systems theory will guide the study to consider the contextual factors and external influences that shape the integration of data science, such as regulatory changes or technological advancements in the industry.

Overall, the systems theory is crucial for the proposed study as it provides a framework to explore the complex, dynamic, and interdependent nature of the healthcare ecosystem and how the integration of data science can optimize the relationships and interactions within this system.

ii. Resource-Based View (RBV)

The resource-based view of the firm is highly relevant to the proposed study as it provides a valuable perspective on how healthcare organizations can leverage their data and analytical capabilities as strategic resources to gain a sustainable competitive advantage.

This theory will inform the study's examination of the specific data science-related resources and capabilities that healthcare organizations need to develop and deploy effectively, including the necessary people, processes, and technologies.

The RBV will guide the analysis of how the integration of data science, supported by complementary organizational resources and capabilities, can lead to superior performance and improved patient outcomes.

By incorporating the RBV, the study will provide insights into the organizational factors, such as leadership, culture, and knowledge management, that contribute to the effective utilization of data science as a strategic resource.

This theoretical framework will help the researchers understand how healthcare organizations can create unique and valuable data science-based capabilities that are difficult for competitors to replicate, thereby sustaining their competitive edge in the industry.

The RBV is particularly relevant as the healthcare industry becomes increasingly data-driven, and the effective integration of data science can be a critical differentiator for organizations seeking to improve patient care and organizational performance.

2.4 <u>SUMMARY</u>

By integrating these two key theoretical frameworks – systems theory and the resource-based view – the proposed study will be able to provide a comprehensive and robust understanding of the complex dynamics involved in the successful integration of data science within the healthcare industry. These theories will guide the research design, the selection of relevant variables and metrics, and the interpretation of the study's findings, ensuring that the conclusions drawn are grounded in well-established theoretical principles and contribute to the advancement of knowledge in this critical field.

With the study grounded in these two complementary theoretical frameworks, the researchers will be able to provide a comprehensive and refined understanding of the complex dynamics involved in the successful integration of data science within the healthcare industry. The insights gained will inform the research design, variables, and interpretation of findings, contributing to the advancement of knowledge in this critical field.

CHAPTER THREE: METHODOLOGY

3.1 INTRODUCTION

The investigation of the integration and fusion of data science and health requires a rigorous and systematic research methodology. This chapter provides an overview of the approach that will be applied to address the research objectives and answer the research question. The methodology encompasses the research design, data collection methods, and the analytical techniques used to derive meaningful insights.

By adopting a mixed-methods approach and employing a range of data collection and analysis techniques, this methodology aims to provide a comprehensive understanding of the integration and fusion of data science and health. It offers a robust foundation for exploring the challenges, opportunities, and potential benefits of leveraging data science in healthcare settings.

3.2 THE RESEARCH DESIGN

To investigate the integration and fusion of data science and health, a well-designed research methodology is crucial.

The research design for this study will adopt a mixed-methods approach, combining qualitative and quantitative data collection and analysis techniques. This approach offers a comprehensive understanding of the complex interplay between data science and health, capturing both subjective experiences and objective measurements. The qualitative component will involve interviews and focus groups with healthcare professionals, data scientists, and relevant stakeholders. These qualitative data will provide insights into the challenges, opportunities, and best practices in integrating data science into healthcare settings. On the other hand, the quantitative component will involve the analysis of large-scale datasets obtained from electronic health records, wearables, and other relevant sources. Statistical analysis and data mining techniques will be employed to identify patterns, correlations, and trends in the data.

The quantitative phase of the research design will involve the analysis of large-scale datasets obtained from electronic health records, wearables, and other relevant sources. This phase aims to quantify and measure the impact and effectiveness of data science integration in healthcare. Statistical analysis, data mining techniques, and machine learning algorithms will be utilized to identify patterns, correlations, and trends within the data. This quantitative analysis will provide empirical evidence and objective insights into the potential benefits and challenges associated with data science integration.

By employing a sequential explanatory mixed-methods research design, this methodology provides a comprehensive and nuanced understanding of the integration and fusion of data science and health. It allows for the exploration of both subjective experiences and objective measurements, ensuring a holistic perspective on the topic.

3.3 POPULATION AND SAMPLING

When it comes to population sampling, it is crucial to carefully select the participants who will be included in the study. The population of interest for this research on the integration and fusion of data science and health consists of healthcare professionals, data scientists, and other relevant stakeholders involved in the implementation and utilization of data science in healthcare settings.

To ensure a representative sample, purposive sampling will be employed. This sampling technique involves deliberately selecting individuals who possess the necessary expertise and experience relevant to the research topic. The goal is to include a diverse range of participants who can provide valuable insights into the integration of data science and health from various perspectives.

Healthcare professionals from different disciplines, such as doctors, nurses, and administrators, will be approached to participate in the study. Data scientists who specialize in healthcare analytics and have experience working with healthcare data will also be included. Additionally, representatives from relevant stakeholders, such as policymakers and healthcare technology vendors, will be invited to participate to gain insights into the broader context of data science integration in healthcare.

The sample size will be determined based on the principle of data saturation, where data collection is continued until no new information or themes emerge during the qualitative phase. This ensures that a sufficient depth of understanding is achieved without unnecessarily extending the data collection process.

By employing purposive sampling and selecting participants with the necessary expertise and experience, the study aims to gather rich and diverse perspectives on the integration and fusion of data science and health. This approach ensures that the findings and insights derived from the research are representative of the population of interest and can provide valuable insights for future implementation and decision-making in the field.

3.4 DATA COLLECTION INSTRUMENTS

Data collection will be carried out through purposive sampling, ensuring representation from various healthcare institutions and data science experts. Semi-structured interviews and focus groups will be conducted to capture rich and diverse perspectives. The interviews will be audio-recorded and transcribed verbatim to facilitate thematic analysis. The quantitative data will be collected from existing datasets, ensuring adherence to ethical and privacy regulations. Data cleaning and preprocessing techniques will be applied to ensure data quality and consistency. Statistical software packages such as R or Python will be used for data analysis, employing appropriate statistical tests, machine learning algorithms, and visualization techniques to uncover meaningful insights.

a. For the qualitative phase of the study, semi-structured interviews and focus groups will be conducted. An interview guide will be developed, consisting of a set of open-ended questions and prompts. These questions will allow participants to share their experiences, perspectives, and challenges related to the integration of data science in healthcare. The interviews will be audio-recorded to ensure accurate data capture, and detailed notes will be taken to supplement the recordings. This combination of audio recordings and notes will facilitate the transcription and analysis of the qualitative data.

b. In addition to interviews, focus groups will be conducted to foster group discussion and generate insights through participant interaction. The focus group sessions will be guided by a moderator who will facilitate the discussion using a predefined set of topics and questions. The interactions and discussions among participants will provide valuable information about the integration of data science and health from a collective standpoint.

c. For the quantitative phase, data will be collected from existing datasets such as electronic health records, wearables, and other relevant sources. Data extraction tools and techniques will be used to gather the necessary variables and measures for analysis. These datasets will provide objective and quantitative information about the impact and effectiveness of data science integration in healthcare settings.

d. By employing a combination of semi-structured interviews, focus groups, and quantitative data collection from existing datasets, this research aims to gather a comprehensive range of data to support a thorough analysis of the integration and fusion of data science and health. These data collection instruments will enable the capture of both subjective experiences and objective measurements, contributing to a holistic understanding of the research topic.

3.5 <u>DATA COLLECTION PROCEDURE</u>

Here is an overview of the data collection procedure for the research on the integration and fusion of data science and health:

1. Participant Recruitment: A purposive sampling method will be utilized to select participants who have expertise and experience relevant to the research topic. Healthcare professionals, data scientists, and relevant stakeholders will be approached and invited to participate in the study. Participants will be provided with information about the research purpose, confidentiality, and voluntary nature of their involvement.

2. Informed Consent: Prior to data collection, participants will be required to provide informed consent, indicating their willingness to participate in the study. The informed consent process will ensure that participants understand the purpose of the research, their rights as participants, the confidentiality of their responses, and their ability to withdraw from the study at any time.

3. Data Collection Instruments:

Semi-Structured Interviews: One-on-one interviews will be conducted with individual participants. An interview guide will be developed, consisting of open-ended questions and prompts related to the integration of data science and health. Detailed notes will also be taken during the interviews to supplement the recordings.

4. Data Analysis: The collected data, both qualitative and quantitative, will be analysed using appropriate methods. The qualitative data from interviews and focus groups will undergo thematic analysis to identify recurring themes and patterns. The quantitative data will be subjected to statistical analysis and data mining techniques to identify correlations, trends, and insights related to the integration of data science and health.

5. Data Privacy and Confidentiality: Throughout the data collection process, utmost care will be taken to ensure the privacy and confidentiality of the participants' data. All collected data will be securely stored and accessible only to the research team. Any identifying information will be anonymized or removed to maintain participant confidentiality.

By following this data collection procedure, the research aims to gather comprehensive and reliable data that will contribute to a thorough analysis of the integration and fusion of data science and health. The procedure emphasizes ethical considerations, participant consent, and data privacy to ensure the integrity and validity of the research findings.

3.6 ANALYSIS AND ORGANIZATION OF DATA

Once the data has been collected for the research on the integration and fusion of data science and health, it is essential to analyse and organize it in a systematic and meaningful way. This process involves transforming the raw data into meaningful insights and structuring it for further analysis and interpretation.

Firstly, the qualitative data from the interviews and focus groups will undergo a process of thematic analysis. This analysis involves identifying recurring themes, patterns, and categories within the data. The recorded interviews and focus group discussions will be transcribed, and the transcripts will be carefully reviewed. By coding the data, either manually or using software tools, key themes and patterns related to the integration of data science and health will be identified. These themes will be organized and grouped together to provide a comprehensive understanding of the qualitative data.

For the quantitative data collected from existing datasets, statistical analysis and data mining techniques will be employed. This analysis aims to uncover patterns, correlations, and trends within the data. Statistical software packages such as R or Python will be used to perform descriptive statistics, inferential statistics, and data visualization. This analysis will provide quantitative insights into the impact and effectiveness of data science integration in healthcare settings.

Once the data has been analysed, the next step is to organize it in a structured manner. This can be achieved by creating a data matrix or database that includes relevant variables and measures. The data matrix will allow for easy comparison and cross-referencing of data points. It is important to ensure that the data is organized in a logical and coherent manner to facilitate further analysis and interpretation.

Furthermore, it is essential to maintain clear documentation throughout the data analysis and organization process. This includes keeping track of the steps taken in the analysis, the software and tools used, and any decisions made during the process. This documentation will enable transparency, reproducibility, and the ability to revisit and validate the analysis if needed.

By conducting rigorous analysis and organizing the data in a structured manner, the research aims to derive meaningful insights and provide a solid foundation for interpreting the integration and fusion of data science and health. This process ensures that the data is effectively utilized to address the research objectives and answer the research question.

3.7 ETHICAL CONSIDERATION

Respecting the rights and safeguarding the wellbeing of the participants is key. Here are some ethical considerations to be taken into account are:

1. Informed Consent: Before data collection, the participants will be fully informed about the purpose, procedures, potential risks, and benefits of the study and the participants will also have the opportunity to ask questions. The informed consent process will ensure that participants know their rights, including the right to withdraw from the study at any time without consequences.

2. Confidentiality and Anonymity: The participants' identities and personal information will be kept confidential and will only be accessed by the research team. For anonymity, any personal identifiable information will not be disclosed and any identifying data from data sets will be removed.

3. Privacy Protection: The participants' privacy will be protected by ensuring the data is securely stored and only accessed by researcher.

4. Data Security: Suitable measures will be taken to prevent the loss, theft or unauthorized disclosure of data collected.

5. Ethical Review and Compliance: The research study will adhere to the ethical guidelines and regulations set by Research Ethics Committee and researcher shall obtain approval from ethics review board.

3.8 SUMMARY

In crafting the research methodology for this study on data science integration in healthcare, our team has taken a thorough and thoughtful approach to explore how data science techniques can be applied within the healthcare sector. Our goal is to delve into healthcare data, seeking valuable insights that can ultimately enhance patient outcomes, improve disease prevention strategies, and boost operational efficiency in healthcare settings. Paramount to our research is the safeguarding of patient privacy and confidentiality. We achieve this by meticulously anonymizing data and ensuring that all participants provide informed consent, thereby upholding the highest ethical standards.

Additionally, our methodology places a strong emphasis on transparency when presenting research findings and interpreting data. It is our commitment to maintain trust and credibility throughout our research process. We aspire to not only glean insights that can drive positive changes in healthcare practices for the benefit of patients but also to remain objective and unbiased in our data analysis. By upholding ethical principles every step of the way, our study aims to make meaningful contributions to the field of healthcare data science while prioritizing the rights and well-being of those involved in our research.

CHAPTER FOUR: DATA REPRESENTATION, ANALYSIS AND INTEPRETATION

4.1 INTRODUCTION

This chapter gives a full representation, analysis and interpretation of research findings. The primary aim of the study is to investigate the factors influencing the healthcare industry's adoption of data science techniques to improve patient care and operational efficiency. Data collection involved the distribution of web-based questionnaires to healthcare practitioners. The use of Google Forms for questionnaire administration provided a practical approach, offering both summarized and individual responses for the researcher to analyse accurate statistics. The data is visually presented in the form of pie charts and aligns with the research objectives indicated in chapter one to derive meaningful insights for improving healthcare practices through data science integration.

4.2 RESPONSE RATE

A total of 24 questionnaires were sent out to selected individuals via email and other means, and a total of 18 participants responded to the questionnaire. 6 questionnaires were not responded to. The formula used to calculate the response rate is as follows:

$$\text{Response Rate} = \frac{Total\ number\ of\ questionnaires\ received}{Total\ number\ of\ questionnaires\ issued}$$

Table 1. Questionnaire Response Rate

Selected Participants	Questionnaires issued	Questionnaires returned	Response Rate	Response Rate %
All	24	18	0.75	75%

4.2.2 Demographic Information of Respondents

Table 2. Professions of Participants

PROFESSION	PERCENTAGE OF PARTICIPANTS	NUMBER OF PARTICIPANTS
Medical Doctor	22.22%	4
Laboratory scientist	5.56%	1
Surgeon	16.67%	3
Technology expert	11.11%	2
Urologist	5.56%	1
Specialist in Medicine	11.11%	2

Medical Director	5.56%	1
Nurse and Researcher	22.22%	4
Public Health	5.56%	1

Figure 1. Professions
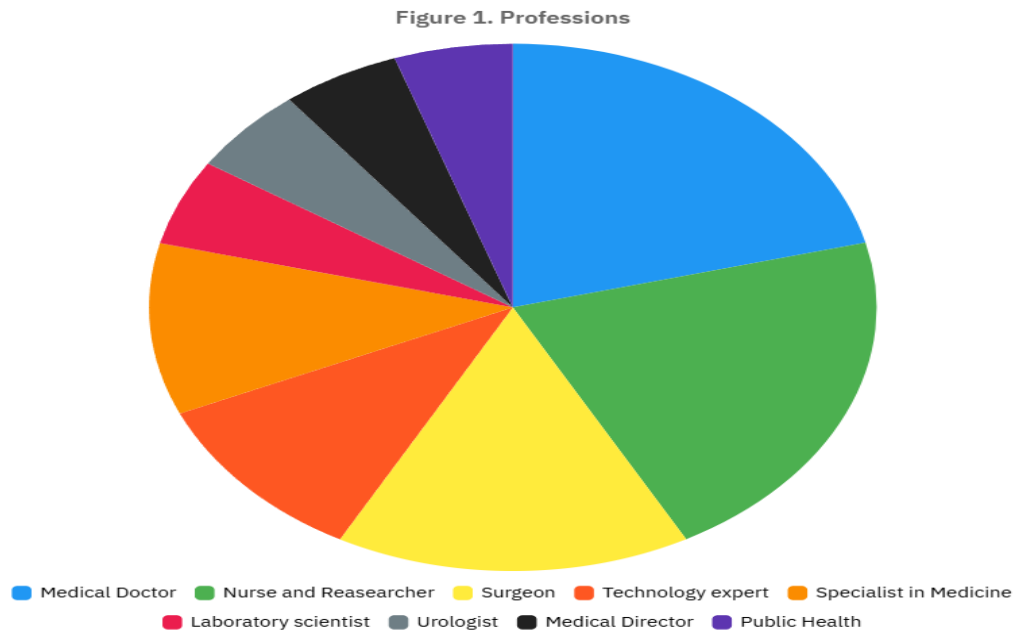
4.3 DATA PRESENTATION AND ANALYSIS

4.3.1 Key Applications of Data Science Techniques in Healthcare

Figure 2. Key Applications of Data Science Techniques currently in use or being explored

Figure 2 shows that 44.4% of the respondents use or are exploring the use of predictive analysis in health while 22.2% are using or exploring all techniques. 11.1% respondents are using machine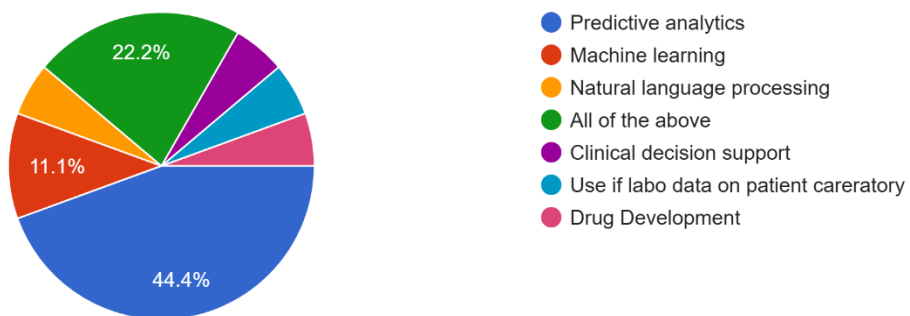 learning and 5.6% of respondents are each making use of natural language processing, clinical decision support, laboratory data on patient careratory and drug development.

4.3.2 Impact of Data Science Techniques on Patient Outcomes, Reduction of Health Costs and Operational Efficiency
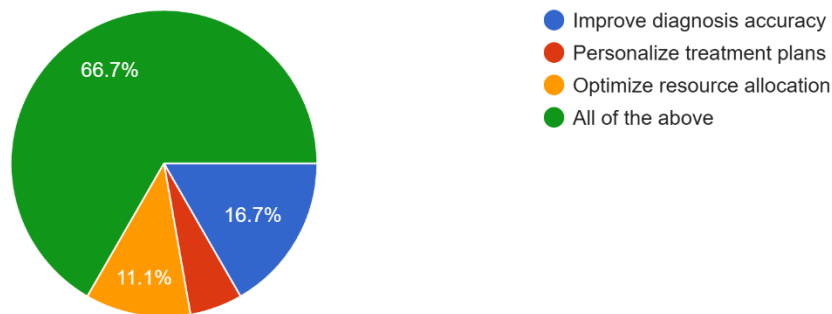
Figure 3. Impact of Data Science Techniques on Patient outcomes and operational efficiency

Figure 3 shows that 66.7% of the participants believe that data science techniques have an impact on patient outcomes, reduce health costs and enhance operational efficiency while 16.7% believe that the techniques only improve diagnosis accuracy. 11.1% believe that the techniques help optimize resource allocation and 5.6% respondents believe that the techniques only help personalize treatment plans.

4.3.3 Technical Barriers that hindered the Effective Adoption and Integration of Data Science Methodologies in Health
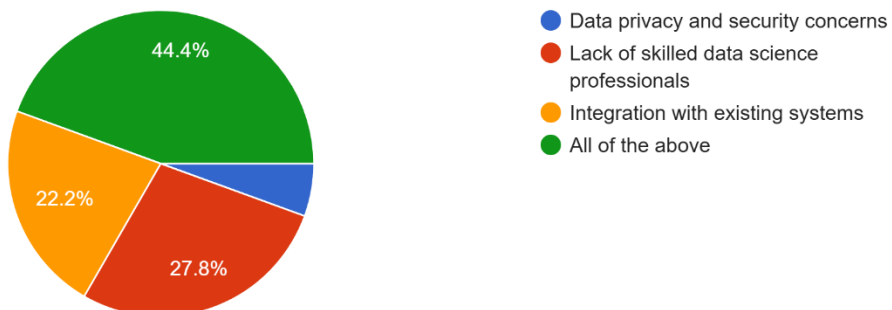
Figure 4. Technical Barriers that hinder Integration of Data Science Methodologies in Health

In figure 4 shown above 44.4% of the respondents agreed that all three points are the barriers that hinder integration of data science and health while 27.8% of respondents have faced lack of skilled data science professionals as a barrier. 22.2% of respondents have encountered a barrier involving the failure to integrate the methodologies with existing systems and 5.6% faced issues related to data privacy and security concerns.

4.3.4 Addressing the Challenges to Facilitate the Successful Implementation of Data Science in Healthcare Organizations
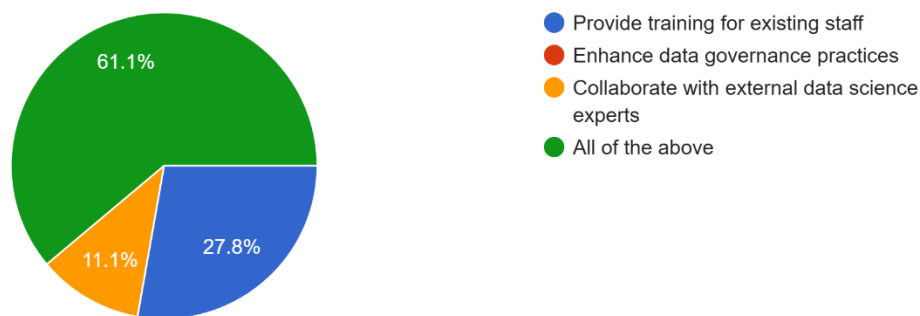
Figure 5. Ways of Addressing Challenges to Facilitate Successful Implementation of Data Science in Health

Figure 5 shows that 61.1% of respondents believe that all the aforementioned points can be used to address challenges faced while 27.8% of respondents believe that providing training for existing staff is enough in their organizations. 11.1% of respondents believe that collaborating with external data science experts is what is needed.

4.3.5 Strategies That have been Implemented or Considered to Build Data Science Capabilities within your organizations
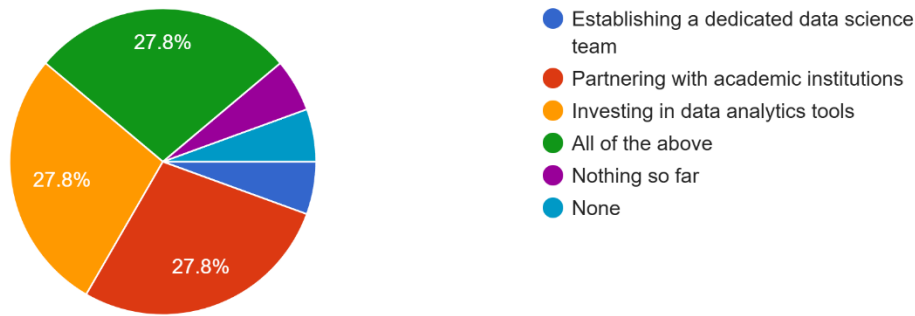
Figure 6. Strategies Implemented or Considered to Build Data Science Capabilities in Organisations

Figure 6 shows that 27.8% of respondents have implemented or considered all mentioned strategies while the other 27.8% have considered partnering with academic institutions and another 27.8% have considered or invested in data analytics tools. The remaining 16.8% of respondents consist of 5.6% respondents who have considered establishing a dedicated data science team and 11.2% who have not considered or implemented any of the strategies.

4.3.6 How to Foster Effective Cross-functional Collaboration between Data Science and Healthcare Professionals to Drive Data-driven Transformation in the Industry
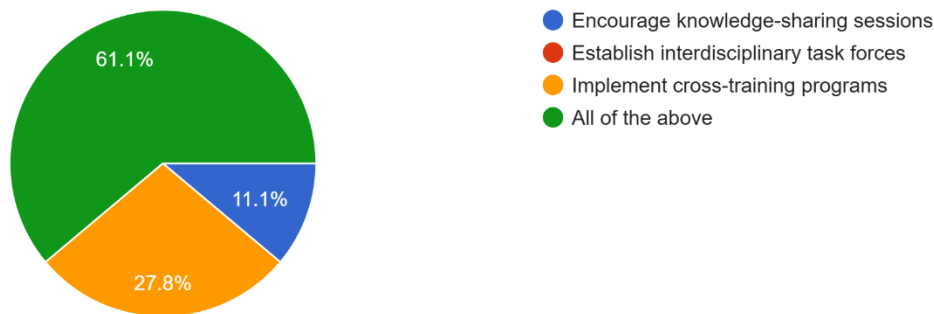
Figure 7. Effective ways to Cross-functional collaboration between data science and healthcare professionals

Figure 7 shows that 61.1% of respondents believe that encouraging knowledge-sharing sessions, establishing interdisciplinary task forces and implementing cross-training programs are all needed to foster effective cross-functional collaboration. 27.8% of respondents believe that implementing cross-training programs is the most effective way while 11.1% believe that encouraging knowledge-sharing sessions is most essential.

4.3.7 Challenges Faced When Integrating Data Science Methodologies into Healthcare Practices
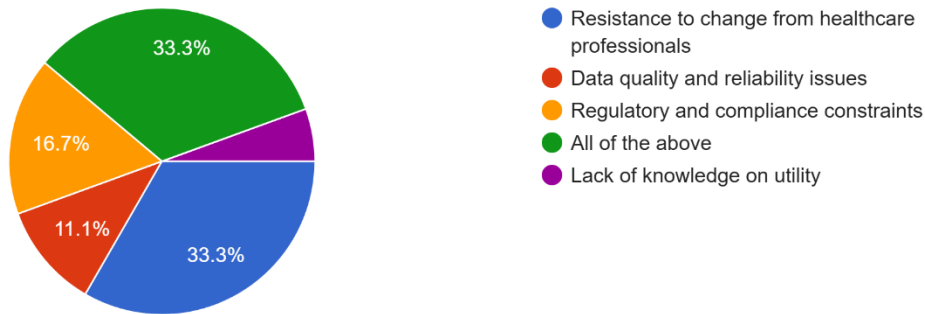
Figure 8. Challenges Faced When Integrating Data Science Methodologies into Healthcare Practices

Figure 8 shows that 33.3% of respondents faced a resistance to change whilst 33.3% of respondents have faced all challenged mentioned. 16.7% of respondents faced regulatory and compliance constraints while 11.1% faced issues with the data quality and reliability. The remaining 5.6% of respondents faced a challenge with the lack of knowledge of utility.

4.3.8 Limitations or Constraints Encountered in Application of Data Science in Healthcare
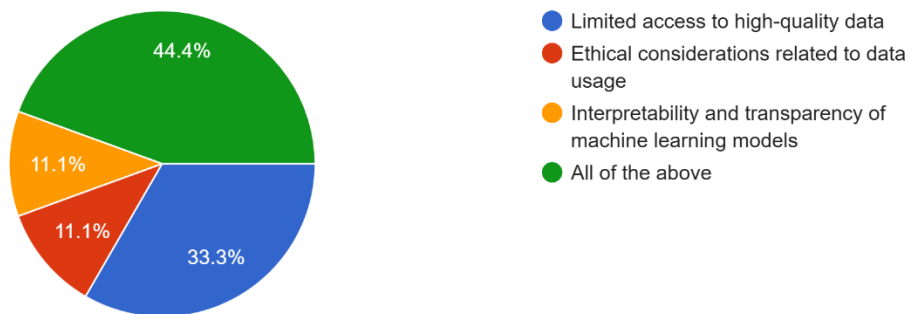
Figure 9. Limitations or Constraints Encountered in Application of Data Science in Healthcare

Figure 9 shows that 44.4% of respondents have faced all the limitations mentioned while 33.3% have faced limited access to high-quality data. 11.1% have limitations associated with ethical considerations related to data usage while 11.1% of respondents have faced interpretability and transparency of machine learning models constraints.

4.3.9 Improvements or Enhancements to Further Advance the Integration of Data Science in Healthcare in Future
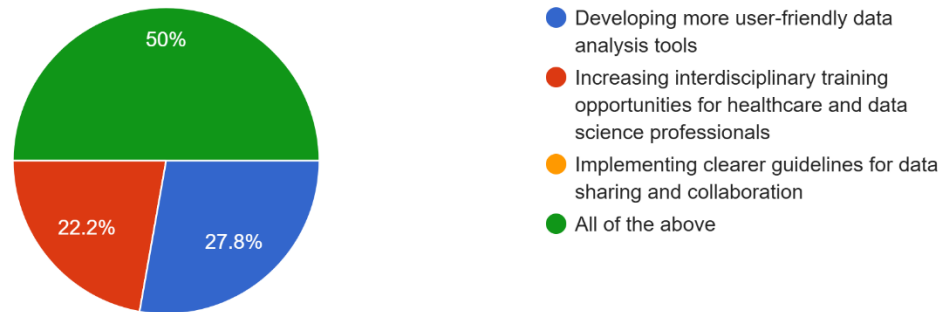
Figure 10. Improvements or Enhancements to Further Advance the Integration of Data Science in Healthcare for the future

Figure 10 shows that 50% of respondents believe that implementing clearer guidelines for data sharing and collaboration and the other two points can be useful in future to advance the integration of data science in healthcare. 27.8% of respondents believe that developing more user-friendly data analysis tools will contribute more while 22.2% of the respondents believe that increasing interdisciplinary training opportunities for healthcare and data science professionals.

4.3.10 Key Insights and Findings Discovered through Study that could contribute to Academic and Industry Knowledge at the Intersection of Data Science and Health
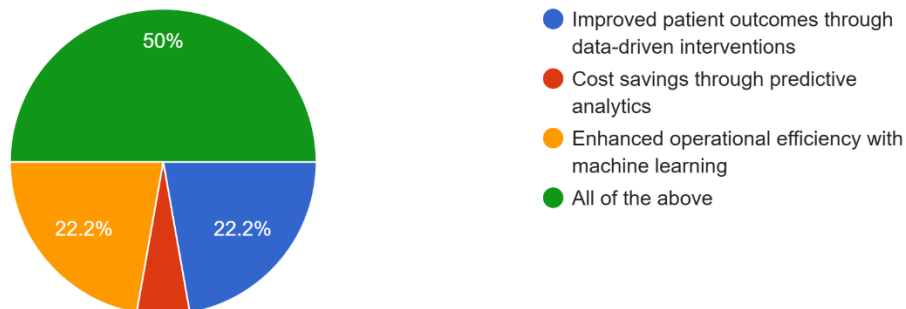
Figure 11. Discoveries in Data Science and Health: Academic and Industry Insights.

Figure 11 shows that 50% of the respondents have discovered the key insights mentioned. 22.2% of the respondents discovered that there are improved patient outcomes through data-driven interventions while the other 22.2% of respondents have discovered an enhanced operational efficiency with machine learning. 5.6% of respondents have discovered cost savings through predictive analytics.

4.4 DISCUSSION AND INTERPRETATION

In Figure 2, the exploration of key applications of data science techniques in healthcare provides a window into the innovative practices adopted by healthcare professionals. Understanding the utilization of predictive analysis, machine learning, and other techniques offers valuable insights into the evolving landscape of patient care and medical advancements.

In Figure 3, the assessment of the impact of data science techniques on patient outcomes, cost reduction, and operational efficiency reveals the aspirations and expectations of healthcare providers. These insights show the potential transformative power of data-driven interventions in shaping the future of healthcare delivery.

Figure 4's discussion on the technical barriers hindering the adoption of data science methodologies in healthcare underscores the challenges faced by organizations striving to embrace technological advancements. These barriers serve as a reminder of the human element behind the integration of complex data science solutions in healthcare settings.

In Figure 5, the strategies proposed to address challenges in implementing data science in healthcare organizations exemplify the proactive and solution-oriented approach of healthcare professionals. These strategies reflect a collaborative effort to navigate obstacles and improve the efficacy of healthcare services through the integration of data-driven solutions.

Lastly, in Figure 6, the strategies implemented to build data science capabilities within organizations demonstrate the forward-thinking mindset of healthcare professionals committed to enhancing patient care. By investing in data analytics tools, forming partnerships with academic institutions, and developing dedicated data science teams, these individuals are actively contributing to the advancement of healthcare practices and the promotion of innovative solutions in the industry.

4.5 CONLCUSION

In conclusion, the data analysis presented in this chapter sheds light on the significant implications of data science in healthcare and its potential to enhance patient outcomes and operational efficiency. Despite the acknowledged technical hurdles, there exists a shared optimism among respondents regarding the transformative effects of data-driven methodologies in healthcare. The proposed strategies to surmount these challenges exemplify the proactive stance taken by healthcare professionals towards embracing innovation and fortifying data science competencies within their respective organizations. By addressing impediments, fostering collaborative initiatives, and investing in advanced data analytics tools, these individuals are actively contributing to the advancement of healthcare practices, with a focus on elevating patient care standards and promoting positive outcomes for both patients and healthcare providers.

CHAPTER FIVE: SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

5.1 INTRODUCTION

This chapter summarizes the key findings of the research on the integration of data science and healthcare. It revisits the study's objectives, evaluates their achievement, and provides actionable recommendations to enhance the application of data science in healthcare. The chapter also highlights the potential of data science to transform healthcare delivery and improve patient outcomes.

5.2 SUMMARY OF KEY FINDINGS

The study explored the role of data science in healthcare, focusing on its potential to enhance patient care, reduce costs, and improve operational efficiency. Key findings include:

1. Applications of Data Science: Predictive analytics, machine learning, and clinical decision support systems are widely used to improve diagnosis accuracy, optimize resource allocation, and personalize treatment plans.

2. Impact on Healthcare: A majority of respondents (66.7%) believe that data science significantly improves patient outcomes, reduces costs, and enhances operational efficiency. These tools are particularly effective in improving diagnostic accuracy and resource management.

3. Challenges to Integration: Key barriers include a lack of skilled professionals, difficulties in integrating data science with existing systems, and concerns about data privacy and security.

4. Strategies for Success: Respondents suggested training staff, collaborating with external experts, and investing in data analytics tools as effective strategies to overcome challenges. Cross-functional collaboration between data scientists and healthcare professionals was also emphasized.

5. Future Enhancements: Recommendations for advancing data science in healthcare include implementing clearer data-sharing guidelines, developing user-friendly tools, and increasing interdisciplinary training opportunities.

5.3 CONCLUSIONS

The research demonstrates that data science has the potential to revolutionize healthcare by improving patient outcomes, reducing costs, and enhancing operational efficiency. However, successful integration requires addressing technical, organizational, and ethical challenges. The study underscores the importance of collaboration, investment in data science capabilities, and ethical data usage.

The findings support the following hypotheses:
- Hypothesis 1: Data science improves patient outcomes and care quality.

- Hypothesis 2: Data science optimizes operational efficiency and cost management.

- Hypothesis 3: Data science fosters innovation and personalized medicine.

- Hypothesis 4: Data science helps address health disparities by analyzing population-level data.

5.4 RECOMMENDATIONS

Based on the findings, the following recommendations are proposed:

1. Invest in Training and Education: Healthcare organizations should prioritize training programs to equip staff with data science skills. Collaborations with academic institutions can help bridge the skills gap and reduce reliance on human decision-making.

2. Enhance Data Integration: Healthcare systems should focus on integrating disparate data sources and ensuring interoperability. This will enable comprehensive analysis and improve the accuracy of predictive models, leading to better patient outcomes.

3. Address Privacy and Security: Robust measures must be implemented to protect patient data. Clear policies and ethical guidelines should be established to ensure secure and responsible data usage.

4. Foster Collaboration: Encouraging collaboration between data scientists and healthcare professionals is essential. Interdisciplinary task forces and cross-training programs can drive innovation and improve healthcare delivery.

5. Develop User-Friendly Tools: To increase adoption, healthcare organizations should invest in user-friendly data analysis tools. These tools should cater to professionals with varying levels of technical expertise.

6. Promote Interdisciplinary Research: Future research should explore the intersection of data science and healthcare to identify new applications and address emerging challenges, such as improving safety, efficiency, and equity.

7. Implement Data-Sharing Guidelines: Policymakers and healthcare organizations should work together to create clear guidelines for data sharing. This will facilitate secure and ethical data exchange, improving accountability and reducing costs.

8. Leverage AI for Non-Critical Cases: AI and machine learning can be used for non-critical cases, allowing patients to receive diagnoses and care at home. This will improve efficiency and reduce the burden on healthcare facilities.

9. Focus on Equity: Data science should be used to address health disparities and improve access to care for underserved populations. Analyzing social determinants of health can help develop targeted interventions.

10. Encourage International Collaboration: Connecting with global experts in data science and healthcare can provide valuable insights and best practices, ensuring that healthcare systems remain innovative and forward-thinking.

11. Promote Public Awareness: Educating the public about the benefits of data science in healthcare can build trust and acceptance of data-driven interventions, ensuring broader community benefits.

12. Adopt AI as a Core Strategy: AI-driven solutions should be prioritized to improve patient outcomes, enhance efficiency, and reduce costs. This will prepare healthcare systems for future challenges.

5.5 CONCLUSION

The integration of data science into healthcare offers significant potential to improve patient care, optimize operations, and drive innovation. However, realizing this potential requires addressing technical, organizational, and ethical challenges. By fostering collaboration, investing in data science capabilities, and ensuring ethical data usage, healthcare organizations can harness the power of data science to transform the industry. The recommendations provided in this study aim to guide organizations toward a more data-driven, efficient, and equitable healthcare system.

LIST OF REFERENCES

- (PDF) the effects of data science and analytics on Healthcare. (n.d.). https://www.researchgate.net/publication/374535216_THE_EFFECTS_OF_DATA_SCIENCE_AND_ANALYTICS_ON_HEALTHCARE
- *Data Science in healthcare: Applications, roles and Benefits*. KnowledgeHut. (2024, June 30). https://www.knowledgehut.com/blog/data-science/data-science-in-healthcare
- Subrahmanya, S. V. G., Shetty, D. K., Patil, V., Hameed, B. M. Z., Paul, R., Smriti, K., Naik, N., & Somani, B. K. (2022, August). *The role of Data Science in healthcare advancements: Applications, benefits, and future prospects*. Irish journal of medical science. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9308575/
- *Data Science in healthcare: Applications, roles and Benefits*. KnowledgeHut. (2024, June 30). https://www.knowledgehut.com/blog/data-science/data-science-in-healthcare
- Mandhir, Mandhir, Richard Russell Marie-Christine Boucher, Richard Russell, Marie-Christine Boucher, Marc-Andre Giguere, Marc-Andre Giguere, Shawn Goheen, Shawn Goheen, Insurance Thought Leadership ITL Partner: FRISS, Insurance Thought Leadership, ITL Partner: FRISS, Insurance Thought Leadership, Mike Seiwert, Mike Seiwert, Stephen Applebaum Alan Demers, Stephen Applebaum, Alan Demers, Marcin Nowak, … Matt Clarenson. (2024, March 26). *Data Science is Transforming Public Health*. Data Science Is Transforming Public Health | Insurance Thought Leadership. https://www.insurancethoughtleadership.com/life-health/data-science-transforming-public-health
- Subrahmanya, S. V. G., Shetty, D. K., Patil, V., Hameed, B. M. Z., Paul, R., Smriti, K., Naik, N., & Somani, B. K. (2022, August). *The role of Data Science in healthcare advancements: Applications, benefits, and future prospects*. Irish journal of medical science. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9308575/
- *Data Science in healthcare: Applications, roles and Benefits*. KnowledgeHut. (2024, June 30). https://www.knowledgehut.com/blog/data-science/data-science-in-healthcare
- Subrahmanya, S. V. G., Shetty, D. K., Patil, V., Hameed, B. M. Z., Paul, R., Smriti, K., Naik, N., & Somani, B. K. (2022, August). *The role of Data Science in healthcare advancements: Applications, benefits, and future prospects*. Irish journal of medical science. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9308575/
- *Data Science in healthcare: Applications, roles and Benefits*. KnowledgeHut. (2024, June 30). https://www.knowledgehut.com/blog/data-science/data-science-in-healthcare
- Team, P. (2023, August 25). *The role of Data Science in Healthcare Advancement: Applications and benefits*. eHealth4everyone. https://ehealth4everyone.com/the-role-of-data-science-in-healthcare-advancement-applications-and-benefits/
- Vivek, J. (n.d.). *Data Science in healthcare industry: Benefits, strategies, applications, tools, and future trends*. Zuci Systems. https://www.zucisystems.com/blog/data-science-in-healthcare/

- Bates, D. W., Saria, S., Ohno-Machado, L., Shah, A., & Escobar, G. (2014). Big data in health care: Using analytics to identify and manage high-risk and high-cost patients. *Health Affairs, 33*(7), 1123–1131. https://doi.org/10.1377/hlthaff.2014.0041
- Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019)**.** Dissecting racial bias in an algorithm used to manage the health of populations. *Science, 366*(6464), 447–453. https://doi.org/10.1126/science.aax2342
- Price, W. N., & Cohen, I. G. (2019)**.** Privacy in the age of medical big data. *Nature Medicine, 25*(1), 37–43. https://doi.org/10.1038/s41591-018-0272-7
- Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. *New England Journal of Medicine, 380*(14), 1347–1358. https://www.nejm.org/doi/10.1056/NEJMra1814259
- Wiens, J., et al. (2019)**.** Do no harm: A roadmap for responsible machine learning for health care. *Nature Medicine, 25*(9), 1337–1340. https://www.nature.com/articles/s41591-019-0548-6
- Davenport, T. H., & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. Future Healthcare Journal, 6(2), 94-98. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6616181/
- Shilo, S., Rossman, H., & Segal, E. (2020). Axes of a revolution: Challenges and promises of big data in healthcare. Nature Medicine, 26(1), 29-38. https://www.nature.com/articles/s41591-019-0727-5
- Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. Nature Medicine, 25(1), 44-56. https://www.nature.com/articles/s41591-018-0300-7
- Price, W. N., & Cohen, I. G. (2019). Privacy in the age of medical big data. Nature Medicine, 25(1), 37-43. https://www.nature.com/articles/s41591-018-0272-7
- Mehta, N., Pandit, A., & Shukla, S. (2022). Transforming healthcare with big data analytics and artificial intelligence: A systematic review. Journal of Biomedical Informatics, 134, 104199. https://www.sciencedirect.com/science/article/pii/S1532046422001759
- Jiang, F., Jiang, Y., Zhi, H., et al. (2021). Artificial intelligence in healthcare: Past, present and future. Stroke and Vascular Neurology, 6(1), 26-38. https://svn.bmj.com/content/6/1/26
- Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2020). From what to how: An initial review of publicly available AI ethics tools. Science and Engineering Ethics, 26(4), 2141-2168. https://link.springer.com/article/10.1007/s11948-019-00165-5
- Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447-453. https://www.science.org/doi/10.1126/science.aax2342
- Matheny, M., Israni, S. T., Ahmed, M., & Whicher, D. (Eds.). (2020). Artificial intelligence in health care: The hope, the hype, the promise, the peril. National Academy of Medicine. https://nam.edu/artificial-intelligence-special-publication/
- Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. New England Journal of Medicine, 380(14), 1347-1358. https://www.nejm.org/doi/full/10.1056/NEJMra1814259

APPENDIX 1: AUREC APPROVAL LETTER

AFRICA UNIVERSITY
A United Methodist Related Institution
"Investing in Africa's future"

AFRICA UNIVERSITY RESEARCH ETHICS COMMITTEE (AUREC)

P.O. Box 1320 Mutare, Zimbabwe, Off Nyanga Road, Old Mutare-Tel (+263-20) 60075/60026/61611 Fax: (+263 20) 61785 Website: www.africau.edu

Ref: AU 3413/24   21 August, 2024

CHIDO KAMBA
C/O Africa University
Box 1320
MUTARE

RE: The Fusion and Integration of Data Science and Health

Thank you for the above-titled proposal that you submitted to the Africa University Research Ethics Committee for review. Please be advised that AUREC has reviewed and approved your application to conduct the above research.

The approval is based on the following.
a) Research proposal
*   **APPROVAL NUMBER**   AUREC 3413/24
    This number should be used on all correspondences, consent forms, and appropriate documents.

*   **AUREC MEETING DATE**   NA
*   **APPROVAL DATE**   August 21, 2024
*   **EXPIRATION DATE**   August 21, 2025
*   **TYPE OF MEETING**: Expedited
    After the expiration date, this research may only continue upon renewal. A progress report on a standard AUREC form should be submitted a month before the expiration date for renewal purposes.
*   **SERIOUS ADVERSE EVENTS** All serious problems concerning subject safety must be reported to AUREC within 3 working days on the standard AUREC form.
*   **MODIFICATIONS** Prior AUREC approval is required before implementing any changes in the proposal (including changes in the consent documents)
*   **TERMINATION OF STUDY** Upon termination of the study a report has to be submitted to AUREC.

> AFRICA UNIVERSITY
> RESEARCH ETHICS COMMITTEE (AUREC)
> **APPROVED**
> P.O. BOX 1320, MUTARE, ZIMBABWE

Yours Faithfully

Chinzou

MARY CHINZOU
ASSISTANT RESEARCH OFFICER: FOR CHAIRPERSON
AFRICA UNIVERSITY RESEARCH ETHICS COMMITTEE

APPENDIX 2: APPROVAL OF SUBMISSION

COLLEGE OF BUSINESS, PEACE, LEADERSHIP AND GOVERNANCE
_____

…29…/..07…/2024

Africa University Research Ethics Committee

Ref: Approval for AUREC Proposal Submission

………Chido Kamba……… has worked on the proposal with the assistance of the supervisor and I confirm that it is ready for reviewed by your esteemed committee.

Respectfully submitted,

Mr T Makambwa   TMakambwa
Supervisor's Name   Supervisor's Signature

APPENDIX 3: QUESTIONNAIRE

The Integration of Data Science and Health Questionnaire

This questionnaire is a part of the study of the impact that data science has in the healthcare field in Harare. It is an exploration of the challenges, opportunities and strategies that can influence the integration of data science and health in order to improve patient care. All responses shared in this questionnaire will help provide good insights and all information will be kept confidential.

* Indicates required question

Please specify your profession in the healthcare industry *

Your answer

1. What are the key applications of data science techniques that you are currently using or exploring in the healthcare industry?
*

- o   Predictive analytics
- o   Machine learning
- o   Natural language processing
- o   All of the above
- o   Other:

2. How do you believe these data science techniques could potentially impact patient outcomes, reduce health costs, and enhance operational efficiency in healthcare settings?*

- o   Improve diagnosis accuracy
- o   Personalize treatment plans
- o   Optimize resource allocation
- o   All of the above
- o   Other:

3. What technical barriers have you encountered that hinder the effective adoption and integration of data science methodologies in healthcare?*

- o   Data privacy and security concerns
- o   Lack of skilled data science professionals
- o   Integration with existing systems
- o   All of the above
- o   Other:

4. In your opinion, how can these challenges be addressed to facilitate the successful implementation of data science in healthcare organizations?*

- o   Provide training for existing staff
- o   Enhance data governance practices
- o   Collaborate with external data science experts
- o   All of the above
- o   Other:

5. What strategies have you implemented or considered to build data science capabilities within your organization?*

- o   Establishing a dedicated data science team
- o   Partnering with academic institutions
- o   Investing in data analytics tools
- o   All of the above
- o   Other:

6. How would you foster effective cross-functional collaboration between data science and healthcare professionals to drive data-driven transformation in the industry?*

- o   Encourage knowledge-sharing sessions
- o   Establish interdisciplinary task forces
- o   Implement cross-training programs
- o   All of the above
- o   Other:

7. What challenges have you faced in integrating data science methodologies into healthcare practices?*

- o   Resistance to change from healthcare professionals
- o   Data quality and reliability issues
- o   Regulatory and compliance constraints
- o   All of the above
- o   Other:

8. How have you overcome or mitigated these challenges to ensure the successful implementation of data science in healthcare?

Your answer

9. What limitations or constraints have you encountered in the application of data science in healthcare?*

- o Limited access to high-quality data
- o Ethical considerations related to data usage
- o Interpretability and transparency of machine learning models
- o All of the above
- o Other:

10. How do you address these limitations to maximize the benefits of data science in healthcare practices?

Your answer

11. What improvements or enhancements do you think could be made in the future to further advance the integration of data science in healthcare?*

- o Developing more user-friendly data analysis tools
- o Increasing interdisciplinary training opportunities for healthcare and data science professionals
- o Implementing clearer guidelines for data sharing and collaboration
- o All of the above
- o Other:

12. What key insights and findings have you discovered through your study that could contribute to academic and industry knowledge at the intersection of data science and healthcare?*

- o Improved patient outcomes through data-driven interventions
- o Cost savings through predictive analytics
- o Enhanced operational efficiency with machine learning
- o All of the above
- o Other:

13. How do you envision these insights shaping the future of data-driven healthcare practices?

Your answer

Submit   Clear form

APPENDIX 4: APPROVAL LETTER 1

Parktown Hospital :: Health Care Within Reach

1074 Sixth Avenue, Waterfalls,
Harare, Zimbabwe
+263 24 2661582
https://parktownh

23 July 2024

Chido Kamba
C/O Africa University
MUTARE

Dear Chido,

RE: REQUEST TO CONDUCT RESEARCH WITHIN PARKTOWN HOSPITAL

Thank you for your interest in conducting research within Parktown Hospital. Your request to conduct research on the topic, "**The Integration and Fusion of Data Science and Healthcare**", has been approved.

You are required to provide the final copy of the research document, which is duly stamped by your university with a clear executive summary, comprehensive research findings and recommendations to me.

Kindly get in touch with me on 0772 214 239.

For and on behalf of Parktown Hospital,

Dr J Mhamba,
CEO and Founder of Parktown Hospital

APPENDIX 5: APPROVAL LETTER 2

Citimed Chitungwiza Hospital t/a
Southmed Hospital Chitungwiza

Zengeza 4, 14656
Hadzinanhanga Road,
Chitungwiza, Zimbabwe

26 July 2024

Chido Kamba
C/O Africa University
MUTARE

Dear Chido,

RE: REQUEST TO CONDUCT RESEARCH WITHIN CITIMED CHITUNGWIZA HOSPITAL

Thank you for your interest in conducting research within Citimed Chitungwiza Hospital. Your request to conduct research on the topic, "The Integration and Fusion of Data Science and Healthcare", has been approved.

You are required to provide the final copy of the research document, which is duly stamped by your university with a clear executive summary, comprehensive research findings and recommendations to me.

Kindly get in touch with me on +263772261519.

For and on behalf of Citimed Chitungwiza Hospital,

Mrs Munaki

Manager of Radiology Department

APPENDIX 6: APPLICATION FOR AUREC INITIAL REVIEW

For office use only	
Protocol no. …………………. ☐	Office stamp
Type of review: Full Committee ☐	
Expedited ☐	

AFRICA UNIVERSITYRESEARCH ETHICS COMMITTEE (AUREC)

APPLICATION FOR INITIAL REVIEW

NB: This form must be completed by all persons/teams applying for ethical review by AUREC. Upon completion by the investigator(s) /researcher(s) it should be submitted electronically to AUREC, aurec@africau.edu. . Application fees (to cover the costs of reviewing proposal) should be paid to the Africa University Business Office, and proof of payment should accompany each application. Please complete all sections of this application form. If there is insufficient space on the form you may use additional pages.

Check list

This checklist is meant to aid researchers in preparing a complete application package and to help expedite review by the AUREC. Please tick all boxes as appropriate (Indicate **N/A** where inapplicable).

CONTACT PERSON'S NAME :   **_____CHIDO KAMBA_____**
CONTACT ADDRESS: **319 FURZE ROAD, WATERFALLS, HARARE**

EMAIL ADDRESS :   **ckamba@africau.edu**

CONTACT NO: **0784900306**

UNDERGRADUATES

		Applicant	AUREC
1	Application form duly completed	✓	
2	Electronic version of research proposal to aurec@africau.edu	✓	
3	Consent forms in English and local language of study population	✓	
4	Advertisement or letter or card used for recruiting participants and any supplementary information *(if applicable).*	✓	
5	Data collection tools being administered during the study in English and local language of study population *(if applicable)* included in the proposal	✓	

		Applicant	AUREC
6	Budget and timeframe included in the proposal.	✓	
7	Approval letter from your academic supervisor/college or institution	✓	
8	Approval letter from authorities where study will be conducted	✓	
9	Application fee paid at AU Business Office and receipt (or copy) attached to application form.	✓	

POST GRADUATES AND OTHER RESEARCHERS

		Applicant	AUREC
1	Application form duly completed		
2	Electronic version of full research proposal (chapter 1 – 3 completed) to aurec@africau.edu		
3	Proposal summary (see guidelines below)		
4	Consent form in English and local language of study population		
5	Advertisement or letter or card used for recruiting participants and any supplementary information (if applicable).		
6	Data collection tools being administered during the study in English and local language of study population (if applicable)		
7	Budget and timeframe		
8	Approval letter from academic supervisor/college or institution (if you are a student)		
9	Approval letter from authorities where study will be conducted		
10	Application fee paid at AU Business Office and receipt attached to application form.		
12	CV's for D Phil and Phd candidates.		

CKamba   CHIDO KAMBA   29/07/2024

-------------------------------------   --------------------------   ---------------------

Signature: Investigator/Researcher   **Name**   **Date**

1. **General information**
   1.1. Study title: THE FUSION AND INTEGRATION OF DATA SCIENCE AND HEALTHCARE

   1.2. Name of Principal Investigator(PI)/ Researcher: _____CHIDO KAMBA_____

   1.3. Nationality of Investigator/Researcher: _____ZIMBABWEAN_____

1.4. Proposed date of start of study: _(dd/mm/yyyy) _____10/08/2024_____

1.5. Expected duration of study: _____5 MONTHS_____

1.6. Study site(s) in Zimbabwe: _____Harare_____

1.7. Sites outside Zimbabwe: _____

1.8. Study budget: ____$100_____   Source of Funding: _____SELF_____

1.9. Is the researcher a student?   Yes

1.10.   If Yes, indicate the following:

    1.10.1. Name and address of institution: _____AFRICA UNIVERSITY_____

    1.10.2. College: _____COLLEGE OF ENGINERRING AND APPLIED SCIENCES_____

    1.10.3. Level of study   Undergraduate/Master's/PhD _____UNDERGRADUATE_____

    1.10.4. Name of Supervisor: _____MR. T. MAKAMBWA_____

1.11.   If No to question 1.10, then indicate the following:
    1.11.1. Name and address of institution: _____
    1.11.2. Academic Title of PI: _____
    1.11.3. Existing Qualifications: _____
    1.11.4. Co Investigators:

	Names:	Qualifications	Institution

40

2. Statement by the investigator

I _____CHIDO KAMBA_____certify that the information in this application document and the accompanying documents is true and complete in all respects. I confirm that the application has NOT been rejected by any other ethics review committee.

Signature _____CKamba_____ Date: _____29/07/2024_____

3. Guidelines for the proposal summary: (Times New Roman, double line spacing, font size 12)
    3.1. Introduction
    3.2. Background ,purpose, statement of the problem, justification, significance of the study
    3.3. Aim(s) and objectives: Outline the main aim(s) and objectives of the study and research questions.
    3.4. Literature review
    4.0 Methodology

    4.1 Research Design *(describe how the research will be carried out including plans for data analysis and dissemination)*

    4.2 Study population and sampling procedure*(give details of the study population and how you will carry out the sampling procedure and NOT general meanings of population and sampling methods)*
    4.3 Inclusion/exclusion criteria*(state who qualifies for selection and who does not)*
    4.4 Devices, Tests, Questionnaires, and Interview Guides:
    4.5 Research participants/subjects
        4.5.1   State the total number of human participants to be enrolled
        4.5.2   State the source(s) of recruitment *(e.g. hospitals, schools, etc.)*
        4.5.3   Age range and sex of participants to be recruited.
        4.5.4   Special or vulnerable populations *(state if vulnerable populations e.g. pregnant women, adolescents, children, prisoners, refugees etc are involved)*
        4.5.5   Payment *(if any)* to be paid to each participant
        4.5.6   Informed Consent Procedure*(describe how this will be carried out)*
    4.6 Potential Benefits of the research *(Describe the benefits of the study both to the participants and to the community)*
    4.7 Potential Risks
        4.7.1   Describe any potential risks, discomforts or harms that may be experienced by the participants. These may be physical, psychological, social, legal, economic or other and state procedures to minimise these.
        4.7.2   Management of Risks*(describe how these risks will be managed/mitigated)*
    4.8 Confidentiality/privacy *(give details of how these will be maintained)*

4.9 Investigator Experience/qualifications *(describe any experience or training/courses that the investigator has/has taken that put him/her in good stead to carry out the study)*

4.10 Explain how research results are going to be disseminated to participants

5. Reference List

6. Attachments

6.1 Approval letter from College Supervisor (if you are a student)

6.2 Data collection instruments *(Include anything you will be using to gather data from human subjects e.g. Tests/Questionnaires/Observation Checklists/interview guides/ FGDs guides etc.)*

6.3 Informed Consent Forms or assent (<u>*informed consent form guide is available from AUREC)*</u>

6.4 Budget and timeframe

6.5 Proof of payment of the review fees.

APPENDIX 7: PROOF OF PAYMENT

STEWARD|BANK

Transaction Complete

Scan QR code to validate

Transaction ID
41c9810a

Transaction Reference
266230018580

Destination Phone
08688002151

Destination Account
01322704290011

Amount
218.0