A visual analytics approach to characterising disease progression among adults with chronic diseases in rural Agincourt northeast South Africa

Mapule Dorcus Nhlapho

Student Number: 1385721

A research report submitted to the Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, in partial fulfilment of the requirements for the degree of Master of Science in Epidemiology (Public Health Informatics)

May 2024

DECLARATION

I hereby declare that this research report is project work carried out by me under the guidance of Dr Chodziwadziwa Kabudula. I have taken care in all respects to honour the intellectual property rights of others and have acknowledged the contribution of others. I further declare that the work reported in this project has not been submitted and will not be submitted, either in part or in full, for the award of any other degree or diploma in this institute or any other institute or university.

23 May 2024

Mapule Dorcus Nhlapho

DEDICATION

I would like to dedicate my research report to my parents, Mrs Selinah, and Mr Paul Nhlapho. Thank you for everything.

ABSTRACT

Chronic diseases pose a significant challenge to the healthcare systems in South Africa, calling for innovative approaches for comprehensive understanding and management. This research study utilizes the Agincourt HDSS-Clinic dataset to design and implement a visual analytics system using the R Shiny web application framework. Focused on adults with chronic diseases, the tool employs dynamic visualizations to show patterns of healthcare utilization and disease progression. Through the R Shiny platform, the system provides a user-friendly interface for exploring and interpreting complex data, offering valuable insights into patient healthcare behaviours and the dynamics of chronic illnesses. The study used data from a total of 26 426 patients consisting of 19 265 (73%) females and 7 161 (27%) males. The study revealed previously unrecognized associations between specific chronic conditions including the existence of a substantial intersection between HIV, Hypertension, and Diabetes with 101 patients experiencing the coexistence of all the three conditions. Notably, the visual analytics system facilitated the identification of distinct healthcare utilization patterns across different demographic groups highlighting the most frequently visited health facility accounted for 5 912 patient visits overall while the least visited health facility accounted for 1 447 patient visits. The findings underscore the effectiveness of visual analytics in uncovering trends within complex datasets. The implications of these findings extend beyond the immediate research scope, influencing healthcare strategies and contributing to the ongoing discussions on innovative solutions for chronic disease management. This study contributes to the evolving field of visual analytics in healthcare, demonstrating the potential for such tools to inform decision-making and enhance patient outcomes.

ACKNOWLEDGEMENT

I wish to acknowledge the following people and organisations:

- I would like to express my sincere gratitude to Dr Chodziwadziwa Kabudula for his guidance, supervision, and support throughout this research journey.
- I would like to express my appreciation to the MRC/Wits-Agincourt Unit for allowing me to use their data.
- I would like to thank the Wits MADIVA team for giving me the opportunity to attend the stakeholders’ meetings that were pivotal in the development of the system.
- I would like to express my deepest thanks to Mrs Irma Maré for being a great mentor and motivating me to pursue this degree. Her continued support is highly appreciated.
- I would like to express my appreciation to Wits Health Consortium – Wits Biomedical Informatics and Translational Science for their financial support and allowing me to take time off during this master’s program.

Table of Contents

List of Figures

List of Tables

List of abbreviations and acronyms

AIDS	Acquired Immunodeficiency Syndrome
CD4	Cluster of differentiation 4
CDC	Centers of Disease Control and Prevention
CDM	Chronic disease management
CRAN	Comprehensive R Archive Network
CSS	Cascading Style Sheets
HDSS	Health and Demographic Surveillance System
EHR	Electronic Health Record
HIT	Health Information Technology
HIV	Human Immunodeficiency Virus
HREC	Human Research Ethics Committee
HRIO	Health Records Information Officer
HTML	HyperText Markup Language
LMIC	Low-and-middle income countries
mL	Milliliter
mm^3	Cubic millilitre
mmHg	Millimetre of mercury
mmol/L	Millimoles per litre
NA	Not Available
NCD	Non-communicable diseases
SQL	Structured Query Language
UI	User Interface
VA	Visual Analytics
WHO	World Health Organization

CHAPTER 1: INTRODUCTION

The introductory chapter begins with a brief background describing chronic diseases in the global, Sub-Saharan Africa and South African contexts. It is followed by a literature review consisting of the following sub-headings: overview of chronic diseases, chronic diseases management, health information technology, visualization, and visual analytics application. The chapter ends with the aims and objectives of the master’s research report.

1.1 Background

Chronic diseases are defined by the Centers of Disease Control and Prevention (CDC) (1) as conditions that require ongoing medical attention and last over one year. These conditions, which comprise of non-communicable diseases such as hypertension, diabetes, heart disease, stroke, and cancers and certain long-term infectious diseases such as Human Immunodeficiency Virus (HIV)/Acquired Immunodeficiency Syndrome (AIDS), are leading causes of poor health worldwide. Each year chronic diseases account for more than 15 million deaths (2). In Sub-Saharan Africa, the burden of death and disability caused by chronic diseases is steadily growing (3). It is projected that chronic diseases will overtake infectious diseases as major causes of morbidity and mortality by 2030 (4).

The rising burden of chronic diseases is one of the major challenges facing the healthcare systems in South Africa. The country is experiencing a double burden of high prevalence of both communicable (HIV/AIDS) and non-communicable chronic diseases as a result of a protracted epidemiological transition (5). The country has the highest prevalence of diabetes and as with HIV, it also has the highest burden of hypertension in Sub-Saharan Africa (6,7).

In order for healthcare systems to effectively respond and lessen the burden of chronic diseases in South Africa and elsewhere, they require accurate information regarding the burden of chronic diseases tailored to local contexts. Visual Analytic techniques provide the means to address this requirement.

1.2 Literature Review

The section presents the review of relevant literature on key concepts of chronic diseases, visualizations, and visual analytics. The section also reviews related works and techniques on the development of visualization applications in public health research.

1.2.1 Overview of Chronic Diseases

“Chronic diseases”, which refer to a group of slowly progressive medical conditions of long duration (8), represent a substantial challenge to global health. In South Africa, chronic diseases have increased over the past 15 years accounting for an estimated 37% of mortality and 16% of disability (9). According to the World Health Organization (WHO) (10), the African continent has been projected to experience the largest increase in death rates from chronic diseases. Moreover, the number of people living with more than one chronic condition known as multimorbidity (11) is also steadily increasing.

Multimorbidity appears to be increasingly prevalent in low-and-middle income countries (LMIC) (12). According to Nunes and colleagues (13), patients experiencing multiple chronic conditions often have poorer health outcomes such as declined physical and mental health functioning and increased mortality rates (14). Prior studies have found that individuals with multimorbidity compared to those with only one condition have their lives impacted in many ways including reduced quality of life (15,16) and increased frequency of healthcare utilisation (17).

In South Africa lifestyle-associated chronic diseases significantly burden healthcare facilities, with hypertension and type 2 diabetes mellitus holding the top and fourth positions, respectively, on the list of the top 25 diagnoses made at primary health care facilities (18). Urbanisation is among the factors attributing to the increase in the burden of chronic diseases in Africa (10). According to Lloyd-Sherlock (7) urbanisation has been thought to be a major determinant of the increasing prevalence of obesity and physical inactivity as people transition from traditional foods to a ‘western’ diet (processed foods, more added sugar, and salt) (19). However, the prevalence of hypertension in rural areas is also high (20). South Africa has a unique burden of disease with the largest number of people living with HIV in the world (21). As stated by Chang and colleagues (22), people with HIV are living longer due to the widespread use of antiretrovirals and are developing age-related non-communicable diseases (NCD), giving rise to disease patterns of coexisting communicable and non-communicable diseases (23). Furthermore, it is not uncommon that people living with HIV also develop other chronic conditions (24).

This research focuses on HIV, Hypertension and Diabetes as they are the most prevalent chronic diseases in South Africa.

1.2.1.1 Diabetes mellitus

Diabetes is a chronic disease with severe late complications. CDC (25) defines diabetes as “a long-lasting health condition that affects how your body turns food into energy”. It occurs either when your body cannot produce enough insulin or cannot effectively use the insulin it produces. This disease is known to impact the quality of life and cause disability, which in some cases affect an individual’s capacity to manage and maintain longer-term health and well-being (26). According to key facts published by WHO (27), diabetes is a major cause of blindness, heart attacks, stroke and lower limp amputation and was the ninth leading cause of death in 2019. In South Africa, diabetes was among the top ten leading causes of deaths in 2018 (5).

Blood glucose level, measured in millimoles per litre (mmol/L), is essential for good diabetes management. According to Mathew and colleagues (28) a random blood glucose of or above 11.1 mmol/L or a fasting blood glucose of or above 7 mmol/L on two or more separate tests indicate that the patient is likely to have diabetes mellitus. There are three main types of diabetes; Type 1: is caused by insulin deficiency and is usually diagnosed in children and young adults; Type 2: occurs when your body is unable to use insulin and can't keep blood sugar levels normal; and gestational diabetes: is a form of diabetes affecting pregnant women.

1.2.1.2 Human Immunodeficiency Virus (HIV)

HIV has become a manageable chronic health condition that enables people living with the condition to lead a long and healthy life. HIV targets the immune system and destroys white blood cells; cluster of differentiation 4 (CD4) cells which is the body's defence against illnesses (29). The function of the immune system is measured by CD4 cell count. Before the recently implemented universal test and treat, the recommended guidelines for Antiretroviral therapy (ART) was <500 cells per cubic millilitre (mm3) (30)

According to WHO (31), at the end of 2020, there were an estimated 37.7 million people living with HIV and over two thirds of whom are in the African region**.** Akullian and colleagues (32) further state that South Africa has the largest HIV epidemic in the world, with 7.7 million people currently living with HIV. Moreover, South Africa is identified as home to the largest concentration of people living with HIV than anywhere in the world (33).

1.2.1.3 Hypertension

Hypertension can be defined using specific systolic and diastolic blood pressure levels (34). Hypertension is also referred to as elevated blood pressure and is the force of blood pushing against the walls of the body's arteries (35). Blood pressure is written as two numbers. The first number (systolic) represents the pressure in blood vessels when the heart beats. The second number (diastolic) represents the pressure in the vessels when the heart rests between beats. According to WHO (35), hypertension is diagnosed when the systolic blood pressure reading is ≥140 millimetre of mercury (mmHg) and/or the diastolic blood pressure reading is ≥90 mmHg, measured on two different days.

Hypertension continues to be a rising global health problem, with over 1 billion people living with hypertension, and accounting for approximately 10 million related deaths globally (36,37). In Sub-Saharan Africa, hypertension is a widespread problem with some communities having prevalence of as high as 38% (38). A study by Gómez-Olivé and colleagues (39) reported that the prevalence of hypertension in South Africa was among the highest at over 77% compared to compared China, Russia, India, Ghana, and Mexico in populations over 50 years of age.

Assessing and understanding the health needs of different communities can contribute to targeted primary healthcare prevention, intervention, and treatment programs for chronic diseases in resource-constrained settings to reduce the prevalence of chronic diseases in South Africa (40). An understanding of the patterns of chronic diseases is required to better manage the chronic conditions. Therefore, the implementation and use of effective chronic disease management is urgently required in different settings in South Africa to tackle the burden of chronic diseases.

1.2.2 Chronic Disease Management

Chronic disease is a primary health problem leading to physical disability especially in the elderly population (41). There are programs used for chronic disease management (CDM). However, they vary in design and effectiveness (42). These programs according to Nolte and Osborne (43), Goodridge and colleagues (44), are designed to monitor symptoms and adherence to treatment for preventing disease progression. It is worth mentioning that although these programs exist, not everyone is eligible to participate in them. Chronic conditions that are poorly managed greatly increase the risk of complications and thus it is important to identify interventions to improve the management of chronic diseases that benefit everyone that utilises healthcare services (45).

As described by Maimela and colleagues (46), CDM is a public health approach that helps people to keep healthy through early detection of complications and management of chronic diseases. Nuño (47) expands the definition by describing CDM as an integrated approach that involves the patient, family, and the community as active participants over a long life of care. Efforts to manage chronic diseases in primary healthcare facilities will require long period of observation and care. According to Siantz and Aranda (48), the successful management of chronic conditions requires behaviour and lifestyle adjustment to minimize disability and functional limitations. Moreover, chronic disease management in primary care helps understand which interventions are effective, for whom and in what context (49).

1.2.3 Health Information Technology

The adoption of Health Information Technology (HIT) is key in improving the quality of healthcare. HIT is described by Rao and colleagues (50) as "the technologies that allow a provider to capture information electronically through Electronic Health Record (EHR) systems and be able to share information among providers and engage patients in innovative ways". HIT is an important tool for improving healthcare quality and safety (51). EHR systems serve an important role in facilitating the making of decisions about health outcomes with current information (52). EHR systems are used to detect hidden patterns and trends, monitor patient conditions (53), reduce medical errors (54), reduce adverse drug events (55) and ultimately improve quality of care (56).

EHR systems can improve patient care by providing information tailored to each individual patient especially patients with chronic diseases (50). EHR systems allow the capturing, storing, and sharing of a patient's health information between healthcare providers (57). However, despite the evidence showing the benefits of

EHR systems, they rarely improve healthcare experts' ability to make better clinical decisions by having access to more comprehensive information (58).

The volume of clinical data being captured for each patient on the EHR systems poses many challenges to (a) clinicians trying to make sense of the patient's condition within the context of the patient's medical history, (b) administrators trying to make decisions grounded in data, (c) researchers trying to understand differences in population outcomes, and (d) patients trying to make use of their own medical data (59). Moreover, African countries have not been able to successfully implement a national EHR system.

In South Africa, different healthcare providers use different EHR systems which makes it difficult to share health information between providers (60). This challenge is further compounded by the fact that in some areas, particularly rural areas, healthcare facilities still use paper-based health records (61). Where in existence, the different EHR systems are built with different underlying architectures that lack interoperability (62) and most of them lack information visualization features to facilitate the provision of personalized care.

To overcome this challenge, health information can be made accessible to users through interactive visualizations.

1.2.4 Visualizations

Visualizations can be defined in several ways. Information visualization is defined by Cairo (63) as "techniques that support the understanding of abstract data" and have been used in ways that present information to improve patient-provider communication (64). Data visualization on the other hand is defined by Kirk (65) as "the representation and presentation of data that exploits our visual perception abilities in order to amplify cognition and involves information exchange".

An effective visualization tool is essential in supporting the human ability to recognise patterns that are not easily detected (66) and provide valuable insights to the data. Backonja, Haynes and Kim (67) further state that visualization can help users overcome their cognitive limitation and help support decision making. West, Borland and Hammond (68), suggested that innovative visualization techniques used to display longitudinal data from EHRs have a potential of discovering new information in the data through the use of exploring other visualization techniques other than tables, charts, and graphs. The effectiveness of a visualization according to Saket, Endert and Demiralp (69) depends on several factors including the task at hand, dataset used and data attributes, as one chart that might be suitable for answering a specific type of question such as examining the correlation between two data attributes, may not be suitable for other purposes, like identifying a data point with the highest value.

According to Shneiderman, Plaisant and Hesse (70) information visualization is highly relevant to healthcare and a key aspect of health informatics. It has also been shown to reduce information overload and improve recall (71). The common characteristics of visualizations focuses on readability, recognizability and meaning (72). Moreover Yarmuluk (73) stated that the purpose of visualization is to quickly communicate data summary or highlight key trends in the data to allow the audience to identify trends and perform comparisons. In addition, a visualization lacking in graphical integrity (a visualization that may appear to be visually impactful but may not be completely truthful) may give a certain impression visually but the same may not be supported at the data level. Furthermore, Caban and Gotz (59) stated that the utilisation of EHR data in interactive visualizations has demonstrated effectiveness for various clinical tasks such as examining disease risk factors and prediction of health outcomes.

Some of the advantages of data visualization according to Li (72) is that it has the ability to represent a vast amount of data immediately, it enables users to identify emergent patterns in the data to formulate new insights and it enhances the understanding of large-scale and small-scale data.

In addition, there is also visual analytics systems as an emerging discipline and a new set of computational tools.

1.2.5 Visual Analytics Approach

Visual Analytics (VA) provide analytical reasoning of abstract data through interactive visual interfaces. It is a term that was formally introduced in the literature over 15 years ago (74). VA is designed to promote knowledge discovery and utilisation of big data through the implementation of visual paradigms and well-designed user interactions (75). Chung and colleagues (76) noted that the implementation and use of VA have bloomed in many sectors of the healthcare system during the past decade.

According to Munzner (77), VA has the potential to enhance the user's confidence in the prediction of results by improving their understanding of the modelling process and results. Rostamzadeh and colleagues (78) further state that VA allows users to access, modify and restructure the displayed data as well as guide the data analytics techniques. VA systems have the ability to offer cognitive and computational assistance to the user by combining machine learning techniques, analytical processes and visualizations (79). Angulo and colleagues (79) summarise VA as being composed of two integrated modules: an analytics module and an interactive visualization module.

The analytics module is technology-independent and it includes the use of processing techniques and data mining algorithms that fits the needs of the domain (78) while the interactive visualization module in VA involves the mapping of processed and derived data from the analytics module to visual structures (80) and it provides the user with new analytical possibilities that can be utilized in an interactive manner (56).

One of the visual analytics tools is Tableau. It is considered to be one of the best platforms for data visualization and is one of the leading software in visual analytics.

It was founded with a mission of helping people understand their data by improving the flow of analysis and make data more accessible to people through visualization (81). Qlickview is another major player in the industry. It provides an easy way of discovering and understating insightful data for visualization (82). SAS has been in the industry for a long time and is quick in interpreting complex data sets with a great degree of accuracy (83). Another tool is R; an open-source statistical and programming environment that incorporates all statistical tests, models as well as analysis with outstanding graphical capabilities (84). In R, there is a package called Shiny that simplifies the process of building interactive web applications directly within the R environment (85).

According to Simpao and colleagues (86) VA systems have the potential to help reduce the complexity of EHRs by integrating automated analysis techniques with interactive visualizations. Furthermore Abdullah (87) stated that VA systems can help with the analysis, interpretation, and making sense of EHR databases by improving the capabilities of the user to accomplish complex data-driven tasks. VA systems show significant potential in analysing and making sense of EHR data from patients with multiple chronic diseases (88).

1.3 Problem Statement

Like elsewhere in Sub-Saharan Africa, South Africa is experiencing a steady increase in the prevalence of chronic diseases. However, patterns of healthcare utilisation and disease progression are poorly characterized especially in rural settings. This is mainly due to lack of official national EHR system to register and monitor patients who seek care for chronic disease in primary healthcare facilities in South Africa. In some cases, EHR systems have been deployed in primary healthcare facilities for patient management and research purposes. However, in most cases the data captured in these systems are usually complex and difficult for healthcare providers to quickly analyse, recognise trends and derive empirical evidence to inform effective decision-making (89,90) that can reduce the burden of chronic diseases. This challenge can in part be addressed by data visualisation systems (dashboards) which as defined by Malik (91) are “rich user interfaces that display information in a graphical form using a variety of elements including charts, tables, and gauges”. As is the case for most rural areas in South Africa, there is a lack of visual analytic techniques applied to data collected from primary healthcare facilities in the Agincourt sub district, rural northeast South Africa to characterise patterns of healthcare utilisation and disease progression among adults with chronic diseases.

1.4 Research Question

The main research question investigated in this study is: What insights do visual analytics provide regarding the patterns of healthcare utilisation and disease progression among adults with chronic diseases in Agincourt sub district, rural northeast South Africa?

1.5 Aim and Objectives

The aim of the study is to utilise a visual analytics system to characterise patterns of healthcare utilisation and disease progression among adults with chronic diseases in Agincourt sub district, rural northeast South Africa. The specific objective include:

i. To develop a visual analytics system for characterising healthcare utilisation patterns and chronic disease progression.
ii. To utilise visual analytics to characterise patterns of healthcare utilisation among adults with chronic diseases in Agincourt subdistrict, rural northeast South Africa.

iii. To utilise visual analytics to characterise the progression of chronic diseases among adults in Agincourt subdistrict, rural northeast South Africa.

1.6 Motivation

There is evidence that interactive visualizations of healthcare and disease outcomes data has the potential to help healthcare providers with the analysis and interpretation of patients' health information and consequently lead to improvements in the provision of care and disease outcomes (92). This study provides insights into patients' interaction with the healthcare system for chronic health conditions and responses to treatment provided to them. This can assist patients and practitioners in making sense of personal health information. The interactive visualizations produced in this study will play an important role in making data regarding utilisation of healthcare services and progression of chronic diseases digestible which can lead to appropriate interventions.

1.7 Outline

This research report is organised into five different chapters. The current chapter (chapter one) is followed by chapter two where we look at the methods used in the study to develop the VAS; chapter three will provide the results of the VAS developed, chapter four will discuss the findings and an overview of how the VAS will be deployed and lastly, chapter five will provide the conclusion derived from the findings of this study and recommendations.

CHAPTER 2: METHODS

This chapter focuses on the methodology used in the study by describing the setting and context where the study took place, the population, and the dataset. It also describes the details of variables used, data pre-processing and an overview of the system. The chapter concludes with an overview of ethical considerations that were taken into account while handling the data.

2.1 Study Setting and Population

The study used data from the Agincourt Health and Demographic Surveillance System (HDSS) study area. The area is situated in a predominantly rural setting in Bushbuckridge municipality in northeast South Africa in Mpumalanga Province (Figure 1). The population of the Agincourt HDSS has been under surveillance since 1992 (93) and as of August 2022 comprised of about 120 000 individuals living in 21 000 households distributed across 31 villages (94). A third of the population are Mozambican immigrants who arrived in the area in the early to mid-1980s, and their descendants (95). Shangaan is the widely spoken language in the area. There are eight primary healthcare facilities (Agincourt, Bhubezi, Lillydale, Xanthia, Belfast, Kildare, Cunningmore and Justicia) within and three district hospitals (Tintswalo, Mapulaneng and Matikwane) 25-60 km outside the Agincourt HDSS area (96).

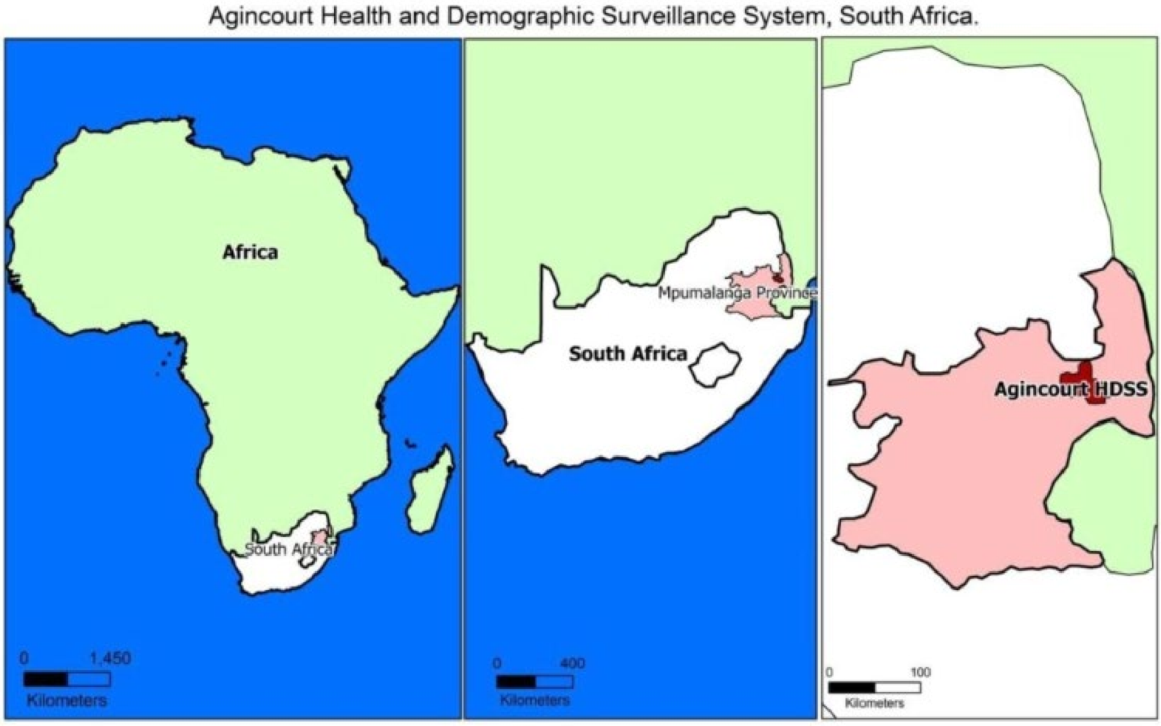

Figure 1: Agincourt HDSS Study Site

Source: https://www.agincourt.co.za/?page_id=1896

2.2 Data Sources

The study used data captured in the Agincourt HDSS-Clinic Link System. The Agincourt HDSS-Clinic Link System has been in full operation since 2015 and is an ongoing longitudinal data platform that captures and stores clinical information collected from primary healthcare facilities in the HDSS study area and links it to the Agincourt HDSS data (97,98). The data was used upon receiving ethical clearance from the University of the Witwatersrand Human Research Ethics Committee (HREC) and a letter of authorization from Agincourt HDSS to use the data.

2.3 Data Pre-Processing

The quality of data-driven research depends on the preparation and pre-processing of the raw data. Data pre-processing plays a fundamental role in ensuring the accuracy of the findings, as it addresses challenges such as missing values, outliers, and inconsistencies that might exist in the raw dataset. Through careful data cleaning, transformation, and feature selection, we aimed to mitigate the potential impact of noise and reduce any bias that might have affected the raw data. In this section, we detail the steps taken to clean, transform, and organize the data before using it in the development of the VAS. We describe the process of acquiring the data and its sources, along with the various techniques used for data cleaning, transformation, and integration. Additionally, we explain the rationale behind the selection of specific variables and the methods employed for encoding categorical variables.

Firstly, data inspection was done to gain insights on the quality of the data. Using the visit year variable, we dropped the visit years outside of the study period (2015-2021) and the health facilities that are not part of the study area.

A new binary variable denoted as "High BP" was created from two existing variables (systolic bp and diastolic bp) within the dataset. The purpose of creating this variable was to categorize individuals based on their blood pressure (BP) status, specifically identifying those with controlled hypertension. Based on the defined thresholds, individuals who met the criteria for controlled hypertension were assigned a value of 0, representing the absence of high blood pressure. Conversely, individuals whose blood pressure measurements fell above the defined thresholds were assigned a value of 1, indicating the presence of high blood pressure.

Another binary variable termed "Viral suppressed" was created by combining information from two existing variables within the dataset, namely "Viral load results" and "Undetectable viral load". The objective of creating this variable was to identify individuals with a viral load less than 400 copies per millilitre (copies/mL) threshold and undetectable viral status, as it holds crucial significance in the context of viral infection management and treatment efficacy assessment. Individuals meeting both conditions, i.e., having a viral load less than 400 copies/mL and an undetectable viral status (Undetectable viral load=1), were assigned a value of 1, signifying the presence of a low viral load. Individuals not meeting both conditions were assigned a value of 0, indicating the absence of a low viral load.

We identified missing values in the dataset by using appropriate functions and techniques available in R Shiny. Missing values are often denoted as "NA" (Not Available) in R programming language, and we also considered cases where empty strings ("") represented missing data.

To ensure the robustness of our analyses, we implemented a standard approach for handling missing values. For numeric variables: we utilized the "na.rm = TRUE" argument in relevant functions to remove rows with missing values from numeric variables. By setting this argument, we ensured that statistical calculations were performed on complete cases, thereby preventing the inclusion of missing values in the computations.

For character or string variables: we treated empty strings ("") as missing values and removed rows with these empty values using the same "na.rm = TRUE" approach.

By employing the "na.rm = TRUE" argument and considering empty strings as missing values, we effectively managed the presence of missing data in our R Shiny application. This data pre-processing step allowed us to maintain the integrity of our analyses, as missing values were appropriately handled during statistical computations and visualizations.

2.3.1 Study variables

Table 1 below shows the key variables used in the study and their descriptions.

Table 1: Definition of study variables

Variable	*Description*
Patient ID	Unique number assigned to each patient
Health Facility	Name of health facility
Sex	Gender of patient (Female, Male)
Date	Date of clinic visit
Diagnosis 1-5	Name of diagnosis made at each visit
Diabetes	Boolean field indicating the presence (Yes) or absence (No) of condition

HIV	Boolean field indicating the presence (Yes) or absence (No) of condition
Hypertension	Boolean field indicating the presence (Yes) or absence (No) of condition
CD4 Count results	Numerical reading for CD4 count
Viral Load results	Numerical reading for Viral load results
Diastolic BP	Numerical reading for Diastolic
Systolic BP	Numerical reading for Systolic
Weight	Numerical reading for Weight
Blood Glucose	Numerical reading for Blood glucose level
Treatment 1-10	Medication given as treatment 1 until 10
Visit year	Year of the visit date
Age group	10-year age group variable
High BP	Boolean field indicating uncontrolled blood pressure (1 for uncontrolled, 0 for controlled)
Viral Suppressed	Boolean field indicating undetectable viral load (1 for suppressed, 0 for unsuppressed)
Controlled Diabetes	Boolean field indicating controlled diabetes (1 for controlled, 0 for uncontrolled)

Note: "High BP" is a created variable that refers to uncontrolled blood pressure readings, while "Hypertension" is a broader term showing the condition of chronically elevated blood pressure.

2.4 Overview of the Visual Analytics System

The system was developed with the aim of assisting healthcare providers in gaining a deeper understanding of disease trends, patient demographics, and healthcare facility performance. The functionalities of the system were sought from healthcare providers as primary stakeholders of the visual system. The system was developed in R (programming language) and utilized R Shiny for the visualizations using the functionalities proposed by the stakeholders.

2.4.1 Functionality Requirements

The researcher was part of two stakeholder meetings held at Wits Rural Facility in Acornhoek, Mpumalanga South Africa in August 2022 and Naivasha, Kenya in August 2023. The objectives of the meetings were to learn about participants' data workflows and discuss different ways of data presentation and visualisation at different levels. There were 24 stakeholders in the meeting at Wits Rural Facility comprising of senior managers, operations managers, professional nurses, and data capturers. While in Naivasha, there were 27 stakeholders comprising of head of units, coordinators, directors, and Health Records Information Officers (HRIOs). These group of stakeholders provided a significant level of expertise to inform the design of the system. There were a few core requirements proposed by the stakeholders including:

- Providing patient summary
- Provide staff performance summary
- Notification of lost-to-follow-up
- Facility summary status

The stakeholders' input was imperative to the final product as attending this type of meeting helped the researcher gain new insights on how to develop an effective VAS with the aim of helping end-users better understand the clinical data they collect.

2.4.2 R (Programming Language)

R is a programming language and environment for statistical computing and graphical representation to analyse and visualize data. R is an open-source language primarily used for data analysis and statistics however it provides a wide range of libraries that make it easy to perform various statical calculations and visualizations. The researcher chose to use R because of its versality in using packages and libraries (discussed in section 2.4.5) to solve specific data tasks including visualization capabilities by creating charts and plots. Furthermore, the many available online resources and tutorials within the R community made it possible for the researcher to quickly learn how to design the VAS.

2.4.3 R Shiny

R Shiny is an R package that allows one to create interactive web applications and dashboards from R. It allows researchers to turn their R code into interactive web applications without needing to know web development languages such as HyperText Markup Language (HTML), JavaScript or Cascading Style Sheets (CSS). When using R shiny, one can create interactive elements such as sliders, select input boxes, buttons, and plots that respond to user input in real-time using their data and has reactivity capabilities when input data changes, it automatically updates the application or dashboards.

2.4.4 Architecture Components

The development of the system's architecture involved the integration of various components to ensure the seamless functioning of the VAS. These components played a pivotal role in designing an interactive and user-friendly platform for analysing and visualizing the Agincourt HDSS-Clinic dataset. In this section, we delve into the key building blocks that constitute the architecture components of the system. Each element's role and contribution to the overall functionality of the VAS is explained, shedding light on the details of its design and implementation. The successful combination of these components facilitated the creation of the final product that users can use to explore the data comprehensively and obtain valuable insights. Below, we detail the integral components that form the backbone of this R Shiny-based VAS.

- User Interface (UI) object: The UI controls the layout and appearance of system. It was designed using R Shiny's layout functions and widgets such as tab panels and fluid rows. The layout structure consists of the following tabs:
    - About
    - Summary
    - Health Facility Dashboard
    - Trends by Time

    - Patient Dashboard
    - Support

  Each tab has its own layout using fluid rows to add specific charts on the tabs.

- Server Function: The server-side of the application contains the instructions needed by the computer to build the system. It is where data processing, computations, and interactions with the dataset takes place. The reactive elements and functions used include:
    - Input – Inputs provided by the user for the server function to perform computations.
    - Output - Defines the outputs that the server function generates such as render plot based on the selected input.
    - Session - Creates reactive expressions and controls reactivity.

- Data Handling: The dataset was loaded using the haven library; an R package for reading dataset from other statistical packages. Data filtering was performed at each reactive stage based on what was expected for that plot.

- Visualizations: The types of visualizations and plots used in the VAS to present the data include Upset plots, value boxes, bar graphs, gauges, line graphs and stacked bar graphs. Each plot was created using libraries such as ggplot2, plotly and upsetR to produce an output function based on the filtered data used.

- Interactivity: The VAS uses interactive features set up on the server function to allow users to interact with the visualizations and manipulate the data by using the input functions on each tab.

- Call to the shinyApp Function: This is the function call that creates and runs the R Shiny application.

Figure 2 below shows how a user can interact with the system. The user engages with the system from their devices where they land on the UI of the system. Using the select input, the user selects for example the year they want to filter the data to display, this goes to the server and renders a plot output back to the UI for the user to see on the devise.

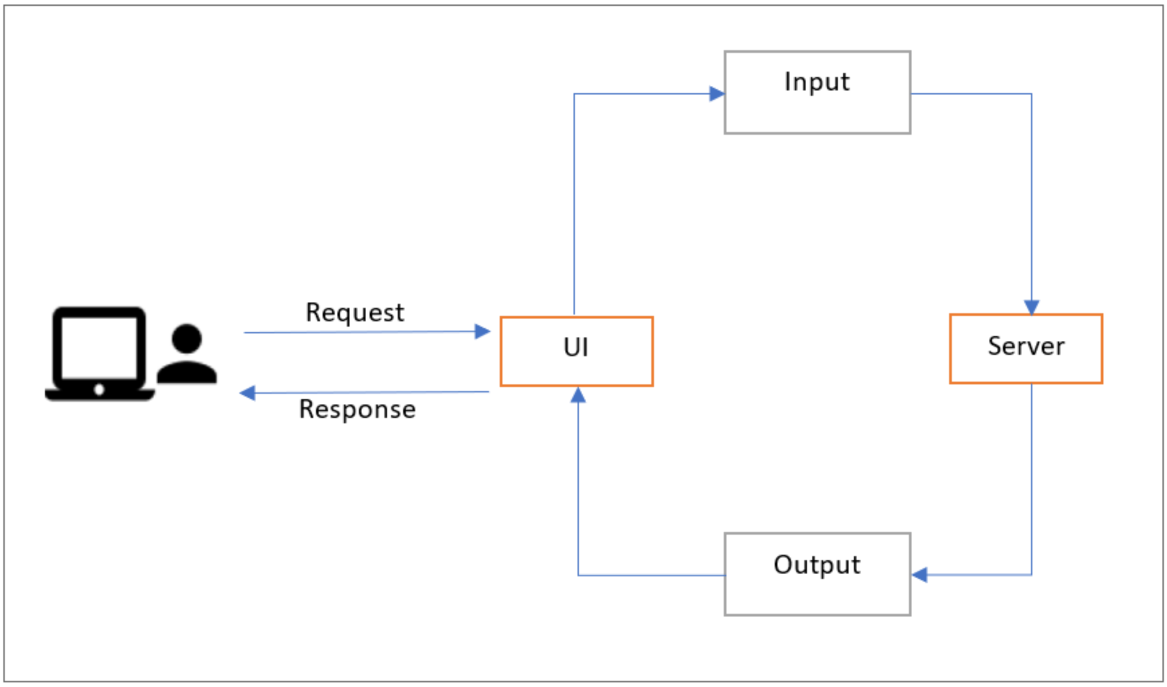

Figure 2: VAS Architecture

2.4.5 Software Dependencies

R is extended using packages available through the Comprehensive R Archive Network (CRAN) family, an R software repository for packages. R packages increase the functionalities of the base R. Table 2 below shows all the packages used in developing the VAS in R. Before using the package in R, one needs to make sure that the package has been installed. To install a package in R, type the install package code on the console: *install.package("name of package").* Once the package has been installed; it should be called before it can be used on your console using for example: library(shiny).

Table 2: R Packages used

Package	Definition
library(shiny)	Builds interactive web applications with R
library(shinyboards)	Used to create interactive dashboards in R Shiny

library(shinywidgets)	Offers additional user interface (UI) elements to enhance functionalities
library(shinythemes)	Provides a selection of different themes to use
library(tidyverse)	Provides a framework for data manipulation, visualization, and analysis
library(ggplot2)	Used for data visualization by creating plots and charts
library(haven)	Allows users to import and export data from statistical packages
library(dplyr)	Is used for data manipulation such as filtering and arranging data
library(gcookbook)	Helps users learn and implement various data visualizations using ggplot2
library(UpSetR)	Creates an UpSet plot that shows intersection and of sets in a dataset
library(ploty)	Supports a wide range of chart types and interactivity options created for web-based visualizations
library(gridExtra)	Is used for arranging multiple ggplot2 plots on a single page by allowing users to combine multiple plots into one
library(reshape2)	Used for data manipulation and transformation (reshaping data frames).

2.4.6 Deployment

The VAS has been deployed on the Agincourt HDSS servers and is accessible using this link: https://pm.agincourt.co.za/. The source code is hosted on GitHub, a web-based platform widely used for version control and collaborative software development. GitHub allows for efficient versioning, enabling easy tracking of changes and facilitating code reuse. Accompanying documentation is provided on the GitHub repository on hosting the application and understanding its functionalities.

2.5 Ethical Considerations

The study received ethical approval from the University of the Witwatersrand Human Research Ethics Committee (Medical), clearance certificate number: M220668. A de-identified dataset was requested and received from the data gatekeeper to ensure confidentiality. The dataset was stored in a secured password-protected computer. The data was used for the purpose of this research and its associated publications at the university. The information about participants was kept private and confidential and used only for this study.

2.6 Summary

In Chapter 2, we detailed the comprehensive methodology employed in this research, encompassing various aspects from study setting to the development of the VAS. The key elements covered in the methods chapter include the study setting where the data collection took place, the study population, an overview of the data sources utilized, how the data was processed before utilizing it in the VAS and an overview of the VAS, it’s functionality requirements, dependencies and ethical considerations while handling the data.

By integrating these various components, our research methodology enabled the comprehensive exploration of the Agincourt HDSS-Clinic dataset while ensuring ethical integrity and providing a user-friendly interface for data visualization. The next chapter looks at the results of the developed VAS that serves as a tool for gaining insights and drawing evidence-based conclusions to address the research objectives.

CHAPTER 3: EXPLORING THE FUNCTIONALITY OF THE SYSTEM

In this chapter, we present the results of our VAS designed in R shiny. We will walk through each tab in detail explaining its purpose, features, and functionalities. Furthermore, we will showcase how users can interact with the system, the visualizations available, and the dynamic capabilities for data exploration.

By providing an in-depth examination of each tab, we aim to demonstrate the full potential of system and its role in fostering a deeper understanding of the dataset.

3.1 About Tab

The "About" tab of the VAS provides information about the system and its purpose. It provides the context, objectives, and key features available on the system.

3.2 Summary Tab

This tab offers a comprehensive overview of the Agincourt HDSS-Clinic dataset, presenting summary statistics and key insights at a glance. Users can interactively select the year and filter data subsets to gain an understanding of the dataset's distribution and characteristics. We will discuss the visualizations and interactive elements that enable users to explore the data from multiple perspectives.

Figure 3 below presents information about the demographic characteristics of the patients. It shows the total number of patients and their gender distribution.

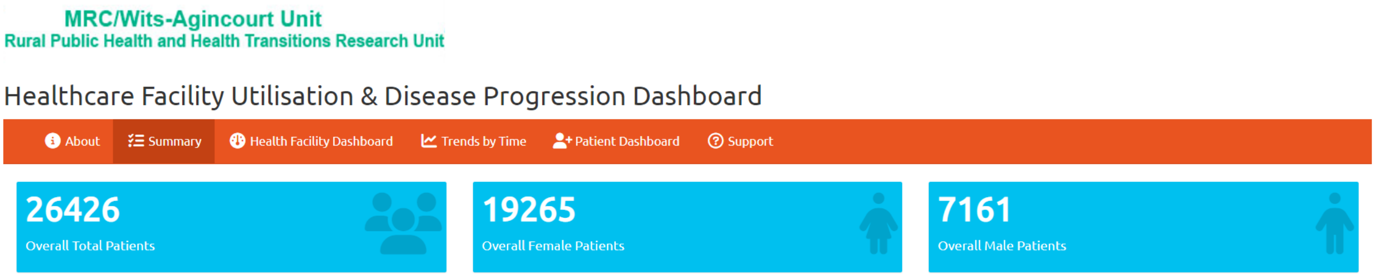

Figure 3: Overall patient count

Using the drop-down field to select the year (filter by year), one can filter the data using the year of visit to display interactivity plots. Figure 4 shows how many patients have the three conditions (HIV, Hypertension and Diabetes) and how often they co-occur within the Agincourt HDSS-Clinic dataset.

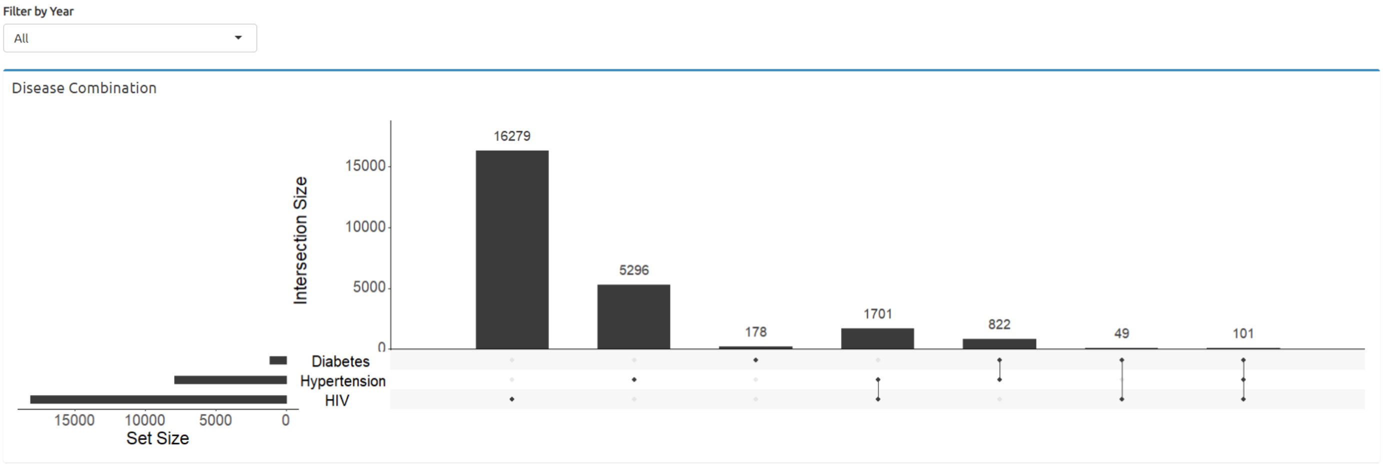

Figure 4: Disease combination Upset Plot

The following figure (Figure 5) presents a set of interactive value boxes that provides the overall patient count and their distribution across different genders. Subsequently, the bar graph that follows shows the total patient count at each healthcare facility. Moreover, we delve into the patient population's age distribution, further categorized by gender.

Figure 5: Patient healthcare utilization, and distribution of patients by age plots

Figure 6 below shows the age distribution of the patient population, with a specific focus on each condition. The bar graphs provide a breakdown of patient ages, stratified by gender, within the context of each condition.

Each plot serves as a detailed window into the age demographics of patients affected by a particular condition. This figure provides an understanding of how age and gender factors interplay in relation to these specific chronic conditions.

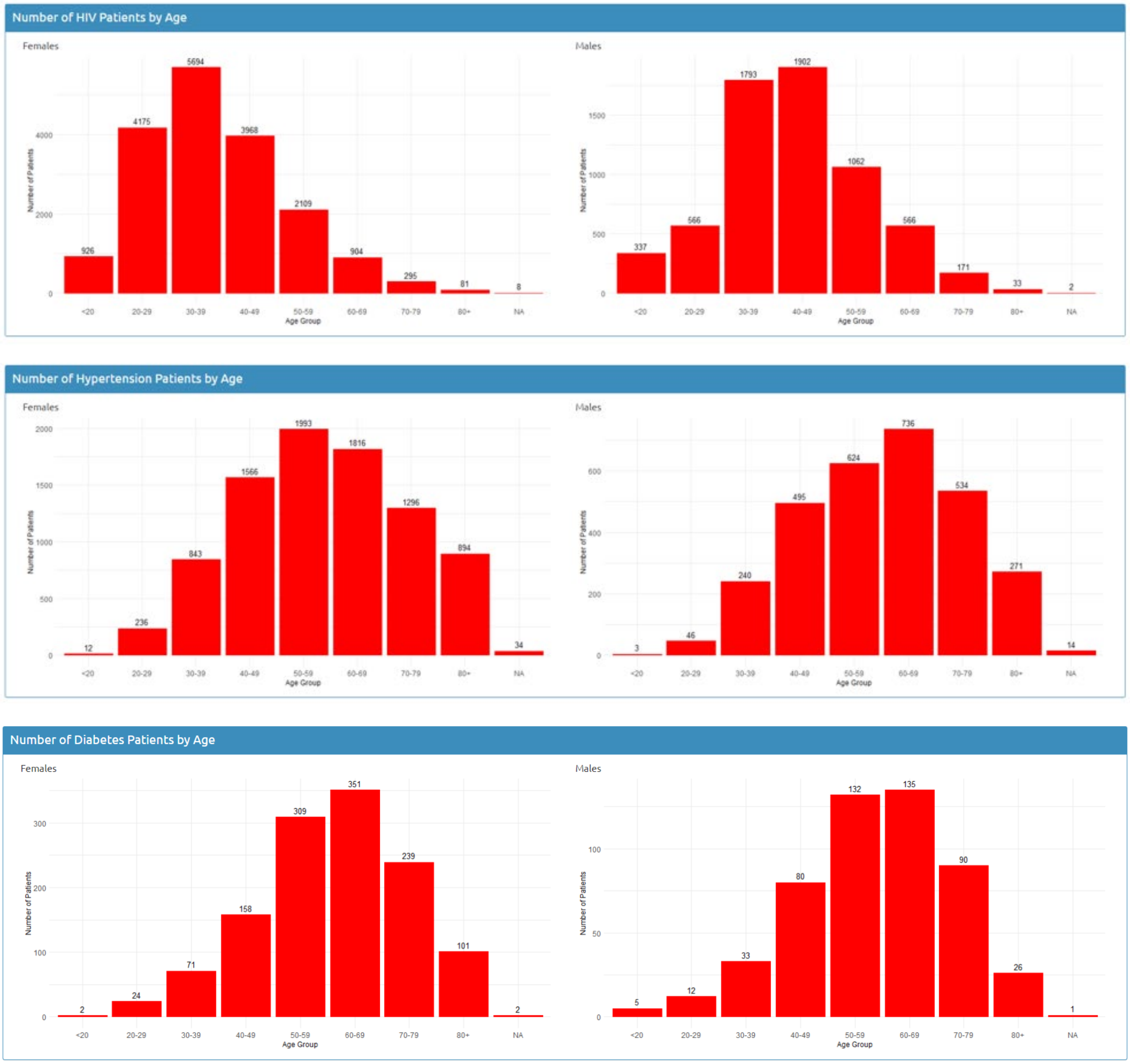

Figure 6: Number of patients by age for each chronic condition

Lastly on this tab, we have gauge plots to effectively communicate key health indicators within the Agincourt HDSS-Clinic dataset. The plots present average values and percentages of critical health metrics offering a visually impactful representation of these measurements.

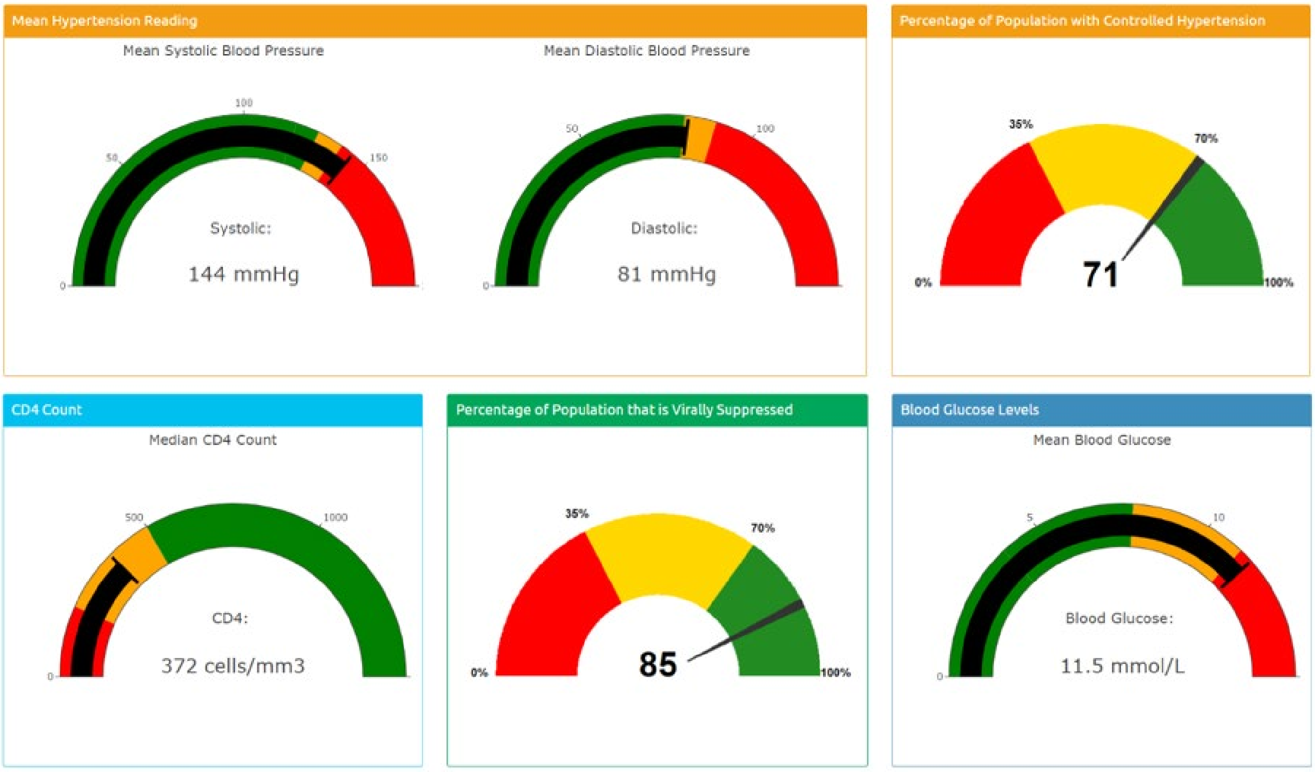

Figure 7: Mean scores and percentage of controlled hypertension and virally suppressed population

3.3 Health Facility Dashboard Tab

This tab starts off with a static bar graph plot that shows a summary of the total patient count at each health facility (Figure 8).

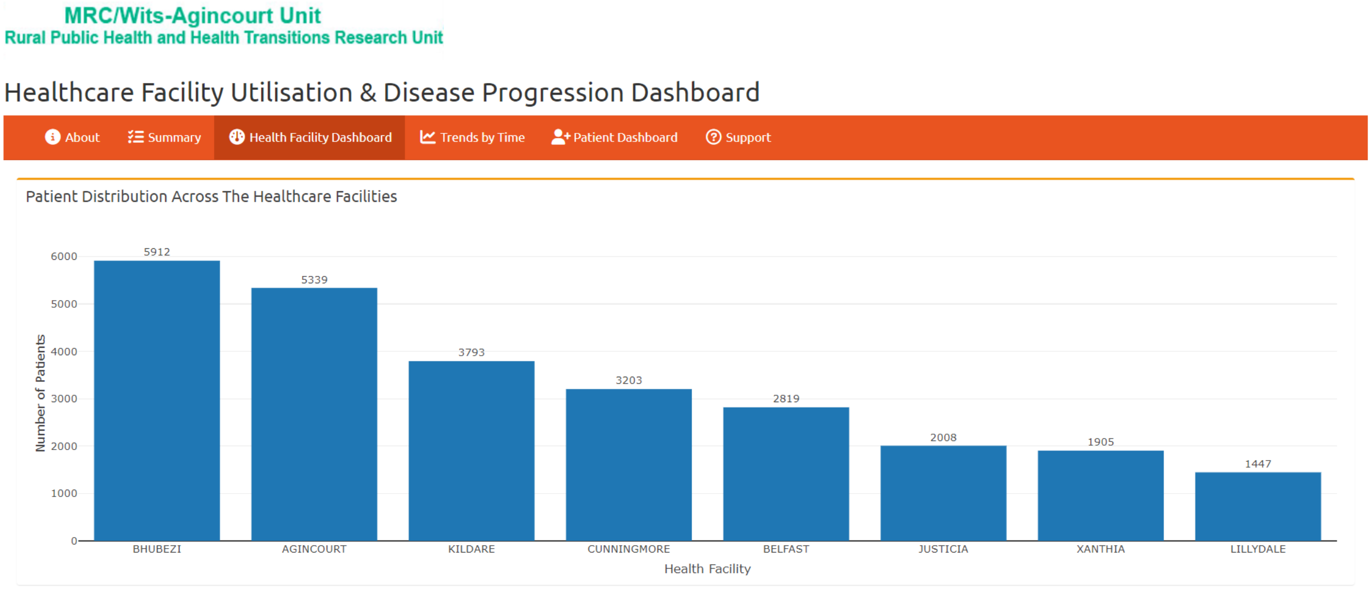

Figure 8: Number of Patients by Health Facility

Figure 9 summarizes the comprehensive visualizations within this tab, using the integration of two dynamic selection inputs: the health facility and the year of visit. These interactive inputs provide the visual elements with interactivity, enabling users to explore the Agincourt HDSS-Clinic dataset through an array of informative metrics. As a result, the figure portrays essential insights, including patient count, age distribution, and the proportion of patients achieving suppressed viral load, controlled hypertension, and managed blood glucose levels. These dynamic attributes offer a subtle perspective on specific chronic conditions for a given year, thereby enhancing the depth and precision of the visual narrative.

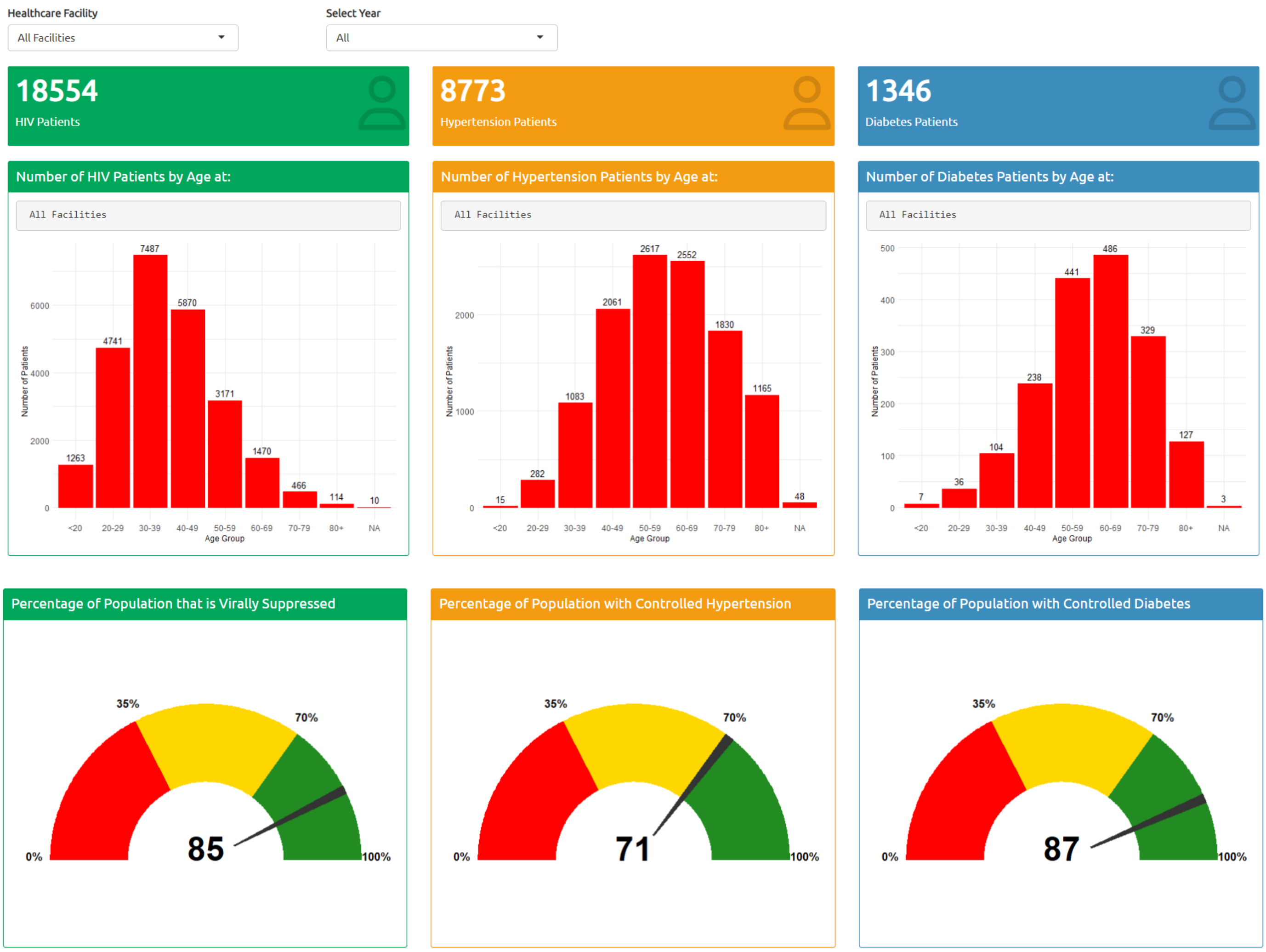

Figure 9: Number of Patients at Health Facilities and percentage of population with viral suppression, control hypertension and control diabetes

3.4 Trends by Time

In the "Trends by Time" tab of the system, we explore the temporal patterns of rates of each condition within the population and healthcare facilities.

Figure 10 displays the condition rates over time for the overall population. Each point on the graph represents a specific period (2015-2021), and the y-axis represents the condition rate (%). This plot helps to assess how the prevalence of the condition changes over time in the clinic population.

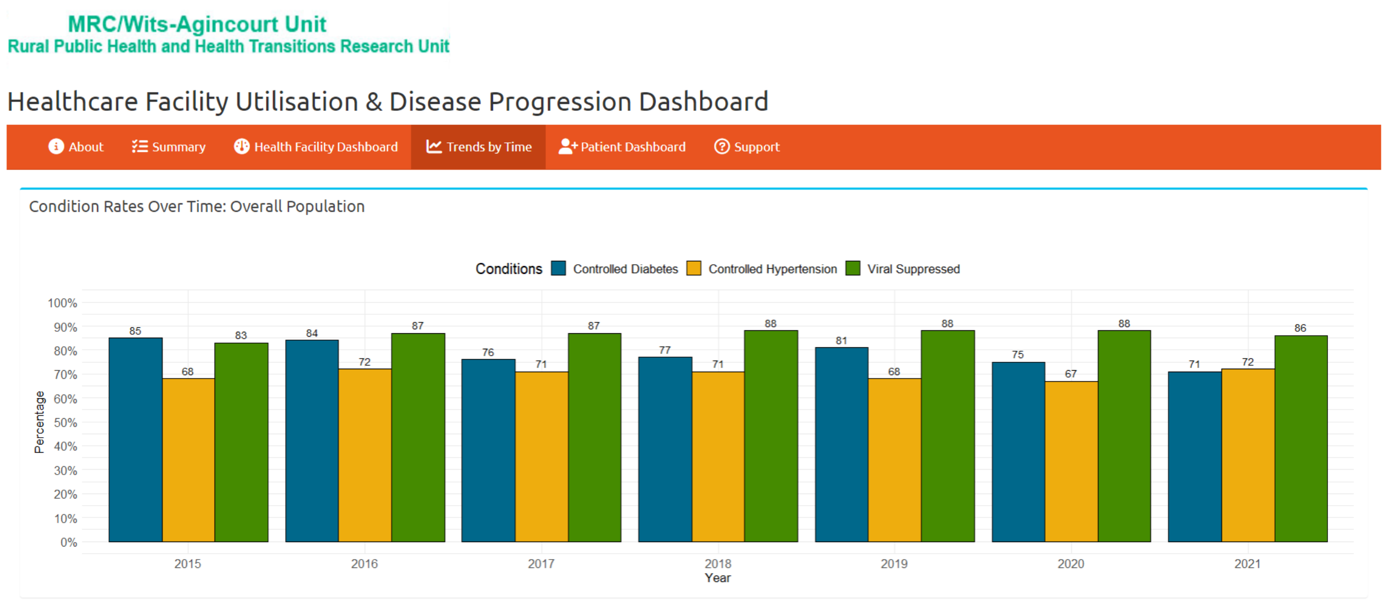

Figure 10: Percentage of condition rates over time for the overall population

Figure 11 focuses specifically on condition rates within healthcare facilities. Similar to Figure 10, each point represents a time period, and the y-axis shows the condition rate (%). This plot provides insights into how the prevalence of the conditions evolve within healthcare settings.

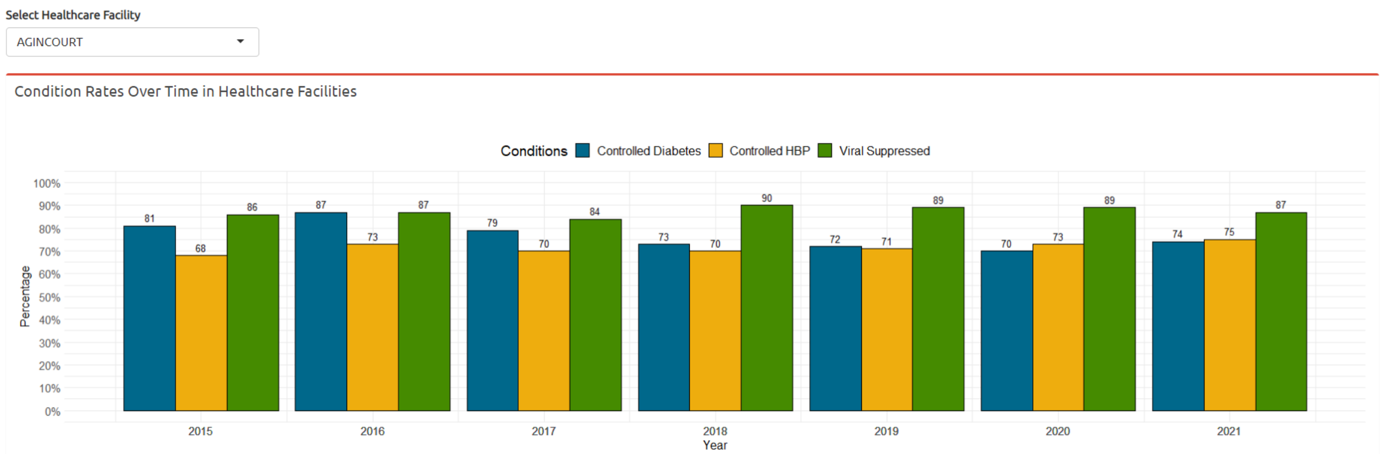

Figure 11: Percentage of condition rates over time by health facility

By exploring these two plots, we gain a better understanding of how condition rates change over the selected time-period and within healthcare facilities.

3.5 Patient Tab

At the heart of the patient tab lies visualization tools that empower the user to explore individual patient journeys with fine granularity. This tab is anchored by two pivotal inputs (Figure 12): ‘Select year’ and ‘Select Patient ID’. By using these two inputs, one can dynamically generate a diverse array of plots that provide a personalized lens into medical histories captured during patients' visits.

The ‘Select Year’ input allows one to pinpoint specific periods of interest, zooming in on the evolution of patients' conditions over time. Complementing this is the 'Select Patient ID' input, which facilitates the selection of individual patients to delve into their unique health trajectories. The resulting interactive plots from manipulating the inputs illuminate the course of patients' health, showcasing, patient information, trends, medication given, and outcomes associated with their visits.

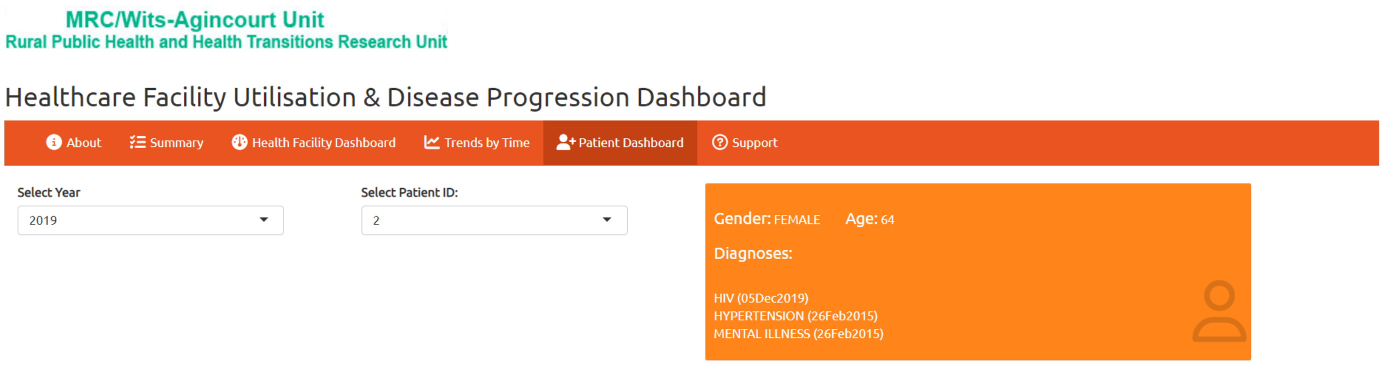

Figure 12: Select year and Patient ID input filters displaying patient information

The above figure shows patient disease progression through the years of being in care. Using the two inputs, we get to display the patients’ diagnosis for the current year and diagnosis for the previous years.

Figure 13 shows two distinct line graphs. The first plot displays the trend in a patient’s weight over time, capturing patterns that illuminate potential correlations with specific dates within the visit year. The second plot combines systolic and diastolic blood pressure trends, presenting a holistic view of the evolution of a patient’s cardiovascular health.

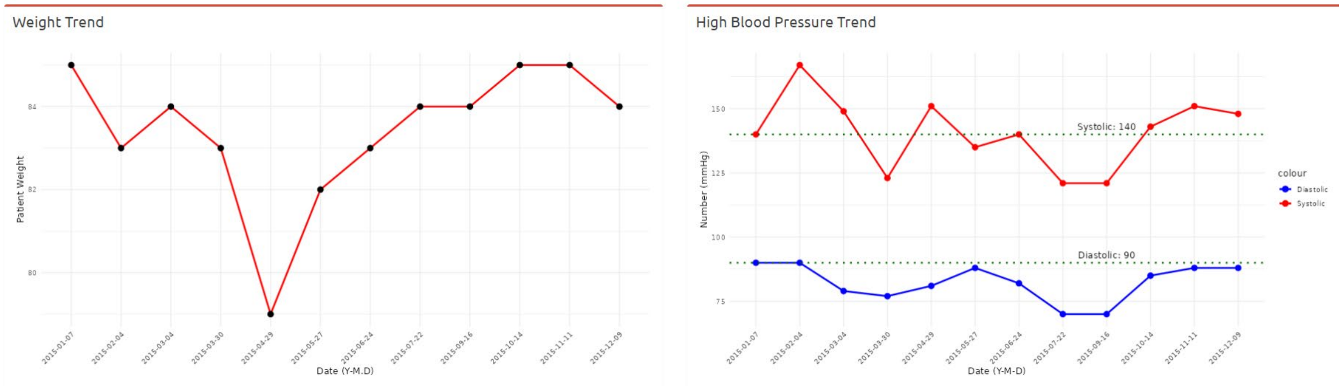

Figure 13: Weight and High Blood Pressure trend plots over time

On Figure 14, the first plot shows the CD4 count trend over time, capturing fluctuations and patterns that shed light on patients' immune response variations across different visit months. The second plot portrays viral load trends, presenting insights into the effectiveness of treatment regimens and disease management.

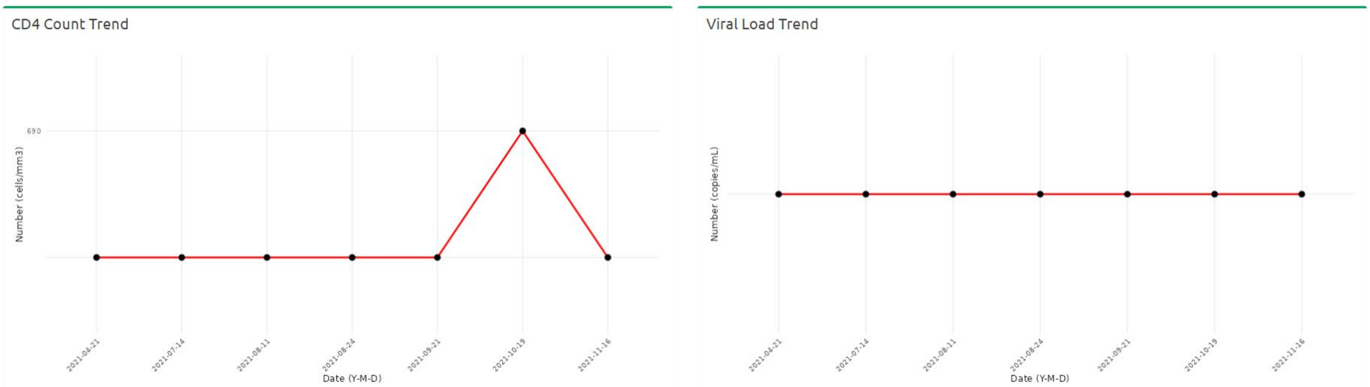

Figure 14: CD4 count and Viral load trend plots over time

Lastly on this tab, figure 15 shows two insightful graphical elements. The first component features a dynamic line graph, illustrating the blood glucose trend across different visit years. This plot unveils patterns, spikes, and trends in blood glucose levels, aiding in the assessment of diabetes management over time.

Complementing this, is a stacked bar graph, showcasing medications administered at each visit. This vivid presentation not only provides a snapshot of the medication landscape but also offers a clear understanding of the distribution and changes in prescribed treatments over time.

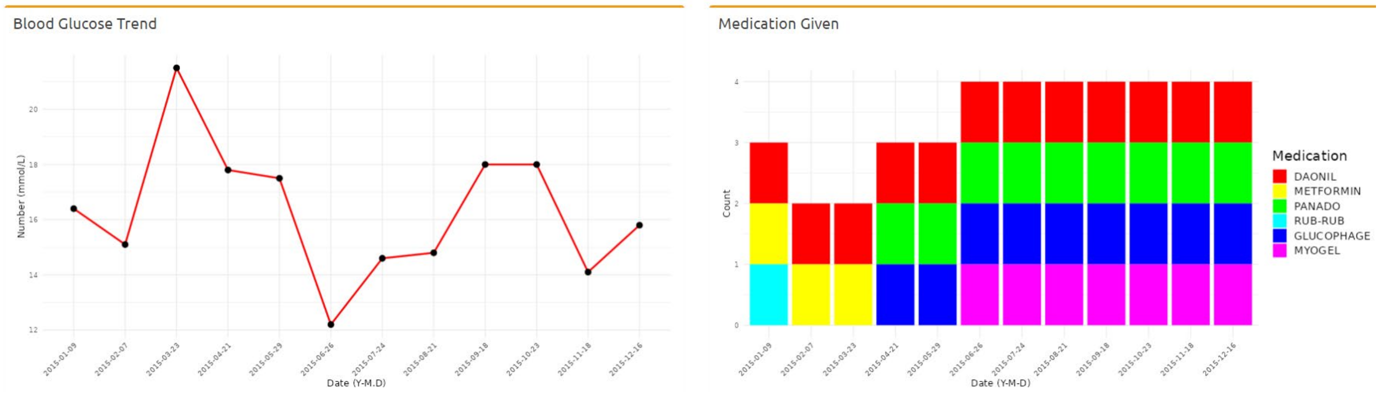

Figure 15: Blood glucose trend and Medication given plots over time

3.6 Support Tab

This tab completes the VAS by providing a support email address and a link to the Agincourt HDSS homepage for more information about the data.

3.7 Summary

Through this chapter, we have demonstrated the rich functionality and interactive nature of our VAS. Each tab offers distinct insights and analytical capabilities, enabling users to gain a deeper understanding of the Agincourt HDSS-Clinic dataset and extract meaningful information for informed decision-making. In the next chapter we discuss the results of the system and the data represented by the plots.

CHPATER 4: DISCUSSION

In this Discussion chapter, we delve into a comprehensive interpretation of the results derived from the development of the VAS. Building upon the insights presented in the preceding chapters, this chapter provides a detailed exploration of the outcomes obtained from the various visualization tabs.

Each tab within the VAS comes under scrutiny, with specific emphasis on the interplay between the different variables and the contextual narratives they paint. With a careful examination of the Summary, Health Facility, Trends by Time, and Patient tabs, we aim to bridge the gap between data and understanding and through a step-by-step exploration of the insights presented, we reveal patterns and trends that are used to shape our comprehension of disease progression, healthcare dynamics, and patient outcomes.

4.1 Summary Tab

Overall, the Agincourt HDSS-Clinic dataset encompassed a total of 26 426 patients, forming the foundation of our exploration.

Within this patient cohort, an examination of the gender distribution further illuminates the demographic composition. Notably, 19 265 (73%) were female patients, while 7 161 (27%) were male patients.

The Upset plot serves as a gateway to comprehend not only the number of patients with each individual disease but also the instances where these diseases co-occur. Through this lens, we're poised to discern significant trends, identifying whether certain diseases tend to cluster together, potentially impacting treatment strategies and healthcare interventions. This visual exploration empowers us to better grasp the complexities of disease combinations and their implications within our patient cohort for the year.

Among these, there were 16 279 patients diagnosed with only HIV, 178 patients diagnosed with only Diabetes, and 5 296 patients diagnosed with only Hypertension. This distribution serves as an insight to comprehend not only the number of patients with each individual disease but also the instances where these diseases coexistence.

This dynamic visualization unveils the complexity of disease combinations and their prevalence within our patient cohort. Furthermore, we observe that 1 701 individuals presented with both HIV and Hypertension, 49 individuals experienced the coexistence of HIV and Diabetes, 822 presented with both Diabetes and Hypertension. Moreover, 101 individuals experienced the coexistence of all three conditions (HIV, Hypertension and Diabetes).

The subsequent bar graph casts a spotlight on the distribution of patient visits across different health facilities providing a comprehensive overview of the patient flow within each establishment.

The graph reveals that Bhubezi health facility accounted for 5 912 patient visits, making it as the most frequently visited facility overall while Lillydale health facility accounted for 1 447 patient visits making it the least visited facility overall. Meanwhile, Kildare health facility serviced 3 793 patients making it the third most visited facility, followed closely by Cunningmore health facility with 3 203 patient visits.

Moving forward with our exploration, the next bar graph offers a captivating comparative exploration of patient demographics through a pair of side-by-side graphs. In these companion visuals, we delve into the age distribution of female and male patients, each column representing the distinct age groups.

The first graph highlights the age distribution of female patients. Notably, the 6 271 females aged 30-39 emerge as a significant demographic cluster. This observation, coupled with noting a decline in the number of female patients for consecutive age groups from 40 years and above, and continues to decline by age group paints a compelling portrait of the female patient age distribution.

In the second graph, the male patient age distribution comes to the forefront. Interestingly, the highest concentration of male (2 249) is situated within the 40-49 age bracket. This contrast with the female distribution prompts us to delve further into potential reasons and implications.

Additionally, it's worth noting that 4 698 females aged 20-29 surpass their male counterparts by 3 863 individuals, revealing a potential health trend warranting closer examination. Interestingly, across all age groups, we note a consistent trend; no age group showcases a notable numerical advantage of males over females. This balanced distribution across age demographics offers an engaging point to the prevailing gender-specific health dynamics observed in other parts of the data. This could also be because there are more females than males in the cohort.

Continuing our discussion, we delve into a trio of visuals, each encapsulating the age distribution of patients diagnosed with distinct health conditions. In parallel to the approach discussed earlier, these sets of two side-by-side graphs; distinguished by gender, offer a comprehensive view of the disease dynamics across varying age groups.

The first of these graphs, examines patients diagnosed with HIV, and shows a notable finding. The age group 30-39 stands out prominently for females, with 5 694 diagnosed cases. For male counterpart, there were 1 902 cases in the age bracket 40-49 making it the highest for male patients.

Transitioning to the Hypertension-focused graph, we unearth another layer of insights. For females, 50-59 emerges as a significant age bracket, hosting 1 993 patients diagnosed with Hypertension followed closely by 1 816 patients from the 60-69 age bracket. On the male side, the corresponding age group features 642 cases in the 50-59 age range and 736 cases for age range 60-69, making it the highest age group. These contrasting figures within the same age segment highlight potential gender-influenced patterns that warrant a closer examination of lifestyle factors and healthcare interventions.

Lastly, we unravel the dynamics surrounding patients diagnosed with Diabetes. Among females, the highest number was the 60-69 cohort exhibiting 351 diagnosed cases, similar in males, the highest number was in the 60-69 group claiming 135 cases.

Advancing in our analysis, it becomes evident that a considerable segment of the patient cohort diagnosed with hypertension amounting to 71% exhibits effective hypertension control, indicative of good medical interventions. In a parallel setting, within the stratum of patients diagnosed with HIV, an encouraging proportion of 85% have suppressed viral load levels, reflecting efficacious antiretroviral treatment approaches.

Focusing on the immunological domain, the computed mean hypertension reading is 144/81mmHg and the computed median CD4 count among the HIV-diagnosed population stands at 372 cells/mm^3, signifying an important quantitative marker for immune competence. Moreover, the computed mean blood glucose level is recorded at 11.5 mmol/L, representing a critical metric in the management of diabetes.

These meticulous figures encapsulate not only vital clinical outcomes but also provide an empirical foundation for understanding disease management efficacy, treatment effectiveness, and underlying health trajectories.

For further analysis on this tab, one can use the ‘Select Year’ input to filter using the different years of interest.

4.2 Health Facility Dashboard Tab

This tab begins with a bar graph offering a comprehensive view of healthcare utilisation across all eight facilities under study. This visualization provides a broad perspective on patient engagement with the healthcare system with Agincourt health facility recording the overall highest number of patients (5 912) visiting the facility.

The next set of visualizations comprises of three frames, each dedicated to one of the three chronic conditions: HIV, Hypertension, and Diabetes. These frames illustrate the prevalence of these conditions and provide insights into the age distribution of patients.

Among our studied population, HIV is the most prevalent condition, affecting 18 554 individuals (prevalence rate 70.21%). The age distribution of HIV patients is concentrated among those aged 30-39, with 7 487 patients in this age group, followed by a decline in older age groups, with fewer patients beyond the age of 60. Notably, 85% of HIV patients achieve viral suppression. Hypertension is moderately prevalent, with 8 773 patients (prevalence rate 33.19%). The age distribution of hypertensive patients is skewed toward older age groups, particularly those aged 40 and above. Approximately 71% of hypertensive patients have their condition under control. Diabetes has a lower prevalence compared to HIV and Hypertension, with 1 346 patients (prevalence rate 5.09%). The age distribution of diabetic patients spans a wide range, with a notable concentration among individuals aged 60-69 (486 patients). A significant 87% of diabetic patients have their condition well-controlled.

4.3 Trends by Time Tab

The 'Trends by Time' tab provides a dynamic view of critical health metrics over a seven-year period, from 2015 to 2021. This section presents an analysis of the trends in the percentages of virally suppressed individuals, controlled hypertension cases, and controlled diabetes cases. The analysis encompasses both the overall population and a facility-specific breakdown for the eight healthcare facilities under study.

In the first graph, we gain insights into the temporal trends in health outcomes across the entire studied population. Notably, the percentage of virally suppressed individuals exhibits adequate growth, increasing from 83% in 2015 to 86% in 2021. This upward trend suggests a positive impact of improved HIV management strategies and increased treatment adherence.

Conversely, the percentage of controlled hypertension cases displays a modest upward trend, rising from 68% in 2015 to 72% in 2021. While the improvement is evident, further efforts may be needed to enhance hypertension control measures.

In contrast, the percentage of controlled diabetes cases demonstrates a fluctuating pattern. It starts at 85% in 2015, dips to 76% in 2017, and then recovers to 81% by 2019 only to see another dip to 71 in 2021. These fluctuations suggest ongoing challenges in diabetes management within the population.

The second graph provides a facility specific perspective, offering granular insights into health outcomes within each of the eight healthcare facilities. It is important to note that facility performance varies considerably. Some facilities consistently outperform others in achieving health goals. For example, several facilities like Cunningmore health facility consistently maintains a high rate of virally suppressed individuals, with percentages exceeding 80% throughout the seven-year period. This demonstrates effective HIV management practices at this facility.

Over the seven-year period, several facilities show marked improvement in health outcomes. Agincourt health facility, for instance, witnesses a substantial increase in the percentage of controlled hypertension cases, rising from 68% in 2015 to 75% in 2021. This suggests successful hypertension control strategies implemented at this facility. However, it is evident that not all facilities experience the same rate of progress. Xanthia health facility, while maintaining relatively stable health outcomes, does not show significant improvement in controlled hypertension or controlled diabetes cases.

While there are positive trends in virally suppression and controlled hypertension, diabetes management remains a challenge with fluctuating percentages. Facility specific disparities highlight opportunities for targeted interventions and resource allocation, ultimately contributing to improved health outcomes within the studied population. Disparities among facilities call for focused strategies to address health inequities.

4.4 Patient Tab

By utilizing the Patient tab, our objective is to gain a deep understanding of the patient's health journey. This tab serves as a dynamic and patient specific tool, offering a comprehensive view of individual health trajectories based on the patient's unique identifier (patient ID). This section provides a detailed breakdown of the patient's medical history, demographic information, and various health metrics over time.

The tab begins by presenting the patient's demographic information, including age and gender, creating a holistic profile of the individual. One critical aspect of the patient's health history is the record of diagnosed conditions. The dashboard displays a historical overview, allowing us to track any changes in health status over time. It provides insight into the evolution of health conditions from the patient's first recorded visit in 2015 to 2021, highlighting the development or remission of specific conditions.

The patient's weight trends are graphically represented, enabling us to observe fluctuations and trends in body weight. This information aids in monitoring weight-related health issues and the effectiveness of interventions. Furthermore, the dashboard showcases blood pressure data, illustrating changes and control levels over time. Monitoring blood pressure is crucial for assessing cardiovascular health.

For HIV-positive patients, viral load trends are essential. The dashboard displays viral load measurements across visits, allowing us to assess the effectiveness of antiretroviral therapy in suppressing the virus. Moreover, the patient's CD4 count, another crucial measure for HIV management, is graphed over time. A rising CD4 count indicates improved immune function.

Monitoring blood glucose levels is vital for diabetic patients. The dashboard presents blood glucose trends, helping to manage diabetes effectively and prevent complications. Lastly, a detailed record of medications prescribed at each visit is available. This information provides insights into treatment regimens and adherence.

This patient-specific approach empowers healthcare providers to make informed decisions and tailor treatment plans to each individual's unique needs. Whether it's tracking the progression of diseases, monitoring vital signs, or assessing medication effectiveness, the Patient Dashboard is a valuable tool for personalized healthcare management.

4.5 Summary

These figures shed light on the varying patient footprints across health facilities, potentially indicating factors such as geographical accessibility, specialty services, and overall patient preference. By deciphering these visit patterns, we gain valuable insights into the utilization of healthcare resources and the potential implications for optimizing healthcare delivery and resource allocation.

The strength of this work lies in the adaptability of the VAS beyond the scope of healthcare. While initially developed to analyse healthcare utilization and chronic disease progression within the context of the Agincourt HDSS-Clinic dataset, the framework used in this study can be readily applied to various domains facing data challenges. However, it's essential to acknowledge the limitations that exist in this study and the potential barriers to generalise the work.

One limitation is how specific the chosen dataset used in this research is. The Agincourt HDSS-Clinic dataset, while rich in healthcare-related information, may not fully capture the diversity of experiences and challenges present in other populations. The patterns identified within this dataset may not be entirely representative of broader populations, limiting the ability to generalise the findings. Additionally, the effectiveness of the VAS in discovering insights relies heavily on the quality of the input data.

CHAPTER 5: CONCLUSION AND RECOMMENDATIONS

5.1 Conclusion

In this research, we presented the design and development of a Visual Analytics System (VAS) utilizing data from the Agincourt Health and Demographic Surveillance System (HDSS). HDSS are pivotal in public health and research, enabling the collection of longitudinal, comprehensive data on populations. They offer insights into demographic trends, disease dynamics, and healthcare planning, serving as foundational resources for improving healthcare and understanding population dynamics within specific geographic areas. VAS, by visually representing datasets, enhance our understanding of patterns and trends. In our research, VAS has played a crucial role in analysing healthcare utilization and disease progression.

In the development of our VAS, careful consideration was given to the design and implementation to effectively handle complex data from the Agincourt HDSS-Clinic. Data visualization techniques, including interactive charts and graphs, were integrated to show the patterns and trends within the data. Moreover, user-centric design principles established during the two stakeholder meetings were applied to make the VAS accessible and intuitive for a wide range of stakeholders, from healthcare professionals to researchers. Our VAS represents not only a powerful analytical tool but also a user-friendly platform that empowers users to explore data, extract insights, and make informed decisions. The insights derived from this analysis will guide future research endeavours and healthcare planning strategies.

In recent years, visual analytics has emerged as a dynamic and evolving field at the intersection of data science and information visualization. This innovative approach harnesses the power of advanced data visualization techniques to transform complex datasets into actionable insights. As our research exemplifies, visual analytics is becoming increasingly instrumental in addressing challenges in public health, research, and beyond. The growth of this field promises exciting possibilities for the future, with the potential to transform how we explore, understand, and utilize data for informed decision making.

5.2 Recommendations

The future direction of the VAS will focus on several key areas to enhance its effectiveness and utility:

1. Data Integration and Quality Assurance: Consider the possibility of integrating additional data sources into the VAS. This expansion can provide a more comprehensive understanding of healthcare trends and patient profiles. Implementing robust data quality assurance processes is essential to ensure that the information presented remains accurate and reliable.
2. Real-Time Data Integration and Scalability: Explore deploying the VAS to a system that can handle real-time data inputs and use a faster framework. This step will assess the scalability of the tool with larger datasets and its compatibility with various data sources. Real-time data integration can provide more up-to-date insights and support timely decision-making.
3. User Feedback and Continuous Improvement: Actively seek feedback from stakeholders who use the VAS. Their input is invaluable for improving user experience, enhancing data visualization, and fine-tuning functionality. A user-centric approach ensures that the VAS remains user-friendly and aligned with user needs.
4. Predictive Analytics for Healthcare: Consider implementing predictive analytics within the VAS to forecast disease progression and healthcare resource utilization. Predictive models can offer early insights, allowing for proactive interventions and optimized resource allocation. This addition can further empower healthcare professionals and researchers.

These future directions represent a strategic approach to evolving the VAS. By expanding data sources, ensuring data quality, enabling real-time capabilities, and incorporating predictive analytics, the VAS can become an even more powerful tool for healthcare management and research. Additionally, ongoing user feedback ensures that the VAS continues to meet the evolving needs of its users, ultimately contributing to improved healthcare outcomes.

REFERENCES

1. Centers for Disease Control and Prevention. National Center for Chronic Disease Prevention and Health Promotion (NCCDPHP). 2019 [cited 2022 Feb 21]. About Chronic Diseases | CDC. Available from: https://www.cdc.gov/chronicdisease/about/index.htm

2. Siddiqi K. Public Health: An action guide to improving health. 2010 [cited 2021 Jul 21]. Non-communicable diseases. Available from: https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases

3. Nyirenda MJ. Non-communicable diseases in sub-Saharan Africa: Understanding the drivers of the epidemic to inform intervention strategies. Int Health. 2016 May 1;8(3):157–8.

4. Holmes MD, Dalal S, Volmink J, Adebamowo CA, Njelekela M, Fawzi WW, et al. Non-communicable diseases in sub-Saharan Africa: The case for cohort studies. PLoS Med. 2010 May;7(5).

5. Stats SA. Mortality and causes of death in South Africa, 2016: Findings from death notification. Statistical release. 2016;(November 2016):51.

6. Peer N, Steyn K, Lombard C, Lambert E V., Vythilingum B, Levitt NS. Rising Diabetes Prevalence among Urban-Dwelling Black South Africans. PLoS One. 2012 Sep 4;7(9).

7. Lloyd-sherlock P, Ebrahim S, Grosskurth H. Is hypertension the new HIV epidemic? Int J Epidemiol [Internet]. 2014 Feb [cited 2022 Feb 21];43(1):8–10. Available from: /pmc/articles/PMC3937981/

8. Puoane TR, Tsolekile LP, Caldbick S, Igumbor EU, Meghnath K, Sanders D. Chronic non-communicable diseases in South Africa: Progress and challenges. South African Health Review 2012/2013. 2013;115–26.

9. Norman R, Bradshaw D, Schneider M, Joubert J, Groenewald P, Lewin S, et al. A comparative risk assessment for south Africa in 2000. South African Medical Journal. 2007;97(8):637–41.

10. Preventing Chronic Diseases: A Vital Investment - WHO, World Health Organization - Google Books [Internet]. [cited 2022 Jun 4]. Available from: https://books.google.cm/books?hl=en&lr=&id=dF00DgAAQBAJ&oi=fnd&pg=PR3&dq=WHO:+Preventing+Chronic+Disease.+A+vital+investment+Geneva:+WHO%3B+2005&ots=8XW7GxzoSx&sig=FgaVC8LwNU90IeiIk0IHf1XCy4o&redir_esc=y#v=onepage&q=WHO%3A Preventing Chronic Disease. A vi

11. Multimorbidity in South Africa: A case of one in five people having more than one chronic illness | South African Medical Research Council [Internet]. [cited 2022 Aug 31]. Available from: https://www.samrc.ac.za/news/multimorbidity-south-africa-case-one-five-people-having-more-one-chronic-illness

12. Nguyen H, Manolova G, Daskalopoulou C, Vitoratou S, Prince M, Prina AM. Prevalence of multimorbidity in community settings: A systematic review and meta-analysis of observational studies. J Comorb. 2019 Aug 22;9:2235042X19870934.

13. Nunes BP, Flores TR, Mielke GI, Thumé E, Facchini LA. Multimorbidity and mortality in older adults: A systematic review and meta-analysis. Arch Gerontol Geriatr. 2016 Nov 1;67:130–8.

14. Wei MY, Mukamal KJ. Multimorbidity, Mortality, and Long-Term Physical Functioning in 3 Prospective Cohorts of Community-Dwelling Adults. Am J Epidemiol. 2018 Jan 1;187(1):103–12.

15. Kanesarajah J, Waller M, Whitty JA, Mishra GD. Multimorbidity and quality of life at mid-life: A systematic review of general population studies. Maturitas. 2018 Mar 1;109:53–62.

16. Makovski TT, Schmitz S, Zeegers MP, Stranges S, van den Akker M. Multimorbidity and quality of life: Systematic literature review and meta-analysis. Ageing Res Rev. 2019 Aug 1;53:100903.

17. Lee JT, Hamid F, Pati S, Atun R, Millett C. Impact of Noncommunicable Disease Multimorbidity on Healthcare Utilisation and Out-Of-Pocket Expenditures in Middle-Income Countries: Cross Sectional Analysis. PLoS One. 2015 Jul 8;10(7):e0127199.

18. Mash B, Fairall L, Adejayan O, Ikpefan O, Kumari J, Matheel S, et al. A morbidity survey of South African primary care. PLoS One. 2012;7(3).

19. Spires M, Sanders D, Puoane T, Hoelzel P, Swart R. South African Health Review. 2016 [cited 2022 Aug 31]. p. 35–42 Diet-related non-communicable diseases in South Africa: determinants and policy responses. Available from: https://journals.co.za/docserver/fulltext/healthr/2016/1/healthr_2016_a6.pdf?expires=1567772270&id=id&accname=guest&checksum=479DB0A52653CF204B2C27C570E0A35E

20. Murphy GAV, Asiki G, Ekoru K, Nsubuga RN, Nakiyingi-Miiro J, Young EH, et al. Sociodemographic distribution of non-communicable disease risk factors in rural Uganda: a cross-sectional study. Int J Epidemiol. 2013;42(6):1740.

21. Roomaney RA, van Wyk B, Turawa EB, Pillay-van Wyk V. Multimorbidity in South Africa: a systematic review of prevalence studies. BMJ Open. 2021 Oct 6;11(10):e048676.

22. Chang AY, Gómez-Olivé FX, Manne-Goehler J, Wade AN, Tollman S, Gaziano TA, et al. Multimorbidity and care for hypertension, diabetes and HIV among older adults in rural South Africa. Bull World Health Organ. 2019 Jan 1;97(1):10.

23. Nojilana B, Bradshaw D, Pillay-van Wyk V, Msemburi W, Laubscher R, Somdyala NIM, et al. Emerging trends in non-communicable disease mortality

in South Africa, 1997 - 2010. South African Medical Journal. 2016 Apr 1;106(5):477–84.

24. Roomaney RA, van Wyk B, Cois A, Pillay-van Wyk V. One in five South Africans are multimorbid: An analysis of the 2016 demographic and health survey. PLoS One [Internet]. 2022 May 26 [cited 2022 Aug 31];17(5):e0269081. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0269081

25. What is diabetes? | CDC [Internet]. [cited 2022 Jun 12]. Available from: https://www.cdc.gov/diabetes/basics/diabetes.html

26. Werfalli M, Kassanjee R, Kalula S, Kowal P, Phaswana-Mafuya N, Levitt NS. Diabetes in South African older adults: prevalence and impact on quality of life and functional disability–as assessed using SAGE Wave 1 data. Glob Health Action. 2018 Jan 1;11(1).

27. Diabetes [Internet]. [cited 2022 Jun 12]. Available from: https://www.who.int/health-topics/diabetes#tab=tab_1

28. Mathew TK, Zubair M, Tadi P. Blood Glucose Monitoring. Medical Devices and Systems [Internet]. 2023 Apr 23 [cited 2023 Aug 15];66-1-66–10. Available from: https://www.ncbi.nlm.nih.gov/books/NBK555976/

29. World Health Organization (WHO). Who. 2018 [cited 2022 Jul 9]. p. 1 HIV/AIDS | WHO | Regional Office for Africa. Available from: https://www.afro.who.int/health-topics/hivaids

30. Lippman SA, El Ayadi AM, Grignon JS, Puren A, Liegler T, Venter WDF, et al. Improvements in the South African HIV care cascade: findings on 90-90-90 targets from successive population-representative surveys in North West Province. J Int AIDS Soc [Internet]. 2019 Jun 1 [cited 2023 Aug 15];22(6). Available from: /pmc/articles/PMC6562149/

31. HIV/AIDS [Internet]. [cited 2022 Jul 9]. Available from: https://www.who.int/news-room/fact-sheets/detail/hiv-aids

32. Akullian A, Vandormael A, Miller JC, Bershteyn A, Wenger E, Cuadros D, et al. Large age shifts in HIV-1 incidence patterns in KwaZulu-Natal, South Africa. Proc Natl Acad Sci U S A. 2021 Jul 13;118(28).

33. Simelela N, Venter WDF, Pillay Y, Barron P. A Political and Social History of HIV in South Africa. Curr HIV/AIDS Rep. 2015 Jun 1;12(2):256–61.

34. World health Organization. Guideline for the pharmacological treatment of hypertension in adults. Vol. 103, American Family Physician. 2021.

35. Hypertension [Internet]. [cited 2022 Jun 12]. Available from: https://www.who.int/news-room/fact-sheets/detail/hypertension

36. Poulter NR, Prabhakaran D, Caulfield M. Hypertension. The Lancet. 2015 Aug 22;386(9995):801–12.

37. Sharma JR, Mabhida SE, Myers B, Apalata T, Nicol E, Benjeddou M, et al. Prevalence of Hypertension and Its Associated Risk Factors in a Rural Black Population of Mthatha Town, South Africa. International Journal of Environmental Research and Public Health 2021, Vol 18, Page 1215. 2021 Jan 29;18(3):1215.

38. Guwatudde D, Nankya-Mutyoba J, Kalyesubula R, Laurence C, Adebamowo C, Ajayi IO, et al. The burden of hypertension in sub-Saharan Africa: A four-country cross sectional study. BMC Public Health. 2015;15(1):1–8.

39. Gómez-Olivé FX, Ali SA, Made F, Kyobutungi C, Nonterah E, Micklesfield L, et al. Regional and Sex Differences in the Prevalence and Awareness of Hypertension: An H3Africa AWI-Gen Study Across 6 Sites in Sub-Saharan Africa. Glob Heart. 2017;12(2):81–90.

40. van Zyl S, van Rooyen FC, Joubert G, Kruger WH, Walsh CM. A Comparison of the Socio-Behavioral-Metabolic Risk Profiles and Associated Factors for Chronic Diseases of Lifestyle in Urban and Rural Communities in Central South Africa. Vol. 8, Frontiers in Public Health. 2020.

41. Chou CY, Chiu CJ, Chang CM, Wu CH, Lu FH, Wu JS, et al. Disease-related disability burden: a comparison of seven chronic conditions in middle-aged and older adults. BMC Geriatr. 2021 Dec 1;21(1).

42. Campbell DJT, Sargious P, Lewanczuk R, McBrien K, Tonelli M, Hemmelgarn B, et al. Use of chronic disease management programs for diabetes: In Alberta's primary care networks. Canadian Family Physician. 2013 Feb;59(2):e86.

43. Nolte S, Osborne RH. A systematic review of outcomes of chronic disease self-management interventions. Quality of Life Research. 2013 Oct 31;22(7):1805–16.

44. Goodridge D, Bandara T, Marciniuk D, Hutchinson S, Crossman L, Kachur B, et al. Promoting chronic disease management in persons with complex social needs: A qualitative descriptive study. Chron Respir Dis. 2019 Mar 4;16.

45. The importance of managing chronic conditions - Discovery [Internet]. [cited 2022 Jul 19]. Available from: https://www.discovery.co.za/vitality/importance-of-managing-chronic-conditions

46. Maimela E, Alberts M, Bastiaens H, Fraeyman J, Meulemans H, Wens J, et al. Interventions for improving management of chronic non-communicable diseases in Dikgale, a rural area in Limpopo Province, South Africa. BMC Health Serv Res. 2018 May 4;18(1).

47. Nuño R, Coleman K, Bengoa R, Sauto R. Integrated care for chronic conditions: The contribution of the ICCC Framework. Health Policy (New York). 2012;105(1):55–64.

48. Siantz E, Aranda MP. Chronic disease self-management interventions for adults with serious mental illness: A systematic review of the literature. Gen Hosp Psychiatry. 2014;36(3):233–44.

49. Reynolds R, Dennis S, Hasan I, Slewa J, Chen W, Tian D, et al. A systematic review of chronic disease management interventions in primary care. BMC Fam Pract. 2018 Jan 9;19(1):1–13.

50. Rao S, Brammer C, McKethan A, Buntin MB. Health Information Technology. Transforming Chronic Disease Management and Care Transitions. Primary Care - Clinics in Office Practice. 2012;39(2):327–44.

51. Alotaibi YK, Federico F. The impact of health information technology on patient safety. Vol. 38, Saudi Medical Journal. Saudi Medical Journal; 2017. p. 1173–80.

52. Monawe M, Chawani MG, Kapokosa G, Moyo C. Strengthening Health Management Information Systems in Malawi: Gaps and opportunities. 2015 IST-Africa Conference, IST-Africa 2015. 2015 Aug 11;

53. Doupi P. Using EHR data for monitoring and promoting patient safety: Reviewing the evidence on trigger tools. Stud Health Technol Inform. 2012;180:786–90.

54. Hoover R. Benefits of using an electronic health record. Nurs Crit Care (Ambler). 2017 Jan 1;12(1):9–10.

55. Graber ML, Johnston D, Bailey R. Report of the Evidence on Health IT Safety and Interventions Final Report. 2016;

56. Rostamzadeh N, Abdullah SS, Sedig K. Data-Driven Activities Involving Electronic Health Records: An Activity and Task Analysis Framework for Interactive Visualization Tools. Multimodal Technologies and Interaction 2020, Vol 4, Page 7. 2020 Mar 3;4(1):7.

57. Katurura M, Cilliers L. The extent to which the POPI act makes provision for patient privacy in mobile personal health record systems. 2016 IST-Africa Conference, IST-Africa 2016. 2016 Aug 2;

58. Heisey-Grove D, Danehy LN, Consolazio M, Lynch K, Mostashari F. A national study of challenges to electronic health record adoption and meaningful use. Med Care. 2014 Feb;52(2):144–8.

59. Caban JJ, Gotz D. Visual analytics in healthcare--opportunities and research challenges. J Am Med Inform Assoc. 2015 Mar 1;22(2):260–2.

60. Katurura MC, Cilliers L. Electronic health record system in the public health care sector of South Africa: A systematic literature review. Afr J Prim Health Care Fam Med. 2018 Nov 20;10(1).

61. Els F, Cilliers L. Improving the information security of personal electronic health records to protect a patient’s health information. 2017 Conference on

Information Communication Technology and Society, ICTAS 2017 - Proceedings. 2017 May 5;

62. Tsegaye T, Flowerday S. A System Architecture For Ensuring Interoperability In A South African National Electronic Health Record System. South African Computer Journal. 2021 Jul 1;33(1):79–110.

63. The Functional Art: An introduction to information graphics and visualization - Alberto Cairo - Google Books [Internet]. [cited 2022 Feb 24]. Available from: https://books.google.com.pk/books?hl=en&lr=&id=xwjhh6Wu-VUC&oi=fnd&pg=PT16&dq=Such+introductions+particularly+are+pretty+much+often+literally+going+with+by+a+web+page+or+slide+show&ots=nNvIlqJL5a&sig=MdfpryqFpw8rji9MXa7ZHIUpSRs&redir_esc=y#v=onepage&q&f=f

64. Lor M, Koleck TA, Bakken S. Information visualizations of symptom information for patients and providers: A systematic review. Vol. 26, Journal of the American Medical Informatics Association. Oxford University Press; 2019. p. 162–71.

65. Kirk A. Data Visualization : a successful design process. 2012. 181 p.

66. De Folter J, Gokalp H, Fursse J, Sharma U, Clarke M. Designing effective visualizations of habits data to aid clinical decision making. BMC Med Inform Decis Mak. 2014;14(1).

67. Backonja U, Haynes SC, Kim KK. Data visualizations to support health practitioners' provision of personalized care for patients with cancer and multiple chronic conditions: User-centered design study. JMIR Hum Factors. 2018 Oct 1;5(4).

68. West VL, Borland D, Hammond WE. Innovative information visualization of electronic health record data: A systematic review. Journal of the American Medical Informatics Association. 2015;22(2):330–9.

69. Saket B, Endert A, Demiralp C. Task-Based Effectiveness of Basic Visualizations. IEEE Trans Vis Comput Graph [Internet]. 2019 Jul 1 [cited 2024 May 16];25(7):2505–12. Available from: https://pubmed.ncbi.nlm.nih.gov/29994001/

70. Shneiderman B, Plaisant C, Hesse BW. Improving healthcare with interactive visualization. Computer (Long Beach Calif). 2013;46(5):58–66.

71. Arcia A, Suero-Tejeda N, Bales ME, Merrill JA, Yoon S, Woollen J, et al. Sometimes more is more: iterative participatory design of infographics for engagement of community members with varying levels of health literacy. J Am Med Inform Assoc. 2016 Jan 1;23(1):174.

72. Li Q. Overview of Data Visualization. Embodying Data. Nature Publishing Group; 2020. 17–47 p.

73. Yarmuluk D. Discover the 4 key traits of a great data visualization [Internet]. 2019 [cited 2022 Jul 18]. Available from: https://www.microsoft.com/en-

us/microsoft-365/business-insights-ideas/resources/discover-the-4-key-traits-of-a-great-data-visualization

74. Wong PC, Thomas J. Visual analytics. Vol. 24, IEEE Computer Graphics and Applications. 2004. p. 20–1.

75. Zhao Y, Parvinzamir F, Wei H, Liu E, Deng Z, Dong F, et al. Visual analytics for health monitoring and risk management in CARRE. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2016;9654:380–91.

76. Chung Y, Bagheri N, Salinas-Perez JA, Smurthwaite K, Walsh E, Furst MA, et al. Role of visual analytics in supporting mental healthcare systems research and policy: A systematic scoping review. Vol. 50, International Journal of Information Management. Pergamon; 2020. p. 17–27.

77. Munzner T. Visualization Analysis and Design. Visualization Analysis and Design. A K Peters/CRC Press; 2014.

78. Rostamzadeh N, Abdullah SS, Sedig K, Garg AX, McArthur E. Visual Analytics for Predicting Disease Outcomes Using Laboratory Test Results. Informatics 2022, Vol 9, Page 17. 2022 Feb 25;9(1):17.

79. Angulo DA, Schneider C, Oliver JH, Charpak N, Hernandez JT. A multi-facetted visual analytics tool for exploratory analysis of human brain and function datasets. Front Neuroinform. 2016 Aug 23;10(AUG).

80. Sedig K, Parsons P, Babanski A. Towards a Characterization of Interactivity in Visual Analytics. Journal of Multimedia Processing and Technologies. 2012;3.

81. What Is Tableau? | Tableau [Internet]. [cited 2022 Jul 18]. Available from: https://www.tableau.com/why-tableau/what-is-tableau

82. Qlik. Qlik | Analytics & Data Integration Platform.

83. SAS Institute Inc. SAS Institute Inc. 2013 [cited 2022 Jul 18]. SAS: Analytics, Artificial Intelligence and Data Management | SAS. Available from: https://www.sas.com/en_us/home.html

84. R: What is R? [Internet]. [cited 2022 Jul 18]. Available from: https://www.r-project.org/about.html

85. Shiny [Internet]. [cited 2022 Jul 18]. Available from: https://shiny.rstudio.com/

86. Simpao AF, Ahumada LM, Desai BR, Bonafide CP, Gávez JA, Rehman MA, et al. Optimization of drug-drug interaction alert rules in a pediatric hospital's electronic health record system using a visual analytics dashboard. Journal of the American Medical Informatics Association. 2015 Mar 1;22(2):361–9.

87. Abdullah S. Visual Analytics of Electronic Health Records with a focus on Acute Kidney Injury. Electronic Thesis and Dissertation Repository. 2020 Jul 15;

88. Nouri MS, Lizotte DJ, Sedig K, Abdullah SS. Visemure: A visual analytics system for making sense of multimorbidity using electronic medical record data. Data (Basel). 2021;6(8).

89. Quinn M, Forman J, Harrod M, Winter S, Fowler KE, Krein SL, et al. Electronic health records, communication, and data sharing: Challenges and opportunities for improving the diagnostic process. Diagnosis. 2019 Aug 27;6(3):241–8.

90. Tseng YJ, Chiu HJ, Chen CJ. Dxpr: An R Package for Generating Analysis-ready Data from Electronic Health Records—Diagnoses and Procedures. PeerJ Comput Sci. 2021;7:1–20.

91. Malik S. Enterprise Dashboards - Design and Best Practices for IT, John Wiley & Sons, Inc. 2005;

92. Faisal S, Blandford A, Potts HW. Making sense of personal health information: Challenges for information visualization. Health Informatics J. 2013 Sep;19(3):198–217.

93. Kahn K, Collinson MA, Xavier Gómez-olivé F, Mokoena O, Twine R, Mee P, et al. Profile: Agincourt health and socio-demographic surveillance system. Int J Epidemiol. 2012 Aug;41(4):988–1001.

94. MRC/Wits Agincourt Unit – Health & Socio-demographic Surveillance [Internet]. [cited 2022 Aug 25]. Available from: https://www.agincourt.co.za/

95. Kabudula CW, Houle B, Collinson MA, Kahn K, Gómez-Olivé FX, Tollman S, et al. Socioeconomic differences in mortality in the antiretroviral therapy era in Agincourt, rural South Africa, 2001–13: a population surveillance analysis. Lancet Glob Health [Internet]. 2017 Sep 1 [cited 2023 Oct 4];5(9):e924–35. Available from: http://www.thelancet.com/article/S2214109X17302978/fulltext

96. Kahn K, Tollman SM, Collinson MA, Clark SJ, Twine R, Clark BD, et al. Research into health, population and social transitions in rural South Africa: Data and methods of the Agincourt health and demographic surveillance system. Scand J Public Health [Internet]. 2007 Nov 25 [cited 2021 Jul 13];35(SUPPL. 69):8–20. Available from: https://journals.sagepub.com/doi/10.1080/14034950701505031

97. Serwaa-Bonsu A, Herbst AbrahamJ, Reniers G, Ijaa W, Clark B, Kabudula C, et al. First experiences in the implementation of biometric technology to link data from Health and Demographic Surveillance Systems with health facility data. Glob Health Action. 2010 Dec;3(1):2120.

98. Kabudula CW, Clark BD, Gómez-Olivé FX, Tollman S, Menken J, Reniers G. The promise of record linkage for assessing the uptake of health services in resource constrained settings: A pilot study from South Africa. BMC Med Res Methodol. 2014 May 24;14(1).

APPENDICES

Appendix A: Plagiarism Policy

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

FACULTY OF HEALTH SCIENCES

PLAGIARISM DECLARATION TO BE SIGNED BY ALL HIGHER DEGREE STUDENTS

SENATE PLAGIARISM POLICY: APPENDIX ONE

I Mapule Dorcus Nhlapho (Student number: 1385721) am a student registered for the degree of MSc in Epidemiology (Public Health Informatics) in the academic year 2.

I hereby declare the following:

- I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.
- I confirm that the work submitted for assessment for the above degree is my own unaided work except where I have explicitly indicated otherwise.
- I have followed the required conventions in referencing the thoughts and ideas of others.
- I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.
- I have included as an appendix a report from "Turnitin" (or other approved plagiarism detection) software indicating the level of plagiarism in my research document.

Signature: ____________ Date. 11/01/2024

Appendix B: Ethics Clearance Certificate

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

R49 Ms M Nhlapho

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)
CLEARANCE CERTIFICATE NO. M220668

NAME: (Principal Investigator) Ms M Nhlapho

DEPARTMENT: School of Public Health
Rural Public Health & Health Transitions Research Unit
Medical School
University

PROJECT TITLE: *A visual analytics approach to characterising disease progression among adults with chronic diseases in rural northeast South Africa*

DATE CONSIDERED: 2022/06/24

DECISION: Approved unconditionally

CONDITIONS:

NOTE: If contact information regarding student study participants is required, please contact the Registrar's office - <Nicoleen.Potgieter@wits.ac.za>

SUPERVISOR:

APPROVED BY: Dr CB Penny, ~~Chairperson~~, HREC (Medical)

DATE OF APPROVAL: 2022/08/15

This Clearance Certificate is valid for 5 years from the date of approval. An extension may be applied for.

DECLARATION OF INVESTIGATORS

To be completed in duplicate and **ONE COPY** returned to the Research Office secretariat on the 3rd floor, Phillip Tobias Building, Parktown, University of the Witwatersrand, Johannesburg.

I/we fully understand the conditions under which I am/we are authorized to carry out the above-mentioned research and I/we undertake to ensure compliance with these conditions. Should any departure be contemplated from the research protocol as approved, I/we undertake to submit details to the Committee. **I agree to submit a yearly progress report.** When a funder requires annual re-certification, the application date will be one year after the date when the study was initially reviewed. In this case, the study was initially reviewed in June and therefore reports and re-certification will be due in the month of June each year. Unreported changes to the study may invalidate the clearance given by the HREC (Medical).

Signature of Principal Investigator ____________ Date ____________