RED CELL DISTRIBUTION WIDTH AS A SURROGATE MARKER OF HAEMOGLOBINOPATHIES AMONG PATIENTS IN AGA KHAN HOSPITAL, WESTERN KENYA

Benard Munguti Mutua

A Thesis Submitted in Partial Fulfilment of the Requirements for the Degree of Master of Medical Laboratory Sciences (Hematology and Transfusion Science) of Masinde Muliro University of Science and Technology

October, 2023

DECLARATION

This thesis is my original work prepared with no other than the indicated sources of support and has not been presented elsewhere for a degree or any other award.

Signature…… Date……

Benard Munguti Mutua, BSc MLS

HML/G/02/2016

CERTIFICATION

The undersigned certify that they have read and hereby recommend for acceptance of Masinde Muliro University of Science and Technology a thesis entitled "**Red Cell Distribution Width as a Surrogate Marker of Haemoglobinopathies among Patients in Aga Khan Hospital, Western Kenya".**

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

Dr. Okoth Patrick, (PhD)

Department of Biological Sciences

Masinde Muliro University of Science and Technology

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

Mr. George Sowayi (MPhil)

Department of Medical Laboratory Sciences

Masinde Muliro University of Science and Technology

COPYRIGHT

DEDICATION

This thesis is dedicated to my only beautiful wife, Violah Jebet Benards' for her steadfast support towards achieving this great work and to my mother Dorcas Nthenya and the late Francis Mutua Munguti for your memory lifts my gaze and fills my heart.

ACKNOWLEDGMENTS

I would like to express my special appreciation and thanks to my academic advisors Dr. Patrick Okoth and Dr George Sowayi for being interested, honest and enthusiastic mentors towards my research work and for allowing me to grow as a research scientist. I pass much gratitude to my beloved brothers; Pastor Nicholas Kithama, Samuel Muteti and my only Sister Esther Mwende for good will support and unrelenting prayers. Words cannot express how grateful I am to my uncle Silas Mutuku, Martha Silas, and Aunt Juliana Mbaika for all the sacrifices they made on my behalf and encouraging me throughout my academic life. A special thanks to my beloved children Patience Nthenya, Glorious Utheu and Honest Mui for being such kids always believing and cheering me up. Special thanks to Raphael Kitonga for allowing me to use his homebased library together with all its accessories that nurtured me as a scientist. Much appreciation to the Aga Khan Hospital, Kisumu for allowing me to carry out the study in their modern health facility. Finally, I thank God for His special grace that has consistently and persistently guided me through challenging situations and helping me this far (1 Samuel 7:12).

ABSTRACT

Red cell distribution width (RDW) measures the morphological diversity of red blood cells (rbcs) per volume rbcs. Such diversity, characterized by significant reticulocytosis and rbcs variation in size(anisocytosis) and shape(poikilocytosis) is common in heightened erythropoiesis, typical of marked haemolysis. Haemoglobinopathies are hereditary haemoglobin disorders characterized by haemolytic anaemia and are highly prevalent among children in malaria-endemic Western Kenya's Lake Victoria basin counties. As in other low-income countries, timely haemoglobinopathy detection in Kenya is hampered by lack of affordable methods, hindering timely detection of many hemoglobinopathy cases. The RDW is cheaper as it is easily measured as a haematological index in complete blood count (CBC), routinely generated by modern haematology autoanalysers. Despite studies showing its potential as a haemoglobinopathy biomarker, RDW has not been rigorously evaluated, including in Kenya. This study aimed to determine the utility of RDW as a biomarker for discriminating individuals with and without haemoglobinopathy in Western Kenya. This was a cross-sectional hospital-based retrospective comparative study of 488 study participants aged 1 month to 66 years, with respective RDW values known and confirmed haemoglobinopathy status, diagnosed at Aga Khan Hospital, Kisumu, and its satellite centers in Western Kenya from January 2015 to December 2020. Sample size was calculated using Cochran's' formula. Data were obtained from the hospital laboratory database and analyzed using SPSS version 23. Haemoglobinopathy profile was summarised as frequencies and percentages and presented in bar-graphs and pie-charts. Demographics variations across groups was analyzed using Chi square. Values for RDW for the haemoglobinopathy-free (control) group were not normally distributed and so the median and interquartile range (IQR) were used as descriptive statistics and in deriving the reference cut-off value for haemoglobinopathy detection. The cut-off value was computed as the upper limit of the 95% confidence interval (CI) and values greater than this indicated a diagnosis of haemoglobinopathy. Kruskal-Wallis tests was used to determine RDW variation among population groups; Dunn's *post hoc* test within groups and Mann Whitney U between two groups. Majority of the patients were from the Kisumu station (49.0%, n=239) followed by Busia (15.4%, n=75), Homabay (12.3%, n=60) while the rest were distributed in small numbers among the rest of the stations ($p<0.0001$**).** Most cases had HbAS, 41.7% (n=103), followed by HbSS β-Thal, 25.1% (n=62); HbSS, 18.2% (n=45); HbSS+HbF, 8.1% (n=20); Hbβ-thal, 3.6% (n=9); HbAS+β-thal, 2.4% (n=6); and HbAS+HbF, 0.8% (n=2). Overall median RDW for control group was 14.5, IQR=2.7;95%CI=9.1-19.9, giving reference cut-off limit of 19.9 while for overall haemoglobinopathies combined was 20.7(IQR=8.3), $p<0.0001$**.** It did not differ significantly between males (14.55[IQR=1.90]) and females (14.2[IQR=2.40]), p=0.089. It varied significantly with age ($p<0.05$) being highest for <5year-olds, 19.0 (IQR=7.1), followed by 6-12year-olds, 15.5 (IQR=8.2) and >12-year-olds, 14.1 (IQR=2.8), respectively. Haemoglobinopathies whose RDW was higher than the cut-off value of 19.9 were: HbSS, HbSS+β-thal, HbSS+HbF, HbAS+HbF, HbAS+β-thal but not simple HbAS and beta thalassemia. For combined haemoglobinopathies, RDW cut-off limit showed, sensitivity, specificity, and *Youden Index* of 55.1%, 94.2%, and 74.6%; positive and negative predictive validity, 90.7% and 67.2%; positive and negative likelihood ratio, 9.5 and 0.476; and odds ratio,19.86 (10.9-36.2 CI). The RDW showed promise as a hemoglobinopathy diagnostic, rather than screening biomarker, with possibility of improved overall performance with cut-off limits reference group drawn from the healthy population thus need evaluation using prospective data.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND ACRONYMS

BTM- Beta Thalassemia Major

BTT- Beta Thalassemia Trait

CBC- Complete Blood Count

EDTA- Ethylenediaminetetraacetic Acid

EQA- External Quality Assurance

HB- Hemoglobin

HBA A- Hemoglobin A A

HB A S- Hemoglobin A S

HB S S – Hemoglobin S S

HB S C- Hemoglobin S C

HCT- Hematocrit

HPLC- High-Performance Liquid Chromatography

IDA- Iron Deficiency Anemia

IQA- Internal Quality Assurance

K-**S** Kolmogorov-Smirnov

MMUST- Masinde Muliro University of Science and Technology

MCI- Matos & Calvalho Index

MCHC- Mean Corpuscular Hemoglobin Concentration

MCH- Mean Corpuscular Hemoglobin

MCV- Mean Corpuscular Volume

MPV- Mean Platelet Volume

CV- Packed Cell Volume

PBS- Peripheral Blood Spear

PS- Peripheral Smear

PDW- Platelets Distribution Width

P-LCR- Platelets Large Cell Ratio

RBC- Red Blood Cell

RDW- Red Cell Distribution Width

SD- Standard Deviation

SCD- Sickle Cell Disease

SCT- Sickle Cell Trait

TT- Thalassemia Trait

UKNEQAS- United Kingdom National External Quality Assessment

UN- United Nations

WBC- White Blood Cell

WHO- World Health Organization

OPERATIONALIZATION OF TERMS

Anisocytosis. Size variation of red blood cells on a blood smear.

Biomarkers. These are biological molecules whose presence indicate presence of a physiological or pathological or disease

Haemoglobinopathy. A group of inherited genetic diseases that result into abnormal protein structure of hemoglobin or reduced production of the Hb protein molecule (thalassemia).

Hydrops fetalis. A life-threatening disorder that results from complete deletion of alpha genes leading into production of gamma tetramer which is incompatible with life.

Poikilocytosis. Shape variation of red blood cells on a blood smear.

Red cell distribution width. A coefficient of variation that is produced as part of complete blood count output that measures red blood cell variation of sizes and shapes.

ROC curve or receiver operating characteristic curve. A graphical representation of a continuous data that measures the optimal potential of a biomarker.

Sensitivity. The ability to identify a disease correctly as a positive test.

Specifity. The ability to correctly identify the absence of a disease as negative test.

Youden Index (*J*). The maximum efficacy of a continuous data being used as a biomarker. It is measured as the Area Under the Roc Curve. It is also the accuracy of a biomarker.

CHAPTER ONE

INTRODUCTION

1.1 Background to the Study

Hemoglobinopathies encompass all inherited genetic diseases of hemoglobin (Weatherall, 2011). Phenotypically they fall into two main groups. One group are the thalassemias, which are among the common recessively inherited autosomal blood disorders. These are characterized by production of inadequate quantities of Hb protein molecules, accompanied with microcytic erythrocytes prone to haemolysis causing anaemia (Karimi *et al.,* 2019, Taher *et al.,* 2021).

The main types of thalassemia are alpha (α)- and beta (β)- thalassemia. The other group are the non-thalassaemia haemoglobinopathies, characterized by the production of structurally and functionally abnormal hemoglobin protein molecules (Vinjamur *et al.,* 2018). The main non-thalassaemia structural hemoglobin variants are HbS, HbE and HbC. They result from deletion mutation of genes encoding the synthesis of specific amino acids comprising the polypeptide chains of the globin protein part of the Hb molecule (Vinjamur *et al.,* 2018). In the homozygous state of these haemoglobinopathies, the red cells containing structurally defective haemoglobin molecules are incompetent in oxygen delivery, and mostly undergo distortion of their shapes in the resultant hypoxia, marking them for hemolysis by the victim's reticuloendothelial system. The hemolysis leads to fatal or life-threatening severe anaemia and hypoxia (Ilesanmi, 2010). As illustrated by several studies done in India and Western Kenya region, genetic inheritance in the vast majority of hemoglobinopathies is heterozygous, manifesting phenotypically as traits

(haemoglobinopathy trait carriers), such as the sickle cell trait (HbAS), and ß-thalassaemia minor and α-thalassaemia minor (Sawaimul *et al.,* 2018, Byrd *et al* 2019, Kifude 2007, Kosiyo et al 2021 and Suchdev *et al* 2014).

A study done by Sawaimul *et al.,* 2018 further, indicates that some victims of haemoglobinopathy carry the semi-normal sickle cell Hb trait (SCT) (HbAS) and are asymptomatic trait (carriers). They thus appear normal and do not know their carrier status unless specifically screened. The authors indicate that if two carriers of the trait come together as couple their union carries 1 in 4 chance that with each pregnancy that their child will inherit the genotype for the clinically manifest HbSS hence sickle cell disease (SCD). This study emphasizes to tackle the problem at regional level due to genetic, ethnic, and regional diversity of the haemoglobin variants gene polymorphism. Owing to its serious medical and socioeconomic impact of haemoglobinopathy, Sawaimul *et al.* (2018), recommends implementation of screening programmes for sickle cell disease gene carriers as an effective approach to reduce incidences and consequently its burden to the society. In another study done to determine the burden of sickle cell trait in Uganda by Ndeezi *et al.,* 2016, documented a prevalence of 13.3% SCT and 0.7% SCD among infants across 112 districts noting variation of this disorder between regions explaining the role of geographical, genetics and ethnicity which could help inform national strategies in combating sickle cell disease.

Hemoglobinopathies are now an increasing and neglected global health problem (Weatherall, 2011; Chakravorty *et al.,* 2015) that needs governments and international countries immediate attention. The world health organization has shown a growing public health problem in 71 % of 229 countries where over 330,000 infants are born

each year with these disorders (Modell & Darlison 2008). The literature reviewed indicates that about 7% of the human population carry a defective Hb gene and that over 300,000 births annually have severe hemoglobin disorders with over 200,000 being born in sub-Sahara Africa alone (Weatherall *et al.,* 2011, Piety *et al.,* 2016). Similarly, it is documented that; haemoglobin disorders are the most common genetic defects in human population with over 269 million carriers majority being found in Southeast Asia with over 90 million carriers followed by 85 million from Sub-Saharan African countries and about 48 million in the West Pacific region (Tent 2006). Studies have therefore, reported 25% risk of passing a major haemoglobinopathy gene to offspring thus making prevention and control difficult (Sawaimul *et al.,* 2018).

In Kenya, research reports spanning several years indicate that the prevalence of haemoglobinopathies is significantly high in malaria holo-endemic regions, that include Western Kenya and coastal region (Uyoga *et al.,* 2019; Kosiyo *et al* 201, Byrd *et al* 2019). In a population survey among children in Kilifi County done by Uyoga *et al.,* 2019 on Kenya's Indian Coast Region, reported prevalence of 8.6% for HbS (HbAS, 7.8% and HbSS homozygous, 0.8%) plus 65.5% for α-thalassemia (α-/α, 48.6% and α/α, 16.9%). A review of studies documented in Western Kenya, reveals a rather different haemoglobinopathy prevalence pattern. For instance, Suchdev *et al.,* 2014 in Kano area of Kisumu County reported significantly higher values for HbS but lower ones for α-thalassemia, i.e HbS, 18.7% (HbAS 17.1% and HbSS 1.6%) and Hb α-thal, 48.1% (α-/α, 38.5% ; α/α, 9.6%), respectively, similar to Kifude *et al.,* 2007 in a survey done in Kombewa, area of same County who reported similarly higher overall HbS prevalence of 19.9% (HbAS, 19.0% and HbSS, 0.9%) and 53.2% for α-thalassemia (α-/α, 44.4%

and α/α, 8.8%). This pattern for Western Kenya is corroborated by two more recent studies. One of these, a hospital-based study targeting sickle cell Hb in children resident in Kisumu County with acute *plasmodium falciparum* malaria, by Kosiyo *et al.*, (2021) which reported a HbS prevalence of 31.8% (HbAS, 20.7% and HbSS, 11.1%). It is noteworthy that this particular study did not target thalassaemia Hb. The other study, by Byrd *et al.*, (2019) targeting rural communities in three counties (Bungoma, Kakamega and Vihiga) in the Western Region of Western Kenya, reported HbS prevalence of 16.4% (HbAS, 16.2% and HbSS, 0.2%) and α-thalassemia, 48.2% (-α/-α, 8.2%; and -α/α, 40%). These reports of high prevalence of haemoglobinopathy in Western Kenya, underscores urgent need for laboratory methods that are financially accessible, thus enabling early diagnosis of haemoglobinopathy diseases and detection of haemoglobinopathy asymptomatic carriers.

Early diagnosis of SCD is essential for provision of appropriate clinical support for victims, in minimizing mortality rate. Similarly, early detection of asymptomatic carriers is important in unmasking these individuals who unknowingly transmit the HbS gene hereditarily through marriage. Indeed, screening for haemoglobin disorders has been recommended by WHO as one of the key haemoglobinopathy control interventions that form part of the maternal and child health care (MCHC) services (Tluway & Makani 2017). However, many Sub-Saharan Africa countries including Kenya have not implemented newborn screening, due to the high cost and significantly long turnaround times of currently available laboratory assay technologies. This has resulted into early childhood mortality estimated at 50-90% among children born with sickle cell disease as reported by a prospective study done in Kilifi area of Kenya which was in tandem with

50-80% mortality documented in Sub-Saharan African countries (Arishi *et al.,* 2021, Uyoga *et al.,* 2019). Majority of those who survive five years without diagnosis, live unproductive life's due to painful sickling crises and fatal anaemia resulting from haemolysis and vaso-occlussive crisis triggered by infections which would have been avoided by early diagnosis leading to proper palliative care thus improved productive life (Uyoga *et al.,* 2019).

The red cell distribution width (RDW) is potentially a suitable haemoglobinopathy biomarker that could enable laboratory testing of haemoglobinopathy in regions with financial challenges, like Kenya. That is because of the easy and ready availability of its assay results as a haematological parameter generated by automated haematology analyzers as part of the complete blood count (CBC). It measures variation in RBC sizes and shapes in a cell volume within the red cell population, and its values have been widely reported to be altered in diseases involving red cell destruction or production, critically ill patients (Ramby *et al.,* 2015). It is a biomarker of disease severity in patients who are critically ill and is also an independent predictor of all-cause mortality in sepsis and congestive heart failure (Ramby *et al.,* 2015; Titcomp 2017; Jandial *et al.,* 2017; Kaori and Kamat 2018). Mathematically, it is the coefficient of variation (CV) of the mean corpuscular (cell) volume (MCV), i.e. RDW= 1SD/MCV x 100% (Titcomp 2017).

Several empirical studies have demonstrated significant elevation in RDW values in the presence of sickle cell and thalassaemia haemoglobin and the potential of the haematological parameter to distinguish haemoglobinopathy cases from other haematological disorders (Ramby *et al.,* 2015; Titcomp 2017; Jandial *et al.,* 2017; Kaori

and Kamat 2018). Some of these investigated uses of RDW to distinguish thalassaemia from iron deficiency anaemia (Vehapoglu *et al.*, 2014, Al-Numan *et al.*, 2021, Hoffman *et al.,* 2015, Song *et al.,* 2020, Jahangiri *et al.,* 2019, and Wickramaratne and Wijewickrama 2021). Other studies, for instance Kosiyo *et al.,* (2020) and others, especially older ones (Qurtom *et al.,* 1989 and Thame *et al.,*1991) demonstrate potential of RDW to serve as laboratory biomarkers for discriminating bearers of sickle cell Hb (HbSS and HbAS) and thalassaemias from other haematological disorders.

Despite having demonstrated potential of its application as a diagnostic marker for haemoglobinopathy, the literature reviewed reports no previous study on the utility of RDW in the detection of haemoglobinopathy in Kenya and indeed Africa. Owing to the reported variation in the biological characteristics between human populations in widely separate geographical locations due to the impact of the physical and sociocultural environment on the expression of genes for various characteristics (Boyce *et al.,* 2020), it is uncertain as to whether the findings of these studies are generalizable to Kenya. Besides, even in these studies, the utility of RDW as a haemoglobinopathy diagnostic biomarker has not been rigorously and formally evaluated. There was need therefore to determine empirically and rigorously, the ability of RDW to discriminate haemoglobinopathy and haemoglobinopathy cases in a Kenyan population.

1.2 Statement of the Problem

Haemoglobinopathies are genetically transmitted red blood cell disorders and research reports shows that their prevalence is significantly high in the malaria-holoendemic Lake Victoria regions of Western Kenya. Prevention of transmission and control of related mortality requires early detection, but this has been hampered by the highly expensive

laboratory testing methods currently in use, unaffordable for resource-limited settings. The currently used laboratory testing methods with high specificity and sensitivity, that have adequate ability to discriminate haemoglobinopathy cases and haemoglobinopathy-free individuals include tests such as HPLC, cellulose acetate paper electrophoresis and the DNA-based polymerase chain reaction (PCR), but they are very expensive, suitable only as gold standard, reference methods.

The available affordable methods for resource-limited settings like, the sickling tests, peripheral blood film and Hb solubility test are either of low discriminatory ability or have long turn-around time. Previous research indicate that the red cell distribution width (RDW) is a promising alternative. Its values are readily, and therefore cheaply available as part of the output for complete blood count or full haemogram generated by automated haematology analyzers, routinely done in the investigation of haematological disorders. However, its utility as a surrogate biomarker for the detection of haemoglobinopathy, has not been investigated in Kenya, specifically. Need for knowledge about the performance of this parameter specifically in Kenya is informed by the empirically demonstrated that the strength of phenotypic expression of genotypes for many biological characteristics can vary with the geographical location, race/ethnicity and associated sociocultural factors of the populations studied. Correspondingly, this causes variation in the values for such biological parameters for population of widely diverse geographical regions, race/ethnicity and associated sociocultural backgrounds. Hence, reference values for any biological characteristics set using the population of a given geographical region and sociocultural and ethno-racial background may not necessarily apply to a population of a widely different geographical setting and

sociocultural and ethno-racial background. The normal references for the various disease biomarkers employed by the many laboratories in Kenya to interpret their assay results come with the assay reagent kits, having been derived from populations of the geographical regions where they were manufactured, usually outside Africa. There is a risk that such interpretations could be causing erroneous clinical diagnostic and patient monitoring findings in concerned Kenyan laboratories. Hence, it was necessary not only to evaluate the utility of RDW for haemoglobinopathy detection, but also to first derive the reference cut-off value to use in the interpretation of the assay values.

Western Kenya is one of the two main malaria holo-endemic regions of the country, the other one being the Indian Ocean coastal region. Research reports have shown that these regions of Kenya, have a significantly high prevalence of sickle cell disease and thalassaemia haemoglobinopathies, the dominant haemoglobinopathies globally. Like most of Kenya, a significant proportion of the population in western Kenya live below the poverty line and cannot afford the cost of the expensive currently used haemoglobinopathy detection technologies. Hence, this study evaluated the potential utility of the cheaper, readily available RDW as a surrogate marker for haemoglobinopathy screening and diagnosis in western Kenya.

1.3 Justification of the Study

Hemoglobin disorders may progress unidentified until they manifest clinically, because clinicians manage patients based mainly on clinical features, leading to missed diagnosis and improper clinical management. While newborn screening for haemoglobinopathy has been recommended by UN for its likely contribution to improved clinical management of patients and prognosis, it has not been implemented in many Sub-

Saharan African countries due to low economic capacity to afford the currently used testing technologies. This has resulted into high morbidity and consequent childhood mortality in these countries as many cases die undiagnosed or are diagnosed too late.

Some hemoglobin disorders especially those manifesting as carrier forms, progresses asymptomatically with mixed clinical presentations and normal full haemogram results. Such individuals seem healthy and unaware of their carrier status unless specifically screened. Thus, the probability of them passing haemoglobinopathy genes to offspring remains high, making the prevention and control of associated haematological disorders difficult. Although the carrier condition poses no immediate medical threat, identification of these disorders is immensely important during armed force recruitment, screening of families with history of hemoglobinopathies for genetic counseling and for patient management.

The RDW is easily obtained as one of the haematological indices routinely generated as the complete blood count by haematology autoanalysers. It measures red cell sizes and shapes (anisocytosis) a common feature in hemoglobinopathies. It also becomes abnormal in Hb disorders earlier than all other hematological parameters thus can serve as an effective biomarker. Having studied in countries outside Kenya and African as a biomarker of haemoglobinopathies, in populations of different geophysical location, it was uncertain as to whether such findings were generalizable to the Kenyan scenario.

It is currently available even in low-ranked public hospitals in Kenya, therefore might enable timely detection of Hb phenotypes even in asymptomatic phases thus, unmask the underlying carriers of the genotype in adults. However, no study has been carried in Kenya to show whether or not it can adequately discriminate between individuals with and those without haemoglobinopathy.

1.4 Significance of the Study

This study aimed at assessing the potential of RDW as an accurate and cost-effective method for use in laboratory detection of the inherited haemoglobin disorders. Its findings have therefore contributed to knowledge on the relationship between the values of this haematological parameter and presence or absence of haemoglobinopathy, thus, there is a potential improvement in the management of haemoglobinopathies in Kenya and other resource limited settings by enabling population access to accurate testing and timely diagnosis of these blood disorders. This could enable easy implementation of the WHO recommendation that neonatal screening for haemoglobinopathy be mainstreamed in childhood healthcare programmes in order minimize the child mortality from these hereditary disorders, by enabling early detection and timely clinical interventions.

1.5 Objectives of the Study

1.5.1 Broad objective

To determine the potential of red cell distribution width (RDW) as a surrogate marker of hemoglobinopathies among patients in Western Kenya.

1.5.2 Specific objectives

1. To determine the distribution of haemoglobinopathies phenotypes among patients in Western Kenya
2. To determine the red cell distribution width (RDW) for patients with hemoglobinopathies and those without haemoglobinopathies in Western Kenya
3. To determine the sensitivity, specificity, and overall accuracy of red cell distribution width (RDW) as a diagnostic biomarker for haemoglobinopathies in Western Kenya.

1.6 Research Question

1. What is the distribution of various abnormal and normal Hb phenotypes among patients in Western Kenya?
2. What is the red cell distribution width (RDW) for the haemoglobinopathy cases and control population groups in Western Kenya?
3. What is the sensitivity, specificity, and overall accuracy of RDW in discriminating individuals with and without hemoglobinopathy among patients in Western Kenya?

1.7 Limitation of the Study

The study had the following limitations:

1. Basing the laboratory diagnosis of haemoglobinopathy on detection of phenotypes rather than genotypes might contribute towards inaccuracy of the prevalence of the disorders in the population.

2. Use of retrospective data, collected over the long period of 5 years may carry the risk of unknowable magnitude of variability in the RDW values because different models of haematology analyzers of even different versions of same model of haematology autoanalyzers or different batches of assay reagents for same analyzer could have affected the summary RDW values used to derive the reference cut-off value/limit

3. Using hospital patients, though confirmed haemoglobinopathy-free, as the source for the RDW normal reference population possibly gave an inaccurate reference cut-off value for RDW, thereby confounding the rating of the haemoglobinopathy diagnostic utility of this haematological parameter.

1.8 Delimitation

The study targeted patients investigated for haemoglobinopathy using high performance liquid chromatography (HPLC) and complete blood count (CBC) at the Aga Khan Hospital, Kisumu, western Kenya, laboratory. These were patients who had actually attended the hospital and its western Kenya satellites blood specimens referred to these facilities, during the period January 1st, 2015, to December 31st, 2020.

1.9 Conceptual Framework

This study was guided by knowledge from empirical studies which have demonstrated a relationship between red cell distribution width (RDW) and haemoglobinopathy, through the effect of the latter on the pace and magnitude of erythropoiesis. Basically, erythropoiesis is triggered by chronic tissue hypoxia that results from chronically compromised ability of haemoglobin to capture and deliver oxygen to the tissues. Briefly, sickle cell Hb and thalassaemia that cause, respectively formation of structurally defective globin moiety and absence or inadequate quantities of haemoglobin (Branchaleoni *et al*., 2016, Machogu and Machado 2018), are the main haemoglobinopathies globally. The thalassaemias are characterized by microcytic, hypochromic erythrocytes (Brancaleoni *et al.,* 2016), meaning erythrocytes of victims have inadequate capacity to capture and deliver oxygen to body tissue cells. Similarly, for sickle cell Hb, defective molecular structure of the oxygen-carrying part of haemoglobin lowers capacity of victim's blood to transport oxygen to the tissues. This limited or absent capacity to transport oxygen causes oxygen deficiency in the tissues (tissue hypoxia).

As demonstrated in studies on mountain climbers, who ascend to higher altitudes suddenly, the resulting tissue hypoxia due to the low oxygen concentration at such altitudes stimulates enhanced erythropoiesis at a faster pace than under conditions of normoxia (Mairbaurl 2018). Thus, hypoxia is the general regulator of erythropoiesis and, hence, increased erythropoiesis the classical response to hypoxia (Vlaski *et al.*, 2009, Haase 2013). Therefore, existence of haemoglobinopathies leads to enhanced erythropoiesis and at rapid pace, as the ultimate physiological compensatory response to the associated hypoxia (Zigot *et al.*, 2018, Vlaski *et al.*, 2009). The rapid pace of erythropoiesis means a significant number of immature erythrocytes being released into the peripheral blood circulation, reticulocytosis (Maibaurl 2018, Mandal and Kartthik 2019). Reticulocytes are macrocytic and, together with the microcytosis of thalassaemia, results into a mixture of erythrocytes of varied sizes and even shapes in the peripheral blood of haemoglobinopathy subjects. This size variation of erythrocytes is known as anisocytosis and leads to elevated values of red cell distribution width (RDW), the major feature of haemoglobinopathy (Ropero 2022, Qurton *et al*.,1989, Matos *et al*., 2015, Webster *et al.,* 1986, Tariq *et al.,* 2019, Dugdale *et al.,* 2018).

The effect of haemoglobinopathy on erythropoiesis can be modified by the presence of the fetal haemoglobin (HbF) and the environment in form of hematopoietic nutrient deficiency, commonly cobalamin, folate and iron. Fetal Hb has higher affinity for oxygen than normal adult Hb, countering the hypoxic effect of haemoglobinopathy and, though meant to be diminished in post-fetal life, can be co-inherited with the other haemoglobinopathies and thus appear postnatally (Mandal and Kartthik 2019). Iron deficiency causes formation of inadequate quantities of Hb leading to microcytic

hypochromic erythrocytes (Piriyakhuntorn *et al.*, 2018), while folate and cobalamin deficiency leads to production of macrocytic, hypochromic erythrocytes (Koury and Ponka 2004). This relationship between RDW and haemoglobinopathy through enhancement of erythropoiesis, together with the effect of HbF and haematopoietic nutrient defficiency is summarized in the conceptual model below (***Figure: 1.1***).

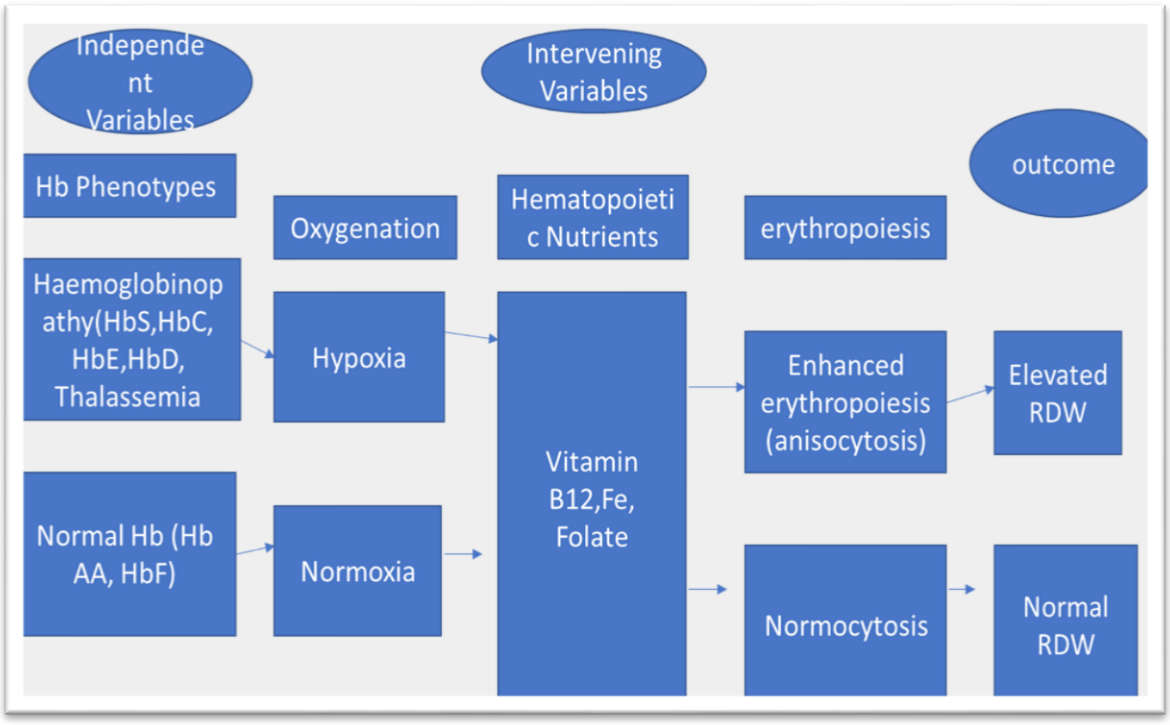


Figure 1.1 Conceptual model illuminating on various Hb phenotypes and their link with RDW, Hb S hemoglobin S, HbC; hemoglobin C, HbE; hemoglobin E, HbD; hemoglobin D, HbAA normal hemoglobin AA, HbF; hemoglobin F

CHAPTER TWO

LITERATURE REVIEW

2.1 Definition and epidemiology of haemoglobinopathies

Hemoglobinopathies are hereditary genetic disorders of hemoglobin that are acquired as recessive autosomal diseases that have resulted as one of the major genetic public health problem globally seen in infancy and childhood (Trent 2006, Suchdev *et al.,* 2012, Sawaimul *et al.,* 2018, Vinjamur *et al.,* 2018) where children born with sickle cell disease living in Sub-Saharan African countries have an alarming mortality rate of 50-80% by the age of five years (Arishi *et al.,*2021). It confers significant morbidity and mortality among the victims, and its epidemiological profile demonstrates an ethnic and regional diversity of gene polymorphism among the haemoglobin variants (Sawaimul *et al.,* 2018), suggesting need to consider ethnicity and geographical location of populations in related research and interventions. With a vast growing majority of haemoglobinopathy carriers, researchers have recommended population screening in potentially susceptible populations to unmask individual unknown cases (Sawaimul *et al.,* 2018).

Kenya has a significant burden of hemoglobin disorders indicating urgent need for a cost-effective method to enable proper and early diagnosis. In a cross-sectional survey involving 858 children aged 6-35 months, Suchdev *et al.,* 2012 determined the burden and the consequences of inherited blood disorders among young children selected randomly from 60 villages in Western Kenya. The study revealed that more than 2 out of 3 children had at least one blood disorder where sickle cell trait (Hb AS) and sickle cell Hb (Hb SS) were found in 17.1% and 1.6% of children respectively; 38.5% were

heterozygotes and 9.6% were homozygotes for alpha-thalassemia. Taken together, these findings presuppose that there could be many other hemoglobin disorders of which existence in the population remains undocumented. Consequently, a simple method to unmask cases needs to be established. To this end, WHO has recommended population and newborn screening for hemoglobin disorders as a key part of countries' health services (Tluway & Makani 2017). However, many Sub-Saharan Africa including Kenya have not implemented these screening due to the high financial cost of existing high-quality laboratory methods for haemoglobinopathy testing that includes haemoglobin electrophoresis and genotyping.

To detect haemoglobin blood disorders, laboratory testing is the best method to use just like other blood disorders. However, it has been reported that children born in Sub-Saharan African countries with haemoglobinopathies die before the age of 5 years due to deficiency of effective detection methods for its early diagnosis (Debaun *et al.,* 2019; Makani *et al.,* 2011). In a study done in Kenya, Kilifi region by Uyoga *et al.,* 2019 reported similarly a high mortality rate of 50-90% of children born with SCD that were enrolled in outpatient clinic, which was in tandem with reports by Arishi *et al.,*2021 who documented 50-80% mortality rate of undiagnosed children with SCD across Africa. These reports recommended priority to be given on diagnosis and proper management of sickle cell disease in clinical research which the present study sought to determine the utility of RDW as biomarker for haemoglobinopathies in resource poor settings. In addition, Debaun *et al.,* 2019 and Makani *et al.,* 2011 approximated that by the year 2050, over 14 million infants will be born with SCD if proper measures are not established to combat this growing problem which they estimated 82% of the 14 million

sicklers will come from Sub-Saharan African countries. The Aga Khan Hospital, Kisumu and its Western Kenya satellite centers sits within malaria-holoendemic region of Lake Victoria Economic Block regions, which is known to have high *plasmodium falciparum* linked to haemoglobinopathies especially sickle cell disease (Suchdev *et al.*, 2012). Kosiyo *et al.,* 2020 in Kisumu County demonstrated differential diagnosis of malaria in SCD using hematological parameters while Kifude *et al.,* (2011) reported a high prevalence of both α-thalassaemia and sickle cell haemoglobinopathy among children in the same region. Therefore, there is need for a less costly methods for laboratory testing for haemoglobinopathies in Western Kenya.

Studies have reported previously the great utility of RDW in differentiating iron deficiency anaemia from other microcytic anaemias (Sarah *et al.,* 2018, Aulakh *et al.*, 2009 and Eldibany *et al.,* 1999) while others have shown its discriminating ability of iron deficiency anaemia from thalassemias (Matos *et al.,* 2016, Miri-Moghaddam *et al.*, 2014, and Sharma *et al.,* 2016). Therefore, RDW is proving to have the ability to differentiate haemoglobinopathies from other red cell disorders associated with anaemia thus, can be utilized as a cost-effective haematological parameter in detection of haemoglobinopathies in resource limited setting of Western Kenya. Okwi *et al.,* 2010 studied the reliability and cost-effective method to diagnose SCD between sickling test, solubility test and PBF. The authors found out that sickling test had a high sensitivity and specificity and recommended it for use in screening of SCD children in poor countries. However, the study targeted only SCD leaving other hemoglobin disorders including thalassemias which the present study covered. The study also recommended screening of children only, while the unknown adult carriers (SCT) continue to transmit

25% SCD offspring as documented by previous studies (Sawaimul *et al.,* 2018) which the present study sought to unmask these carriers from the general population. In another study by Piety *et al.,* 2016 validated use of low-cost paper-based screening test for sickle cell haemoglobinopathy whose results had high accuracy and therefore could be effective in poor clinical settings. However, the procedure was not satisfactory as it would need well trained staff especially in the visual translation of Hb AA, Hb AS and Hb SS colors in addition to frequent re-constitution of hemoglobin solubility buffer which was usable per day thus, consuming large volumes of reagents in the process. Stringent temperature demanding reagents as the test is performed at 18-22 ^{0}C room temperature will be inhibitory as most clinical laboratories in poor countries especially sub-Sahara Africa cannot afford. Its suggested long turnaround time of 35 minutes to prepare the test, would be ineffective for fast screening of large population.

Essentially, red cell distribution width measures erythrocytes sizes (anisocytosis) and shapes (poikilocytosis) in red cell population which are very common erythrocytes phenotypic features in haemoglobinopathies especially sickle cell disorders (Ramby *et al.,* 2015). In clinical laboratory, RDW is generated routinely as part of complete blood count (CBC) output by haematological analyzers. Therefore, RDW is a simple, faster, cheaper, and widely used in routine practice as part of a full haemogram report. RDW has been studied well as a significant entity in various disease pathogeneses including a marked variation in size (anisocytosis) and shape (poikilocytosis) in diseases associated with red cell destruction or heightened compensatory production in response to the RBC destruction. It is a biomarker of disease severity in patients who are critically ill and is also an independent predictor of all-cause mortality in sepsis and congestive heart failure

(Ramby *et al.,* 2015). RDW along with MCH and MCV improved identification of Hb E hemoglobinopathy in Sri Lanka (Nishad *et al.,* 2014) with significant clinical value in sickle cell disease (Thame *et al.,* 1991). It is derived and presented as a coefficient of variation, CV.

Despite studies showing the great utility of RDW as a biomarker of haemoglobin disorders, data concerning its use in Kenyan population remain poorly documented. Gene expression that influences phenotypic characteristics is known to be altered by environmental and sociocultural epigenetic changes bringing about variation in haematological, immunological and biochemical characteristics in humans (Saulnier & Dupras 2017, Boyce *et al.,* 2020). It is uncertain therefore as to whether the findings of studies of RDW values in other populations in different geographical locations can be applied in Kenyan Scenario, especially in malaria-holoendemic region of Western Kenya. The overall goal of the present study was to offer timely diagnosis of haemoglobinopathies by use of accessible but affordable laboratory assay thus improving chances of survival of infants and children born with haemoglobinopathies in Western Kenya population.

Studies have shown that hemoglobin disorders pose a major public health risk in future if proper programmes are not established since some inherited haemoglobin disorders, usually results into death by the age of 5 years (Weatherall, 2011). Unmasking the carriers is equally an important part of the profiling of the major hemoglobin disorders for family counselling, to avert translation into major hemoglobinopathies (Sawaimul *et al.,* 2018, Tluway & Makani 2017). A study done by Weatherall, 2011 documents that

the sickle cell haemoglobinopathy is widely spread out in the whole of Sub-Saharan African countries, parts of the Indian sub-continent and Middle East where carriers have higher prevalence of 5% to 40% in their population (Taher *et al.,* 2021). These studies indicate that thalassemias have high prevalence's in Mediterranean basin, Africa, Middle East, South-East Asia, the Indian sub-continent, Melanesia and into the Pacific Islands (Muncie & Campbell 2009). βeta-thalassaemia have 1%-20% in these areas while α-thalassaemia has a higher prevalence in sub-Saharan Africa (Weatherall & Clegg 2001). The α-thalassemia majors are commonly restricted in South-East Asia and Mediterranean basin thus, poses less of a global problem (Muncie & Campbell 2009). Therefore, the author recommended to international health agencies and governments of countries to use this information to develop strategies on how to combat these diseases (Weatherall, 2011).

To that effect, UN and WHO recommended screening of adult population and newborn to reduce incidences as an effective approach to reduce the burden of these diseases in society (Modell & Darlison 2008). However, this has not been implemented in many sub-Saharan Africa countries due to financial constraints (Tluway & Makani 2017) resulting into a high childhood mortality rate of about 50-90% on children born with SCD (Arishi *et al.,*2021, Uyoga *et al.,* 2019) in addition to lack of adequate documentation on carriers who remain asymptomatic therefore, risking to reach the predicted enormous burden of over 14 million children born with major haemoglobinopathies by the year 2050 as estimated by previous studies (Grosse *et al.,* 2011, DeBaun &Najibah 2019). This would be an overwhelming burden to families and

healthcare system thus the present study sought to intervene by evaluating the ability of RDW that would enable early detection thus timely clinical intervention.

2.2. Classification of Hemoglobinopathies

Hemoglobinopathies are a group of inherited genetic disorders that results from mutation or deletion of gene encoding the synthesis of specific amino acids comprising the polypeptide chain of the hemoglobin protein part (Vinjamur *et al.,* 2018). Currently there are more than 1000 hemoglobinopathies that have been identified and characterized. They are classified into two major groups: Thalassemia syndromes and structural hemoglobin variants (Forget & Bunn 2013). The authors define thalassemia syndromes as disorders that affect the rate of production of the globin molecule due to deletion or inactivation of genes that make up the globin chain. The deletion or inactivation of these genes can be heterozygous or homozygous which determines the degree of the resulting thalassemia syndrome. Structural hemoglobin abnormalities are a group of hemoglobin inherited disorders resulting from substitution of amino acids, amino acid deletion or amino acid addition in the coding sequence of one or more of the globin chains and are majorly Hb S, Hb C and Hb E (Vinjamur *et al.,* 2018). Since Western Kenya is a *plasmodium falciparum* endemic zone, the present study focused more on hemoglobinopathies whose frequency and distribution are more prevalent in regions endemic for *plasmodium falciparum* (Kosiyo *et al.,* 2020). The reason is that RBCs in Thalassemia and some Hb variants such as sickle cell Hb (Hb S), Hb C and Hb E are protective to malaria, therefore are less likely to develop severe malaria (Trent 2006 & Gonçalves *et al.,* 2016).

2.2.1 Thalassemia syndromes

Thalassemia is a group of hemoglobin disorders characterized by a defect in the rate of production of one or more of globin chains. In alpha (α) Thalassemias, α -globin subunits are absent or reduced and, in the beta, (β) Thalassemias, β-globin subunits are absent or reduced (Forget & Bunn 2013, Vinjamur *et al.,* 2018, Sadiq *et al.,* 2021).

2.2.2 Beta thalassemia

Beta chain genes are found on chromosome 11 which also carries other gene loci that includes: G-gamma and A-gamma chains, and embryonic epsilon chains (Needs *et al.,* 2020). Gene deletion or inactivation in chromosome 11 determines the form of beta thalassemia that develops depending on whether it is fully or partly affected. The greater the degree of deletion, the greater the severe form of thalassemia that develops and the greater the degree of anemia (Mansilla *et al.,* 2016, Sadiq *et al.,* 2021). There are neighboring genes that border beta genes referred to as delta genes that can be deleted together with beta genes thus resulting into delta-beta thalassemia syndrome (Forget & Bunn 2013). Beta thalassemia is classified based on genotypic and phenotypic forms. In genotypic notation of beta thalassemia, a "$^{+}$" represents a reduction in beta chain production, a "^{0}" represents a complete deletion whereas '^{sc}' represents silent carrier, thus, giving six forms of beta thalassemia as follows: B^0/B^0, B^0/B^+, B^+/B^+, B^0/B, B^+/B, B^{sc}/B, (Galanello & Origa 2010) In the phenotypic system, there are four types of beta thalassemia that depend on the degree of clinical presentation experienced by the patient which includes: Beta thalassemia major, Beta thalassemia intermedia, Beta thalassemia minor, Beta thalassemia minima (Fibach *et al.,* 2017, Forget & Bunn 2013, Needs *et al.,* 2020).

Beta Thalassemia Major is also referred to as Cooley's anemia and genotypes associated to it are: B^0/B^0, B^0/B^+, or B^+/B^+. Deletion of these two genes results to very few or no beta chains being produced, and Hb A is almost zero percent (Fibach *et al.,* 2017). It is the most severe form of beta thalassemia and children with this form of disorder develop clinical signs during their first year of life. They appear to be malnourished and may exhibit abdominal girth expansion referred to as Mongoloid facial features (Needs *et al.*, 2020). Beta Thalassemia Intermedia is attributed to a wide variety of genotypes including B^+/B^+, B^0/B^+, or B^0/B. Production of beta globin chain is significantly reduced in all the genotypes and subsequent reduction in the quantity of Hb A. Patients with this disorder have a normal life span. However, some patients may have facial bone deformity and/or splenomegaly (Forget & Bunn 2013).

Beta thalassemia minor is expressed as B^0/B or B^+/B, and these individuals are usually healthy and unaware of their status and rarely show clinical presentations (Galanello & Origa ,2010). The body is able to generate sufficient Hb A thus oxygen delivery is near normal and red cells have normal life span, therefore patients do not need treatment (Fibach *et al.,* 2017). This is similar to Beta Thalassemia Minima (Silent Carrier) B^{sc}/B whose beta genes are partially deleted or inactivated or mutated therefore, hemoglobin A produced is normal or near normal and cannot be recognized unless unmasked during family screening. Deletion of Delta-$beta^0$ / Delta-$beta^0$, results into Delta-Beta thalassemia major where only Hb F is produced (Needs *et al.,* 2020). This type of thalassemia is common in many ethnic groups globally, however, persons from Africa, Greece and Italy are majorly affected.

2.2.3 Alpha thalassemia

Alpha thalassemia, results from the deletion or inactivation of either of the four genes in chromosome 16 which codes for zeta and alpha hemoglobin chains and each of the chromosome has two loci alpha chains, α1 and α2 (Farashi & Harteveld 2018). This gives a total of 4 genes that codes for the alpha globin chain. There are four phenotypic forms of alpha thalassemia that are developed based on the degree of deletion or inactivation of the alpha genes which includes: silent carrier (only one gene is deleted out of four), thalassemia minor (two genes out of four are deleted), hemoglobin H disease (three of four gene loci deleted) and Alpha thalassemia major or hydrops fetalis (the four gene loci deleted) (Harewood & Azevedo 2021). In alpha thalassemia genotypic notation an "α" represents the presence of an alpha locus. A ‘‘-’’ represents a deletion of a locus. The notation for the normal number of alpha loci is αα/αα that produces a normal Hb A of 95-98 % (Muncie & Campbell, 2009).

In Alpha Thalassemia Silent Carrier, deletion or inactivation involves only one gene (-α/αα) and Hb A is produced at its fullest potential of 95-98% (Galanello & Cao 2011) and individuals who have this disease do not have any clinical presentation. Deletion or inactivation of two genes in one chromosome (homozygous) or in either of the two chromosomes (heterozygous) results into alpha thalassemia minor. In the homozygous genotype (-α/-α), both parents contribute one missing locus while in heterozygous state (--/αα), one parent contributes a normal gene, but the other parent gives deleted gene in both alpha chains (Farashi & Harteveld 2018). When three genes have been inactivated or deleted, Alpha thalassemia intermedia forms (--/-α) which have 70-90% hemoglobin A produced where the surplus beta globin chains combine with each other forming a

beta tetramer referred to as Hemoglobin H (β4) ((Harewood & Azevedo 2021) which is lethal hemoglobin disorder that usually result into stillborn infants. Alpha thalassemia major, involves complete deletion of alpha chain loci (--/--) resulting in the production of gamma chain tetramers γ4 (hemoglobin Bart's) which is usually incompatible with life (Farashi & Harteveld 2018, Harewood & Azevedo 2021).

2.2.4 Haemoglobin S

A normal Hb A has glutamic acid at the 6th position in the gene that code for the beta globin chain (Ribeil *et al.*, 2017). Forget & Bunn 2013, defines sickle cell hemoglobin (Hb S) as an inherited disorder that results from Substitution of valine for glutamic acid as the sixth amino acid. Sickle cell trait has been demonstrated to show resistance against most species of malaria in different regions of the world (Gonçalves *et al.*, 2016). Sickle cell disease develops when an offspring inherits two mutated alleles S/S (homozygous) or inheriting mixed mutated alleles (heterozygous) such as sickle-thalassemia, sickle-hemoglobin C (Hb SC) and other combinations (Arishi *et al.*, 2021). Sickle cell trait is also a form of heterozygous whose one allele is affected (Hb AS) but produces a normal haemoglobin. In sickle cell disease, polymerization of hemoglobin S is triggered by low oxygen making the cells deformed. Repeated polymerization destroys red cell membranes and shape irreversibly (Dufu *et al.*, 2016). The structure of haemoglobin molecule is functionally impaired molecule, including inability of affected erythrocytes to transport oxygen in the body effectively thus are marked for haemolysis by the victim's reticuloendothelial system (Ilesanmi, 2010).

Arishi *et al.*, 2021 documented that this increased cell destruction reduces red cell life span by 75% in addition to small vessels blockage i.e. vaso-occlusion that leads to

painful crisis and acute chest problems which is the major cause of death in SCD patients. Reticulocytes that contain Hb SS are sticky and on interacting with proteins of the vascular endothelium triggers inflammation that leads to vascular occlusion resulting into tissue necrosis, with consequential fatal or life-threatening severe anaemia and hypoxia (Ilesanmi, 2010). Organs that are majorly affected include brain, bone marrow, lungs, liver kidneys and the spleen. Predisposing factors that trigger this mechanism includes infection, acidosis, fever, cold temperatures, dehydration, anxiety, depression, and stress (Dufu *et al.,* 2016). These results in death of infants in the early years of life in regions that lack proper management of SCD especially Sub-Saharan African countries (Arishi *et al.,* 2021).

2.2.5 Haemoglobin C (Hb C) and Hemoglobin E (Hb E)

Gonçalves *et al.,* 2016 published Hemoglobin C (Hb C) as a disorder caused by a substitution of glutamic acid by lysine mutation at the 6th position of the amino acid sequence of beta globin. The authors suggested that homozygotes for (genotype C) are strongly protected against severe malaria with heterozygous states being mildly protected. Piel & Weatherall (2015), defines hemoglobin E (Hb E) as a disorder that results from substitution of glutamic acid for lysine at the 26th position in the gene that make up the beta globin chain (Nienhuis & Nathan 2012). Hemoglobin E is majorly restricted in Southeast Asia and other Asian countries whose pathophysiology remains unknown (Fucharoen & Weatherall, 2012).

2.3 Laboratory Detection of Haemoglobinopathies

2.3.1 Hb electrophoresis

Hemoglobin electrophoresis is defined as a laboratory assay that involves movement of hemoglobin proteins in an alkaline or acid pH with sieving absorbent materials such as gel or paper165 that is able to identify normal and abnormal hemoglobin types (Santhosh & Sampath (2011). The hemoglobin types comprise of hemoglobin A_1 (HbA_1), hemoglobin A_2 (HbA_2), hemoglobin F (HbF; fetal hemoglobin), hemoglobin C (HbC), and hemoglobin S (HbS). The normal range of haemoglobin types in a normal healthy adult includes: HbA1: 95%-98%, HbA2: 1.5%-3.5%, HbF: < 2% (age-dependent) HbC: Absent HbS: Absent (Khanam *et al.*, 2018). Since different types of haemoglobin molecules have different charges and sizes, they show different degree of mobility. Other factors like the pore size of the medium and ionic concentration of the buffer also determines how far a molecule migrates. Hemoglobin A, A2, S, F and C have net negative charge under alkaline condition and moves towards the anode (positively charged electrode) (Kim *et al.*, 2017). In alkaline electrophoresis, a charged hemoglobin proteins which are in form of ions (Zwitterion) move from cathode (negatively charged electrode) to anode (positively charged electrode) in a buffered cellulose acetate medium at a pH of 8.4-8.6 after which a dye is applied to the medium to be able to visualize the bands that are also measurable by densitometry (Kim *et al.*, 2017).

In a normal adult, Hb A (hemoglobin A) migrates the furthest, followed by Hb F (hemoglobin F). Hemoglobin A2 (Hb A2) moves slightly away from the cathode (the point of origin). In individuals with abnormal hemoglobin, migration of these proteins follows the following order: Hb C and Hb A moves together slightly away from the

cathode, Hb S lies between Hb A2 and Hb F. Hemoglobin H moves the furthest followed by hemoglobin Bart near the anode (Kim *et al.,* 2017).

Hb electrophoresis is considered to be effective in rapid screening of small number of samples and may be inaccurate when dealing with small sample concentrations. Hb electrophoresis carried out in capillary tube is referred to as capillary zone electrophoresis (CZE) 165 (Kim *et al.,* 2017). It was recommended therefore, by Adu *et al.,*2017, that hemoglobin electrophoresis data be presented as 'haemoglobin Phenotypes' at a specified pH but not' haemoglobin genotype'.

Thus, the present study compared red cell distribution width to Hb phenotypes performed using HPLC. Red cell distribution width is simple, reliable, and less costly and will serve as an adjunct pointer for screening hemoglobinopathies in societies with limited financial capacity.

2.3.2. High performance liquid chromatography—HPLC

High Performance Liquid Chromatography (HPLC) is a sensitive technique that is able to separate molecules based on their sizes and charges using cation exchange chromatography thus identifying hemoglobin proteins subtypes in the blood sample (Shah *et al.,* 2021). Granular silica or other polymers are used as sieving medium in HPLC where fluid is passed through by the pressured pump then the densitometry is able to detect the separation. Therefore, HPLC is a better sensitive in separation of hemoglobin variants than electrophoresis since proteins have different interactions with the stationary phase (Arishi *et al.,* 2021) and has specific retention time that can be compared to hemoglobin fractions that are known (control). Arishi *et al.,* 2021 documents that HPLC can detect as well as quantify Hb A_1, Hb A_2, Hb F, Hb S, Hb C

Hb Barts and other variation. HPLC requires less labor and is more reliable in quantification of Hb levels making it useful in monitoring patients under blood transfusion or hydroxyurea than Hb electrophoresis (Alapan *et al.,* 2016, Roy *et al.,* 2019). Bio-Rad D10 is one of these machines that applies the HPLC principal, however, it is an expensive machine that is not accessible in resource poor setting and cannot differentiate among all variants with the same retention time (Arishi *et al.,* 2021, Alapan *et al.,* 2016, Kim *et al.,* 2017). The haemoglobinopathies data collected in the present study was done using HPLC (Bio-Rad D10) machine manufactured by Bio-Rad Laboratories based in Hercules, California United States

2.4 Red Cell Distribution Width and Haemoglobinopathy Phenotypes

RDW measures degree of anisopoikilocytosis (RBC sizes and shapes) in cell volume within the red cell population and can be directly calculated from the RBC histogram as part of full haemogram. However, it is currently generated by automated haematology analyzers as part of the CBC. This erythrocyte/red blood cell parameter has been studied fairly well and has been linked to many different disease pathogenesis (Ramby *et al.,* 2015). This article notes that diseases resulting in erythrocyte haemolysis or production can increase anisocytosis thus lead to RDW elevation. RDW might therefore enable the detection of the Hb phenotypes in children's asymptomatic phases and thus unmask the related/underlying genotype carriers in adults.

Webster and Castro (1986) studied red cell distribution width in sickle cell disease of adult patients that included HBSS, HBSC, and HBS-thal and provided useful information on how to project the severity of sickle cell anemia in adults. A study done by Thame *et al.,* 1991 found out that RDW was moderately increased in sickle

cell+minor beta thalassemia and in sickle cell+hemoglobin C (SC) disease while marked increased in homozygous sickle cell disease and sickle cell+major beta thalassemia therefore demonstrating the importance of red cell width in diagnosis of sickle cell disease. Another study Qurtom *et al.,* 1989 compared the mean of RDW in iron deficiency anemia with that of normal individuals and patients with beta thalassemia trait. It showed that there was statistical difference between thalassemia and iron deficiency anaemia. Clearly these studies are dated and, 30 years since the latest of them was done, significant changes in the human population structures and other environmental condition that can affect genotypes and their phenotypic expression might have occurred. These conditions have potential to alter epidemiologic profiles of genetically affected health problems including haemoglobinopathies. Besides, the studies targeted only adults, and yet clinical problems associated with haemoglobinopathies are mostly common in childhood (Debraun & Galadana 2019). Having no similar study done in Kenya and also in the entire African continent, there was need therefore to study the potential of RDW in Western Kenya to close this critical gap in knowledge.

2.4.1 Derivation of the normal reference cut-off value (limit) for RDW

In a retrospective study done in Chennai Southern India by Subhashree *et al.,* (2012) to develop gender reference ranges for MPV, RDW, PDW and other complete blood count parameters for the healthy and adult population in that region, they discovered that existing haematological reference values were different from their values indicating the importance of developing region- and –gender specific reference intervals. Therefore, this present study generated ninety-five percent (95%) confidence intervals/limits for

RDW based on age and gender in normal control group in Western Kenya population. Another study in Turkey by Tonbul *et al.,* 2011 determined RDW in newborn and found out that the normal range of RDW differs among newborns in terms of gestational age. The authors suggested that RDW values be evaluated according to these specific results for diagnosis of newborn blood disease without respect to adult or child values. This is because, RDW in neonates is elevated as compared with that in adults and reflects significant variability in the size of the RBCs. The article recommended further studies be carried out for determining baseline of RDW in newborn. Therefore, there is need to consider age and gender while setting reference values for RDW.

In order to use any laboratory analyte as a biomarker for a pathological condition, the requisite normal reference limit or, other, cut-off value is necessary (Placzkowska *et al.,*2022). The reference limit or cut-off value for RDW was therefore a necessary step in the evaluation of this haematological index as a possible haemoglobinopathy surrogate biomarker. The first step in this regard was choosing of the method for derivation of the RDW limit or cut-off value. This includes the type of source or, other, reference, population for the relevant individual RDW values and, the applicable statistical technique in the computation of the normal reference interval at the appropriate confidence level. The source for the reference or control value had to be a population of patients with normal haemoglobin phenotype (HbAA) and considered free from clinical conditions known to significantly alter RDW values towards the pathological conditions (Bakan *et al.,* 2016). Previous studies described two types of source populations for laboratory reference values. One is the healthy populations as far as the disease for which the reference values are to be derived and is considered to be the ideal source for the requisite data, and their use has been referred to as the direct method for the purpose

(Arzideh *et al.,* 2021, Placzkowska *et al.,* 2022). However, gathering the requisite numbers of the necessary mix of healthy subjects is always a daunting task in terms logistical feasibility and financial affordability. The second, and more readily obtainable population used as source or reference population for this purpose, is the hospital patients routinely investigated for the relevant analytes by the concerned laboratory. Use of this type of the stored laboratory data from this type of population constitutes the indirect method for deriving laboratory reference values the various disease biomarkers and this is widely reported in research literature (Placzkowska *et al.,* 2022, Arzideh *et al.,*2021, Farrell and Nguyen 2019, Yan *et al.,*2022, and Ozarda *et al.,*2021).

Accordingly, the RDW and corresponding haemoglobinopathy-status data from the laboratory database were used for this purpose. That is because laboratory data for patients confirmed free from a given pathological conditions, tend to be similar to those of the healthy population (Battacharya 1967, Bakan *et al.,* 2016). Subjects HPLC-confirmed to be haemoglobinopathy-free (bearers of HbAA phenotype) by the Aga Khan Hospital's Kisumu laboratory during the same period as the haemoglobinopathy cases (2015-2020), and had their RDW values as well, were used. Red cell distribution width data are usually generated as part of the complete blood cell count (CBC) by the modern automated haematology analyzers and form part of the laboratory's database thus, their use to derive data for computing reference values was easier.

Secondly a choice had to be made about the statistical approach to use in the computation of the reference limit. The RDW was expressed as the median and interquartile range (IQR) of the percentage coefficient of variation (CV) of the arithmetic mean. The choice depends on whether the data conform to the Gaussian or

normal distribution. The Shapiro-Wilks and Kolmogorov-Smirnov tests (Nahm, 2016) revealed that the values of RDW for this control group were not normally distributed ($p<0.05$) *(See appendix VII).* Therefore, the non-parametric, rather than parametric, statistical approach was found appropriate for use in the analysis of these RDW data and thus computation of the reference value, as recommended by Nahm (2016). Accordingly, the median and interquartile range (IQR) were used to determine the reference limit or cut-off point for discriminating between the haemoglobinopathy-positive and haemoglobinopathy-free patients. The reference limit for the RDW was in this case defined as the upper limit of the 95% confidence interval (CI) of the median (based on the corresponding interquartile range, IQ) of the control group. This cut-off value, sensitivity, specifity and Youden index in haemoglobinopathy determination was evaluated using Receiver operating characteristic curve (ROC curve) (Ruopp *et al.* 2008).

2.5 The Sensitivity, Specificity and Overall accuracy of RDW as Diagnostic Biomarker for Haemoglobinopathies

The receiver operating characteristics (ROC), principal parameters for evaluating the utility of a biomarker for laboratory assay methods for various biomedical analytes are sensitivity, specifity and Youden Index (Accuracy). This evaluation of a method performance parameters in a continuous data, is based on reference cut-off values (ROC curve coordinates) to correctly detect the analyte and/or discriminate individuals with or without the concerned abnormality (Ruopp *et al.,* 2008). Several studies have investigated the sensitivity, specificity and predictive validity of RDW and other

hematological parameters, for the various haematological disorders (Habibzadeh *et al.*, 2016, Kallner 2018).

In differential diagnosis of iron deficiency anemia from other microcytic anemias using RDW, Sharma *et al.,* 2016 found out that this haematological parameter has a high sensitivity in discriminating IDA from Thalassemia. They therefore concluded that it can be used as a cost-effective biomarker in early detection of iron deficiency anemia. Sarah *et al.,* 2018 shows that red cell distribution width is an early predictor of iron deficiency anaemia in pregnancy where it was noted to be elevated in iron deficiency anemia of pregnancy. This study showed RDW sensitivity of 92% and a specifity of 84.7% in marking the presence of iron deficiency anemia in pregnancy thus proving to have diagnostic utility in clinical diagnosis of disorders. The authors concluded that RDW is affordable compared to iron profile testing, hence could be used as the first screening tool in iron deficiency of pregnancy. To evaluate diagnostic reliability of hematological indices Miri-Moghaddam *et al.,* 2014 compared ten discrimination indices in distinguishing IDA (Iron deficiency Anemia) from β Thalassemia (BTM) on 100 BTM and 70 cases with IDA in southern Iran. Since the two are the most frequent hypochromic microcytic anemias, they have distinct prognosis and treatment, so they need to be differentiated quickly using automated blood count analyzers. The authors calculated sensitivity, specificity and the overall accuracy of cut-off values for every formula in that population. This article found out that, discrimination indices showed significant diagnostic value based on the area under the ROC curve (AUROC) where Green & King (MCV^2 RDW/100HB), England & Frazer (MCV-RBC-(5HB)-3.4) and Sirdah formulae (MCV-RBC-(3HB) had their respective AUROC of 0.909, 0.907, 0.904

in South-east of Iran, thus developing a relatively different cut-off values for every formula. They also noted that the spectrum of β thalassemia mutations, have different effects on various RBC indices, therefore, it suggested to determine cut-off value for these formulae in different populations (Miri-Moghaddam *et al.,* 2014).

A study published by Matos *et al.,* 2016, developed index formula for discriminating microcytic hypochromic anemias which is majorly iron deficiency anemia (IDA) and thalassemia trait (TT). Discriminating the two microcytic anemias is vital because they have different treatment and prognosis. Their diagnosis requires several gold standards tests that require advanced technology and are not available in resource limited settings. Matos *et al.,* 2016 urges that though several indices to have been suggested as initial simple screening tools to differentiate IDA from thalassemia trait, some discriminative indices use parameters in the formulas that are produced by advanced blood counters that are not always affordable to the small laboratories. Therefore, the authors developed Matos & Carvalho Index (MCI) that utilizes MCHC and RBC which are available in all blood counters. This article demonstrated that MCI is a useful tool in discriminating Iron deficiency anemia from Thalassemia carriers which are both microcytic anemias. The index formula Matos & Carvalho Index (MCI) = (1.91 × RBC) + (0.44 × MCHC). The ROC curve generated an MCI cut-off value of 23.85 which meant that any individual whose MCI was < 23.85 was classified to have iron deficiency anaemia while > 23.85 was expected to be a thalassemia trait carrier. In their validation exercise using the gold

standard diagnostic techniques for IDA and TT, MCI performance was analyzed giving a sensitivity of 99.3% and a specifity of 76%. Its Youden index of 95% proved to be a good method for screening. Though this was an excellent discriminant for microcytic anemias, other hemoglobin disorders were not included in the study. RDW ability to discriminate hemoglobin disorders was also not captured in the study.

Another study by Nishad *et al.,* 2014 in Sri Lanka, developed red cell distribution width cut-off value of 14.45 that increased the screening sensitivity for Hb E trait. They found out that, in Sri Lanka National screening policy for hemoglobinopathies, a cut-off values MCV > 80 and MCH > 27 values were the most valued in Hb E trait screening neglecting RDW, a strategy that would miss some individuals with HB E trait. However, including RDW improved the diagnosis of cases rather than taking MCV or MCH alone. The study concluded that RDW Cut-off value of 14.45 improved the sensitivity of MCV and MCH from 86.6% to 98.2% in detecting Hb E (EBT). This study demonstrated the importance of using the three parameters: MCV, MCH, and RDW that targeted specifically Hb E trait alone and failed to show the significance of RDW cut-off values in detecting other hemoglobin disorders (Hb SS, HbAS, β-and α-Thalassemia) (Nishad *et al.,* 2014).

So, despite studies showing the significance of RDW in disease detection, no similar study has been done to test its ability on hemoglobin disorders. Therefore, this is the first attempt in Kenya to assess the role of RDW on hemoglobinopathies in a resource poor setting of Western Kenya and indeed the entire continent of Africa.

2.5.1 Use of receiver operating Characteristics (ROC) curves and Youden index in evaluation of clinical utility of Diagnostic tests and biomarkers

The (ROC) curve are used to evaluate how effective a certain biomarker or test is in discriminating diseased and non-diseased populations (Ruopp *et al.,* 2008, Habibzadeh *et al.,* 2016, Kallner 2018). A ROC curve shows how clinical sensitivity is related to specificity at various possible cut-off points of an indicator or biomarker for a given physiological/biological state. The best cut-off value has the highest sensitivity and specificity, giving the lowest false positive rate (1-specifity), the maximum point at the left upper corner of the ROC curve. The Youden Index (J) is the measure of the overall diagnostic accuracy or other, ability of a biomarker or test to differentiate diseased from disease-free populations (Parikh *et al.,* 2007). It is estimated as the area under the ROC curve (AUROC) of a test and is the efficacy of the test to differentiate true from untrue value (accuracy). The greater the area under the curve, the greater the accuracy of the test. ROC curve is a plot of Sensitivity versus (1-Specifity) at all possible cut-points (c). i.e. The x-axis showing 1 – specificity and the y-axis showing sensitivity (= true positive fraction = TP/ (TP+FN)).

A test to be perfect, it needs to differentiate diseased from disease-free individuals with a sensitivity of 100% and a specifity of 100%, however, such tests do not exist in reality because attainment of one is achieved at the expense of the other, in clinical practice. A high sensitivity means being able to pick as many positives as, possible and increases the risk of including false positives among true positives, which reflects low *specificity.* On the other hand, high *specificity* for, say a cut-off point or biomarker, will demand a very narrow confidence interval and thus low statistical confidence level so as to include

among the typical bearers of the given state only the true ones. This quest for high specificity comes at the risk of missing some individuals with latent or trace levels of the concerned state, meaning having low sensitivity. The desirable levels of sensitivity and specifity of a test in a ROC curve is obtained at the point (optimal point) nearest to the top left-hand corner which is the most accurate in diagnosis of a given disease (Hoo et al., 2017). A worthless test in clinical diagnosis, has a ROC curve that flows along the diagonal line and has both sensitivity and specifity of about 50% and a Youden index of 0.5 (Ruopp M. D *et al.*, 2008, Lalkhen & McCluskey 2008, Parikh *et al.*, 2017, Kallner 2018).

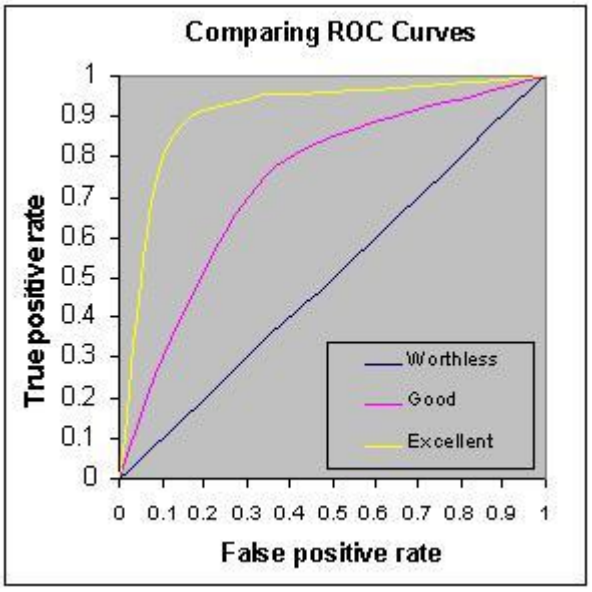


Figure 2.1 Comparing Roc Curves

Figure 2.1 shows three ROC curves representing excellent (yellow curve), good (pink), and worthless tests (blue curve) plotted on the same graph. The ROC curve flowing at the left upper corner, the more the test is accurate. The more the ROC curve flows along

the diagonal line, the more useless the test is in clinical practice whose sensitivity and specifity is about 50% and usually has a Youden index of 0.5 (Ruopp *et al.,* 2008).

2.6 Summary and Conclusions

This chapter has reviewed previous studies of relevance to the topic for this study. The RDW was shown to be markedly elevated during sickling crisis in adult sickle cell disease (Webster & Castro (1986). The literature revealed that, despite the increasing burden of hemoglobin disorders in African countries, there is paucity of data on haemoglobinopathies. Further from the literature it is possible that besides fundamental genetic variation underlying racial and ethnic differences among populations of different parts of the world, epigenetic differences caused by residence in different geographical environment plus sociocultural variations consequent to residence in these environments tend to impose variation in the expression of the genes for various human biological characteristics. This implies that normal RDW reference cut-off values derived for a population in a particular geographical setting, may not necessarily be generalizable to populations in other physically far-removed geographical environments. However, none of the studies reviewed has evaluated the RDW as a biomarker for detection of haemoglobinopathy in not only in Kenya but in Africa generally, leave alone Western Kenya.

In addition, the literature revealed that no research has been conducted on the prevalence of haemoglobinopathy in wider Western Kenya, served by the Aga Khan Hospital, Kisumu main laboratory. Further, the studies were limited to populations in specific areas of Western Kenya, mainly Kisumu County and areas served by Moi Teaching and Referral Hospital, such as Uasin Gishu mainly and surrounding parts of Kakamega and

Nandi Counties. Though there are broad similarities, still there are variations in the environmental conditions, levels of urbanization and hence possible sociodemographic characteristics of the resident populations of different regions of the wider western Kenya. Consequently, it is not certain that results of these studies are generalizable to the wider western Kenya regions, especially those served by the Aga Khan, Kisumu, Hospital laboratory.

CHAPTER THREE

MATERIAL AND METHODS

3.1 Study Site

The study was conducted in Aga Khan Hospital, Kisumu, and its satellites (Outreach Centers) that cover Western Kenya region within the Lake Victoria basin. The health care facility offers both inpatient and outpatient services with over 100 in-patient bed capacity and 10 ICU bed capacity. This part of western Kenya is a malaria-endemic zone and has a high percentage of hemoglobin disorders especially sickle cell and thalassaemias. The specific parts of the region of the wider Western Kenya where the Aga Khan Hospital, Kisumu, and its satellites are located, is geographically divided into four areas with different climatic conditions which has shown to influence malaria heterogeneity (Matsushita *et al.,* 2019). These regions are: highlands of western Kenya, lowlands of Western Kenya (adjacent to Lake Victoria), Upper midland areas and the upper rift valley regions (Trans-Nzoia) (Matsushita *et al.,* 2019) (*see appendix I*).

The highland areas of western Kenya (Kakamega, Bungoma, Vihiga, and Upper Busia) have an altitude of 1500mm above sea level. The annual rainfall is about 2538mm with a mean average temperature of 19.8^0c that supports farming and forests. The major land use is conversion of forests into crop-farming mainly tea growing, sugarcane, maize farms thus, draining swamps into farms , hence, reduce mosquito bleeding an ecological element that has shown to cause fluctuations of malaria in these areas (Matsushita *et al.,* 2019, Omukunda *et al.,* 2012). This region is characterized by high population density predominated by the Abaluhya ethnic group, with Luo and Teso forming minority ethnic communities (Hassan 2017, Alwy & Schech (2004).

Lowland areas in Western Kenya (Kisumu, Homabay, lower Busia region) stands adjacent on the shores of Lake Victoria. (Rakama 2017), whose altitude ranges from 1100-1200m above sea level and its rainfall annually is about 1200mm (Omukunda, *et al.,* 2013). The climate in these areas is relatives hot and dry (Ojowa *et al.,* 2001) whose maximum annual temperature ranges from 25^0 C to 30^0c (Rakama 2017). This region tends to develop many swamps during the rainy season resulting to positive association between rainfall and malaria in locations near the Lake Victoria (Matsushita *et al.,* 2019). However, Homa Bay covers the lower midland along the Lake Victoria shores to the upper areas bordering Kisii and Nyamira. The rainfall here is bi-modal ranging from 800-1400 mm per annum and much of this land is arable land with marshy, rocky and eroded section that is too steep for cultivation (Ojowi *et al.,* 2001). Geographical reports, documents that the luo ethnic group occupies the major region of lowland areas of western Kenya. (Hassan 2017, Alwy & Schech (2004).

The Upper midland areas of western Kenya include Kisii, Nyamira and Migori with an altitude of 1200-2000 above sea level with the highest altitude being in Kisii county which is characterized by undulating hills whose rainfall is approximately 1500 mm per annum in a bimodal pattern that supports crop production and livestock farming. The major cash crops in Migori area are maize, tea, cassava, and sugarcane (Njoroge *et al.,* 2020, Ojowi *et al.,* 2001). Kisii and Nyamira is predominated by the Abagusii ethnic community while Migori is predominated by a Luo speaking and a significant number of the Abakuria, an ethnic community of bantu extraction **(**Hassan 2017, Alwy & Schech, 2004).

Trans-Nzoia County is in the upper North Rift Valley Region and, though a largely cosmopolitan settlement area. It is dominated by the Abulahya and Kalenjin ethnic communities (Hassan 2017, Alwy & Schech 2004) with some Kikuyu and Abagusii, among others. It is located on latitude 1° 01' north, longitude 35°7.5' east, and altitude 1890 m above sea level whose annual mean day temperature ranges from 17.9 to 19.4°C with an average annual rainfall of 1050 to 1100 mm per year that supports maize production as the major cash crop (Masinde *et al.,* 2011). Studies have shown that geographical differences between these regions of western Kenya play a major role in the spatial heterogeneity of malaria incidences where malaria is consistent in lowlands but fluctuates on highland areas of western Kenya with other reports linking malaria to haemoglobinopathies (Kosiyo *et al.,* 2020; Matsushita *et al.,* 2019). Therefore, the Aga Khan Hospital Kisumu and its satellite centers in Western Kenya was the best catchment area for haemoglobin disorders as they are represented in the four areas of the wider Western Kenya.

3.2 Study Design

This was a cross-sectional hospital-based retrospective comparative study of 488 study participants aged 1 month to 66 years, study comparing RDW in a sample of normal subjects (as the Control or reference population group, consisting of high-performance chromatography (HPLC)-confirmed hemoglobinopathy-free individuals) and those confirmed by same method assay to have hemoglobinopathy (as the Case). HPLC was needed for phenotyping the haemoglobinopathies into HbSS, HbSS-HbF, HbSS-beta thalassemia, HbAS, Hb AS-HbF, HbAS-beta thalassemia, beta thalassemia. HPLC

It was a hospital-based study utilizing an existing database of patients consecutively attending or blood specimens referred for investigation at Aga Khan Hospital Kisumu and its satellites in western Kenya from January 2015 to December 2020.

3.3 Study Population

This was a two-arm comparative study and therefore two populations were included. One population consisted of 241 patients confirmed to have normal Hb phenotype, based on HPLC and values for the various hematological indices based on the respective reference cut-off values in use at the Aga Khan Kisumu hospital laboratory. These served as the control group. The other population consisted of 247 patients confirmed to have various haemoglobinopathies (i.e. abnormal Hb phenotypes), based on HPLC and, correspondingly, presumably abnormal values for the RDW

3.3.1 Inclusion Criteria

All the patients that had HPLC reports and their respective full haemogram reports performed from 2015 to 2020 met eligibility criterion to be included in the study.

3.3.2 Exclusion Criteria

The subjects excluded from the study were those HPLC results that lacked their respective complete blood counts, those confirmed to have leukemia, those transfused in the past three months, those confirmed to have any haematopoietic nutrient deficiency (iron deficiency anemia, cobalamin, or folate deficiency anaemia) or those that were tested for blood cultures and antinuclear antibodies (ANA). Normal Hb phenotypes whose Hb was less than 9.5g/dl for under 5 years, less than 10.5g/dl for less than 12 years and less than 11g/dl in age above 13 years were excluded from the study to tame iron deficiency anaemia and other microcytic anaemia from the control group. The

normal Hb phenotype with MCV >97 ft as also excluded to rule out megaloblastic anemias caused by lack of folate or vitamin B 12. This selection criterion was highlighted by Wacholder *et al* (1992).

3.4 Study Variables

The HPLC and hemogram reports were selected based on age, gender and the site from where the Aga khan health care facility was located in Western Kenya.

3.4.1 Dependent variables

The dependent variable was red cell distribution width (RDW)

3.4.2 Independent variables

Independent variables were the Hb phenotypes, that included normal Hb phenotype (HbAA), provided by the control group and haemoglobinopathies (from the case group). These were the abnormal haemoglobin phenotypes: Sickle cell Hb (Hb AS and HbSS), β-Thalassemia and heterozygous HbS+β thalassemia phenotypes.

3.5 Sampling Design

Study participants were obtained by census. That is, first the Aga Khan Hospital in Kisumu and all its satellites in Western Kenya were included as the source for the participants. Secondly, the participants studied consisted of all the 488 eligible patients investigated at the hospital's Kisumu laboratory for haemoglobinopathy and RDW, as part of the complete blood count (CBC) from January 2015 to December 2020. The 488 participants included a case and control group. The case group consisted of 247 patients confirmed through high performance liquid chromatography (HPLC) to have a haemoglobinopathy and had their RDW values from the automated haematology

analyzer generated as part of the complete blood count (CBC). The control group, on the other hand, was a population consisting of all the 241 patients confirmed, similarly, through HPLC to have the normal haemoglobin phenotype (HbAA) and, similarly, had their RDW values.

Choice to include all the Aga Khan Hospital's outreach centers located in western Kenya, was to ensure the representation of the entire Western Kenya geographical area served by the facility. Since the geographical setting of habitation of a population may provide physical environmental conditions and influence sociocultural practices of populations by virtue of them being bounded together in same physical environment. Further, Kenya is an ethnically diverse country and given geographical areas tend to be inhabited predominantly by people of particular ethnicities. Research reports have reported variation in the same biological characteristic between human populations inhabiting widely diverse geographical locations (Boyce *et al.,* 2020). People bounded together in a given geographical setting tend to share ethnicity, and their shared geophysical environment and sociocultural practices provide epigenetic factors likely to influence the phenotypic expression of the genes that encode the various biological characteristics. It is uncertain therefore, as to whether the findings on population RDW for inhabitants of one Kenyan region are generalizable to all other regions. The inclusion of all the centers, therefore, was meant to ensure, as far as practicable, the representation these population characteristics known likely to influence the red cell distribution width in the study subjects were capture.

Regarding the source for the study subjects the decision to use the patients for the haemoglobinopathy cases and, for the control group presumed haemoglobinopathy-free

patients, investigated at the Aga Khan Hospital's laboratory was because this was logistically and financially more manageable as it enabled easy access to a large number of subjects to make the two study populations. Use of patients presumed free from a given disease as a control group in place of the healthy population in the derivation of reference values for biomarkers of the disease is known as the indirect approach for this purpose, while use of presumably healthy subjects is the direct approach (Yan Ruohua *et al.,* 2022, Farrell and Nguyen 2019). The indirect approach has been demonstrated by many studies to offer clinically valid results, if appropriately executed (Yan Ruohua *et al.*, 2022, Farrell and Nguyen 2019, Arzideh *et al.*, 2021).

3.6 Sample Size Determination

The sample size determination for the present study was calculated using Cochran (1963:75). adapted to a two-sample (two-arm) study design to give sample required to estimate a proportion with 5% (0.05) margin of error, e.

$$N = \frac{z^2 pq}{e2}$$

N - The sample size

Z- The abscissa of the normal curve that cuts off an area α at the tails (1 - α equals the desire confidence level, e.g., 95%). is the value for the selected alpha level, e.g. 1.96 for (0.025 in each tail) a 95 percent confidence level

e - The acceptable sampling error

p- α-thalassemia and sickle cell among children enrolled in a malaria vaccine clinical trial at Kombewa in Lake Victoria basin, western Kenya (Kifude C.M

(2007) reported a sickle cell prevalence of 19%. This prevalence, p value, was adopted for this present study giving p=0.19 and q=1- p=1-0.19; q =0.81. Substituting these in the formula would yield n=237. Thus, each study arm had a sample size of 237, giving a total of 474 study participants in both study arms.

$$N = \frac{z^2 pq}{e2}$$

$$N = \frac{1.96^2 \times 0.19 \times 0.81}{(0.05)^2}$$

$$N = 237 \text{ Study participants}$$

$$\text{Two-Arm study} = 237 \times 2$$

$$= 474 \text{ Study participants}$$

$$\text{The minimum sample size} = 474 \text{ study particpants}$$

3.7 Data Collection and Management

3.7.1 Data collection

Data were obtained from the laboratory database for patients examined at the hospital's haematology laboratory for the past five years, from January 2015 to December 2020. For each eligible subject (both case and control group members), the data obtained from the database were the demographics (gender and age), values for RDW out of the CBC output from the haematology autoanalyzers and then the corresponding haemoglobinopathy status reports. These values of the CBC parameters and Hb phenotypes were obtained from patient specific database that has also all patients'

demographic information and is easily retrievable using patient unique number, as per the laboratory's usual protocol.

3.7.2 Determination of red cell distribution width

Values of red cell distribution width were obtained from the output of an automated blood haemogram analyzer consisting of a complete blood count (CBC). The automated haemogram analyzers employed were various models of the machine (XP 300, SYSMEX XS1000i, SYSMEX XNL 330, KX-21N and SYMEX XS 500i, manufactured by Sysmex Corporations Kobe, Japan) located at different selected stations in Aga Khan Hospital, Kisumu, and its western Kenya satellites centers. To ensure accurate and reliable patients results, equipment calibrations are properly monitored using daily internal quality controls (IQC) and monthly external quality assurance (EQA) therefore, owing to sustained quality, the Aga Khan Hospital, Kisumu laboratory has attained ISO 9001:2015 certification, ISO 15189:2012 accreditation and is at advanced stage of attaining Joint Commission International Accreditation.

3.7.3 Determination of haemoglobin phenotype

Data on the various haemoglobinopathy phenotypes were obtained from haematology laboratory database. These had been determined using high performance liquid chromatography (HPLC) Bio-Rad D10 machines manufactured by Bio-Rad Laboratories (based in Hercules, California United States of America). Internal quality controls are performed before running patient samples and the machines are under stringent preventive maintenance to monitor calibrations thus ensuring reliable tests results.

3.8 Data Analysis

Data were first keyed into Excel program, then cleaned and coded to create a database. They were then imported into and analyzed by use of the (statistical package of social sciences) *SPSS* program, *version 23* (IBM in Chicago)

For the continuous variables RDW and age of participants, the non-parametric approach was used in the analysis of data, as recommended by Nahm (2016). The reason being that, Shapiro-Wilks and Kolmogorov-Smirnov tests revealed that the RDW values for the control group were not normally distributed ($p<0.05$) *(See appendix VII).* Hence, the median and interquartile range (IQR) were used as the summary statistics for the assay RDW values for both the control and case groups. The age of participants was similarly analyzed. On the other hand, frequencies percentages and proportions were the descriptive statistical measures used to summarize the data for nominal variables, like gender. Similarly, the haemoglobinopathy phenotypes profiles, as distributions or prevalence of cases in the various population subgroups based on demographic characteristics in Aga Khan Hospital sites from where blood specimens came, were summarized as frequencies and proportions.

For the same reason of the distribution of the RDW data of non-Gaussian distribution, non-parametric inferential statistical techniques were employed. These included Pearson's Chi square test of association, Mann Whitney U test was used for inference about difference in RDW between two groups, Kruskal-Wallis H test differences across multiple population variables and Dunn's *post hoc* test, within group variations. The Chi-square (χ^2) test was used to determine the variation in the frequency of occurrence of the haemoglobinopathy phenotypes with the various population characteristics,

including the Aga Khan Hospital centers that provided blood specimens/patients and age-groups.

The receiver operating characteristic (ROC) curve was used to evaluate the diagnostic power of RDW as a biomarker for haemoglobinopathies in terms of the sensitivity, specificity and Youden Index (accuracy) of the derived reference limit, including odds ratio (OR), likelihood ratio (LR) and predictive values. The sensitivity of RDW was measured as the proportion of individuals correctly identified by the derived reference cut-off limit or value as having any haemoglobinopathy compared with HPLC. Conversely, specificity of the RDW for haemoglobinopathy was measured as the proportion of individuals correctly identified as haemoglobinopathy-free by the derived reference cut-off limit, compared with HPLC. Statistical inference was judged at the 5% significance level.

3.9 Ethical Considerations

All the Ethical clearance was sought at Masinde Muliro University Ethical Committee, NACOSTI and data collection permit granted by Aga Khan Hospital, Kisumu Ethical Review Committee. Ethical considerations were anchored on the four ethical basic confidentiality principles that include: autonomy, justice, Beneficence and Non-maleficence. To maintain confidentiality of the study subjects, all staff and any other person who were in access to patient information were required to sign a confidentiality log according to hospital policy to prevent any harmful publicity of the patient. To this regard, there was strict use of password proof computers and restricted access to rooms that store patient information only to authorized persons to ensure the four basic ethical principles are maintained as documented by Jahn (2011).

To ensure that the present study complied to the four ethical principles, firstly, the autonomy section was automatically realized since contacting patients was not possible within our context limiting patient choice of participation. However, permission to collect data was sought from the facility management. Secondly, Justice was maintained in that all the patients who met inclusion criteria were sampled equally and included irrespective of gender or ethnicity. Regarding excluded patient, there was enough justification explained to hospital on sample type. Thirdly, to achieve beneficence, the data obtained were used for research purposes with the view of improving patient's early and accurate diagnosis of hemoglobin disorders through evidence-based information to practitioners. It has helped immensely to guide the choice of confirmatory test leading to correct diagnosis and treatment especially advantageous in resource poor setting where screening for population has not been implemented. The finding was to be shared with the facility if they needed it. The results of the study were shared through presentation, printed copies and publications. Finally, the study did not involve direct contact with patients, which eliminated risks of physical and minimized risks of psychological harm to the study participants thus achieving non-maleficence goal. This data was not used regarding place of residence or next of kin.

Compensation

There was no monetary or any other inducement to study participants selected, as the researcher did not meet them or need their consent directly to obtain data about them.

Conflict of Interest

The principal investigator funded this study independently; there was no financial support from anywhere or individuals who might have undue interest in the results. There was therefore no conflict of interest during all the phases of the study.

CHAPTER FOUR

RESULTS

4.1 Introduction

This health-facility based cross-sectional retrospective study investigated the potential of the red cell distribution width to serve as a diagnostic marker for haemoglobinopathy among patients attending the Aga Khan Hospital, Kisumu, and its satellites in western Kenya. The subjects were all the 488 patients of HPLC-confirmed haemoglobin phenotype comprised of 247 haemoglobinopathy victims (as case group) and 241 haemoglobinopathy-free persons (as the control group) seen by the hospital and its western Kenya satellites and investigated by its laboratory in Kisumu from January 2015 to December 2020. The RDW data were obtained from the hospital's Kisumu laboratory database of routine assay results existing as part of the complete blood count (CBC) output from the haematology autoanalysers. The utility of RDW was determined by evaluating the accuracy, predictive validity, and likelihood ratio, in addition to diagnostic sensitivity and specificity of a reference cut-off point derived as the upper limit of 95% confidence interval of the median for the control group. These biomarker or assay method receiver operating characteristics (ROC) for the reference RDW cut-off limit were evaluated against the RDW values for the various HPLC-confirmed haemoglobinopathies in the case group.

4.2 Demographic characteristics and origin stations of the study subjects

The study participants were 488 individuals aged 1 month to 66 years divided into two arms consisting of controls and cases. The controls were 241 (49.4% of the total) individuals, consisting of 102 (20.9%) males and 139 (28.5%) females. The cases were

247 (50.6% of total), consisting of 112 (22.9%) males and 135 (27.6%) females respectively, p=0.786). The vast majority of the participants were drawn from Kisumu station (49.0%, n=239), followed by Busia (15.4%, n=75) and then Homa Bay (12.3%, n=60) (p<0.0001). The other stations shared rest of haemoglobinopathies in small proportions of less than 5%. Haemoglobinopathies proportions varied significantly in the three age groups: 0-5years, 6-12 years and 13 above (42.4%, n=207; 19.9%, n=97; 37.7%, n=184 respectively, (p<0.0001). These results are summarized in *Table 4.1*

Table 4.1. Sample distribution of study subjects based on demographic characteristics

		Study Groups Frequencies (N=488)		***P***
Characteristics		CONTROL (n=241) 49.4%	CASE5 (n=247) 50.6%	0.786
STATIONS	Busia (n=75) 15.4%	22 (4.51%)	53 (10.9%)	
	Bungoma (n=42) 8.6%	28 (5.74%)	14 (2.9%)	
	Kitale (n=17) 3.5%	7 (1.43%)	10 (2.1%)	
	Kakamega (n=20) 4.1%	10 (2.1%)	10 (2.1%)	
	Kisumu (n=239)49%	137(28.1%)	102 (20.9%)	
	Kisii (n=24) 4.9%	14 (2.9%)	10 (2.1%)	
	Homabay (n=60)12.3%	22 (4.51%)	38 (7.8)	**<0.0001**
	Migori (n=11) 2.3%	1 (0.21%)	10 (2.1%)	
Gender	Male (n=214) 43.9%	N=102 (20.9%)	N=112 (22.9%)	0.502
	Female (n= 274) 56.1%	N=139 (28.5%)	N=135 (27.6%)	
Age	<5 Years (n= 207) 42.4%	N=82 (16.8%)	N=125 (25.6%)	
	6-12 Years (n=97)19.9%	N=37 (7.58%)	N=60 (12.3%)	**<0.0001**
	>12 Years (n=184)37.7%	N=122 (25%)	N=62 (12.7%)	

Note: *P*-values in bold define a statistically significant difference†, This table shows distribution of study subjects based on demographic characteristics in terms of frequencies and proportions.

4.3 Distribution of haemoglobinopathy phenotypes in Western Kenya

4.3.1 Overall haemoglobinopathy phenotype profile

Of the total of 247 haemoglobinopathy cases, 7 different haemoglobinopathy phenotypes were detected. They were homozygous sickle cell disease (SCD) Hb(HbSS), 18.2%, n=45; HbSS with fetal Hb (HbSS+HbF), 8.1%, n=20; and HbSS with β-thalassemia (HbSS+β-Thal), 25.1%, n=62. On the other hand, 41.7% (n=103) of the individuals had sickle cell trait (SCT) haemoglobin (HbAS); 2.4% (n=6) had HbAS+β-thal, while the lowest proportion, 8% (n=20) had HbAS+HbF. Similarly, 3.6%(n=9) of the participants had the homozygous β-thalassemia (β-thal) phenotype. These results are summarized in *Table 4.2* and bar graph *Figure 4.1*

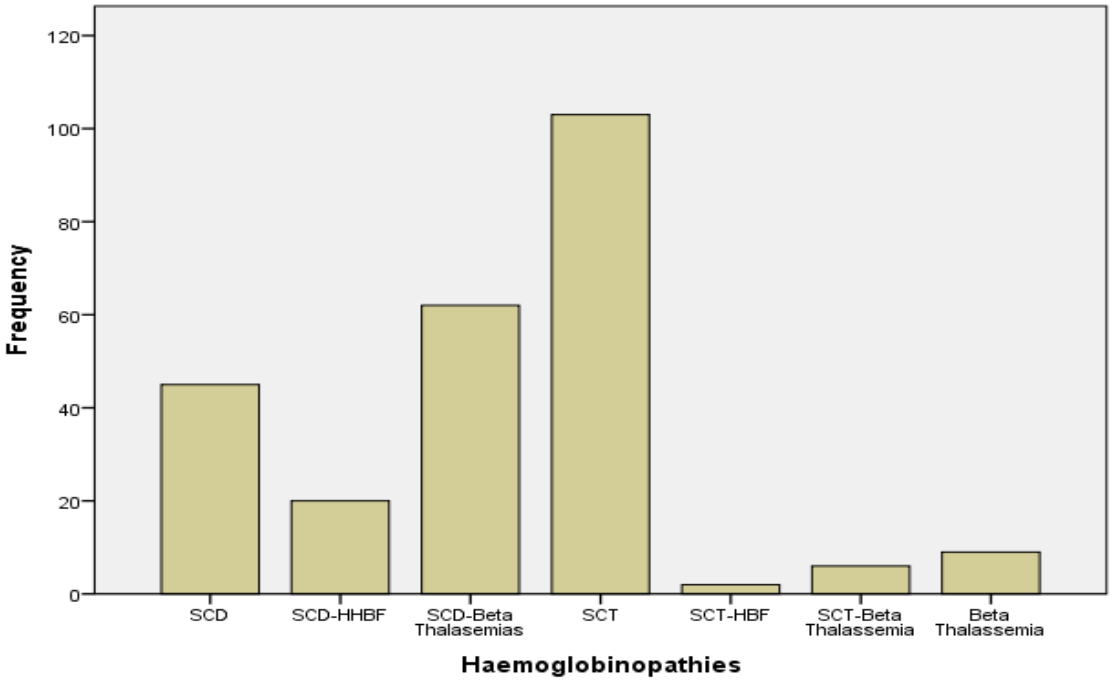


Figure 4.1 Distribution of haemoglobinopathies in Aga khan Hospital, Kisumu, and its Western Kenya satellite centers. SCD, sickle cell disease, SCD-HbF, sickle cell disease with hemoglobin F, SCD-beta thalassemia, sickle cell disease with beta thalassemia, SCT, sickle cell trait, SCT-HbF, sickle cell trait with hemoglobin F

4.3.2 Haemoglobinopathy phenotype profile in Aga Khan Hospital, Kisumu, and it Satellites in western Kenya

The satellites within the Western Kenya counties closest to Lake Victoria (Kisumu, Busia and Homa Bay) had the highest overall proportions of haemoglobinopathies of 41.3% (n=102), 21.5% (n=53) and 15.4% (n=38) respectively. The rest of the counties had low proportions of haemoglobinopathy, with Bungoma having 5.7% (n=14) while Kitale, Kakamega, Kisii and Migori had 4% (n=10) each. (*Table 4.2 and figure 4.2).*

In terms of specific haemoglobinopathies, Busia had 10.9% (n=27) for HbAS; 2.8% (n=7), HbSS; HbSS+β-Thal, 4.5% (n=11); HbSS+HbF,1.6% (n=4); <1%, HbAS+HbF and HbAS+β-thal, each; and 0.4% (n=1), β-thal. In Bungoma, the HbSS+β-thal phenotype was the commonest, at 2.8% (n=7), followed by HbAS, 2.0% (n=5); HbSS and HbSS+HbF, 0.4% each. Kitale center, serving Trans Nzoia County, accounted for less 2% of all the haemoglobinopathy cases. Of these, HbSS+β-Thal and HbSS+HbF were the commonest, each accounting for 1.2% (n=3); while HbSS and HbAS each accounted for 0.8% (n=2). Kakamega County had 3 haemoglobinopathy phenotypes: HbSS, 1.6%(n=4); HbSS+HbF, 1.2%(n=3); and HbAS 1.2% (n=3).

The Kisumu site, being the center for the Aga Khan Hospital's network of health facilities in Western Kenya, contributed the vast majority of the study subjects, that had the highest proportion of the haemoglobinopathies by phenotype diversity as well as the overall total number. It accounted for 41.3% (n=102) out of the 247 comprising the haemoglobinopathy case group for this study. Those with HbSS+β-thal haemoglobinopathy phenotype made the highest proportion, 12.6% (n=31) of these. These were followed by HbAS phenotype, 12.1% (n=30); HbSS, 10.1% (n=25);

HbSS+HbF, 2.8% (n=7); β-Thal, 2.4% (n=6); and HbAS+β-Thal, 1.2% (n=3). It is noteworthy that the proportion of Beta-thalassaemia (β-Thal), 2.4% (n=6), in the participants drawn from this station was the highest out of all the case group for this study.

From this part of Western Kenya, after Kisumu, the Homa Bay County site showed third highest diversity and overall proportion of haemoglobinopathies. Of these HbAS made the highest proportion, 9.7% (n=24) and it was followed by HbSS+β-Thal, 2.4%(n=6); HbSS, 1.6%(n=4); HbAS+β-thal, 0.8% (n=2); while HbSS+HbF and β-thal each made a proportion of 0.4%(n=1). Among the cases drawn from Kisii and Migori County sites each had only 3 different haemoglobinopathies out of the total of 7 reported for the totality of 247 cases included in the study. Kisii had HbAS accounting for 2% (n=5), followed by HbSS+β-thal, 1.6% (n=4); and HbSS+HbF comprising the lowest proportion, 0.4%(n=1). Migori, on other hand reported HbAS, of 2.8% (n=7); HbSS, 0.8% (n=2); and β-thal, 0.4% (n=1). These results are summarized in *table 4.2* and *figure 4.2*

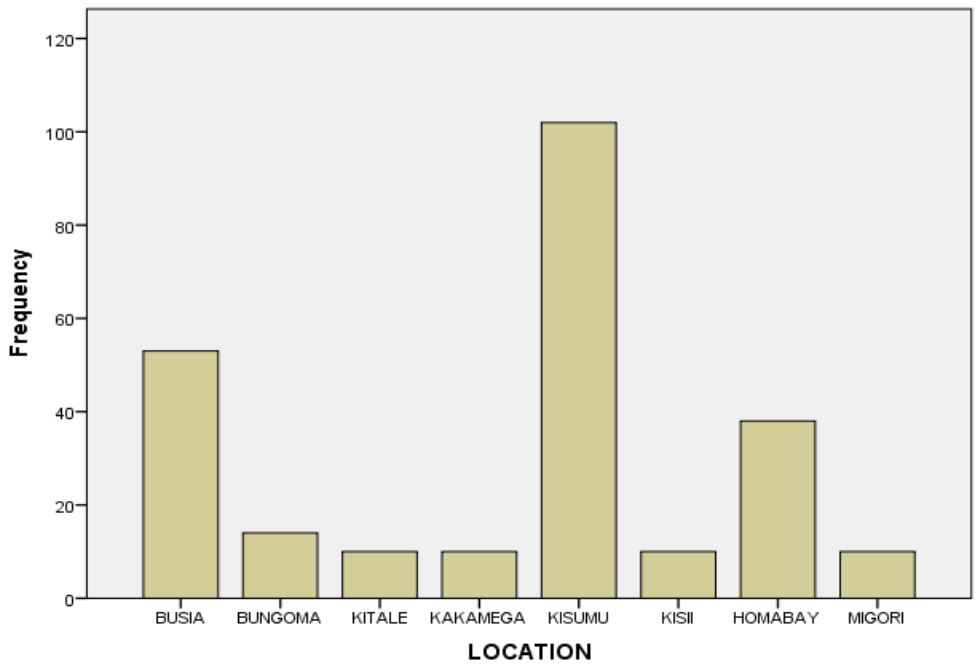


Figure 4.2 Distribution of haemoglobinopathies in Aga khan Hospital, Kisumu, and its Western Kenya satellite centers (Busia, Bungoma, Kitale, Kisii, Homabay and Migori)

4.3.3 Distribution of haemoglobinopathies proportions based on gender and age

It was important to note that generally, HbAS formed the highest proportion, 41.7% (n=103), followed by HbSS+β thalassaemia, with 25.1% (62); HbSS, 18.2%(n=45); HbSS, 8.1% (n=20); while HbSS+HbF and the rest of the other phenotypes were distributed in small proportions of less than 4% each. (*Table 4.2* and *Figure 4.1).* Similar trend was observed in gender and across the age groups and in the three regions that included: Kisumu, Busia and Homabay regions. There was no statistically significant variation (p=0.293) in proportion of hemoglobinopathies with gender (*Table 4.2),* though, females recorded a higher proportion, 54.7% (n=135) as shown on the pie chart *Figure 4.3.* The total of female cases of haemoglobinopathy with HbAS dominated with a proportion of 22.3%(n=55), followed by HbSS+β thalassemia, 13.8% (n=34);

HbSS, 10.1%(n=25); HbAS+HbF 6.1%(n=15); and the remaining 2% (n=3), as shown on the *Table 4.2*. Similarly, in males HbAS made the highest proportion of the haemoglobinopathies, 19.4% (n=48). This was followed by HbSS+β-thalassemia, 11.3% (n=28); HbSS, 8.1%(n=20); HbSS+HbF, 2% (n=5); while the rest recorded about 5% of the 247 cases in the study.

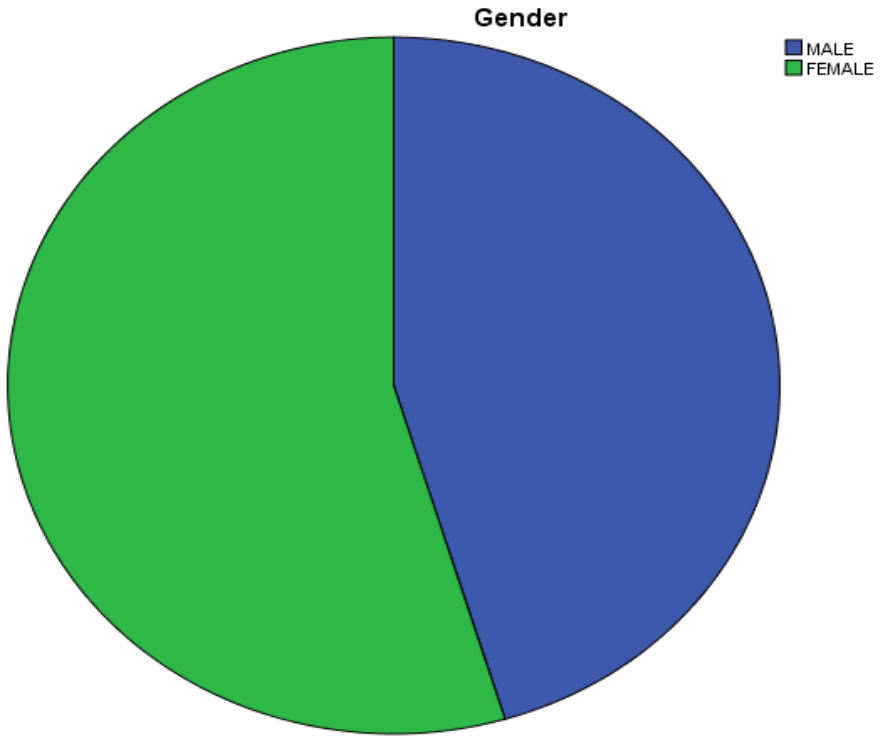


Figure 4.3 Proportions of haemoglobinopathies in gender

There was no statistically significant variation (p=0.202) in the proportions of the various haemoglobinopathy phenotypes across the three age groups, but children under 5 years formed the highest proportions, 50.6% (n=125); 6-12 years, 24.3%`(n=60); and those above 12 years, 25.1%(n=62).

Regarding the individual phenotypes, the highest proportion of subjects under the age of 5 years had HbAS, 21.5% (n=53); followed by 13.4%(n=33), HbSS+β thalassemia; 6.9% (n=17), HbSS, 4.9% (n=12); and 4%, HbSS+HbF. ß-thal among the age set of 6-12 years, the highest proportion, 7.7% (n=19), had HbSS+β-thalassemia, followed by 7.3% (n=18), HbAS; 6.5% (n=16), HbSS 2.0% (n=5), HbSS+HbF and then 0.8% (n=2), ß-thalassemia. For the study subjects >12 years,13.0% (n=32) had HbAS this being the dominant phenotype in the group, followed by those with HbSS, 4.9% (n=12), HbSS+β-thalassemia, 4.0% (n=10); HbSS+HbF 1.2% (n=3) and finally β-thalassemia, 1.2% (n=3). All these results are summarized in *Table 4.2,* and *Figure 4.4.*

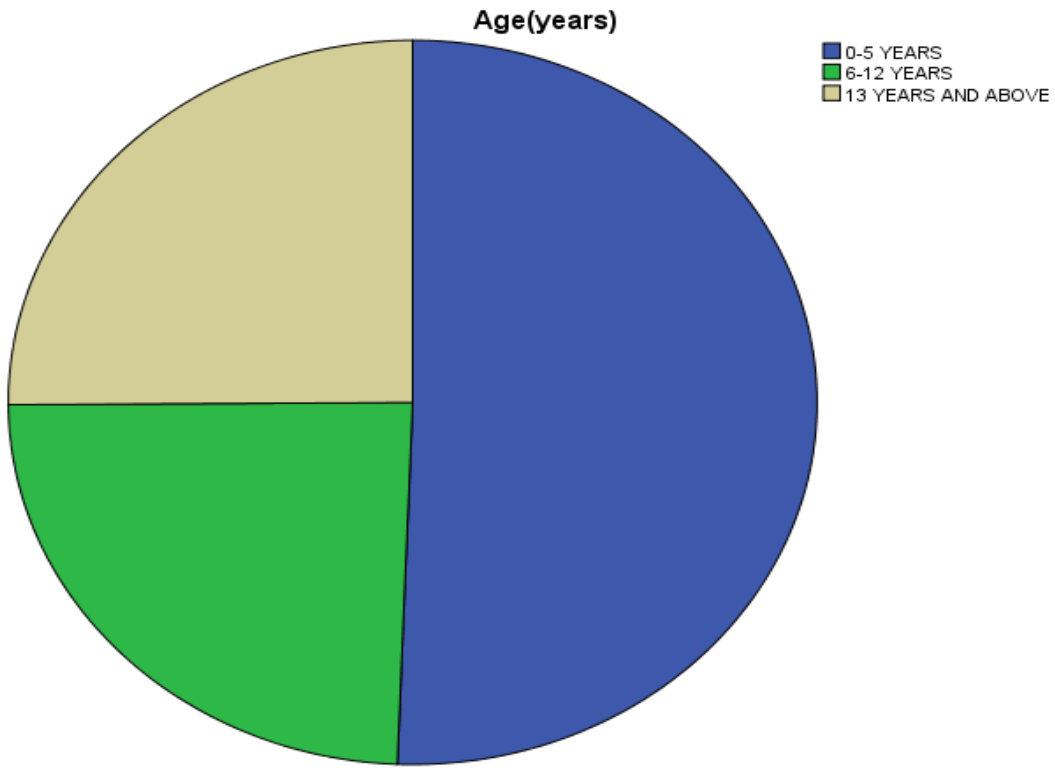


Figure 4.4. Distribution of haemoglobinopathies in Age groups (years)

Table 4.2 Distribution of haemoglobinopathies in western Kenya patients investigated at Kisumu Aga Khan Hospital

		Haemoglobinopathies								
	Hospital Sites	HbSS	HbSS+HbF	HbSS+ß- thal	HbAS	HbAS+HbF	HbAS+ß-thal	ß- thal	Total	p
	BUSIA	7 (2.8%)	4(1.6%)	11 (4.5%)	27 (10.9%)	2 (0.8%)	1(0.4%)	1 (0.4%)	53 (21.5%	
	BUNGOMA	1 (0.4%)	1 (0.4%)	7 (2.8%)	5 (2.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	14 (5.7%)	
	KITALE	2 (0.8%)	3 (1.2%)	3 (1.2%)	2 (0.8%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	10 (4.0%)	
	KAKAMEGA	4 (1.6%)	3 (1.2%)	0 (0.0%)	3 (1.2%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	10 (4%)	**0.017**
	KISUMU	25 (10.1%)	7 (2.8%)	31 (12.6%)	30 (12.1%)	0 (0.0%)	3 (1.2%)	6 (2.4%)	102 (41.3%)	
	KISII	0 (0.0%)	1 (0.4%)	4 (1.6%)	5 (2.0%)	0 (0.0%)	0(0.0%)	0(0.0%)	10 (4.0%0	
	HOMABAY	4 (1. 6%)	1 (0.4%)	6 (2.4%)	24 (9.7%)	0 (0.0%)	2 (0.8%)	1 (0.4%)	38 (15.4%)	
	MIGORI	2(0.8%)	0 (0.00%)	0 (0.0%)	7 (2.8%)	0 (0.0%)	0(0.0%)	1 (0.4%)	10 (4.0%)	
Gender	MALE	20 (8.1%)	5 (2.0%)	28 (11.3%)	48 (19.4%)	1 (0.4%)	5 (2.0%)	5 (2.0%)	112 (45.3%)	0.293
	FEMALE	25 (10.1%)	15 (6.1%)	34 (13.8%)	55 (22.3%)	1 (0.4%)	1 (0.4%)	4 (1.6%)	135 (54.7%)	
Age	0-5 YEARS	17 (6.9%)	12 (4.9%)	33 (13.4%)	53 (21.5%)	2 (0.8%)	4 (1.6%)	4 (1.6%)	125 (50.6%)	
(years)	6-12 YEARS	16 (6.5%)	5 (2.0%)	19 (7.7%)	18 (7.3%)	0 (0.0%)	0 (0.0%)	2 (0.8%)	60 (24.3%)	0.202
	>12 Years	12 (4.9%)	3 (1.2%)	10 (4.0%)	32 (13.0%)	0 (0.0%)	2 (0.8%)	3 (1.2%)	62 (25.1%)	
Combined haemoglobinopathies		45 (18.2%)	20 (8.1%)	62 (25.1%)	103 (41.7%)	2 (0.8%)	6 (2.4%)	9 (3.6%)	247 (100%)	**<0.0001**

Note: P-values in bold define a statistically significant difference†, This table shows frequencies and proportions of haemoglobinopathies distributions based on site, age and gender'

4.4 Red Cell Distribution Width among study groups and participants' demographics

The overall median RDW for control group was 14.5, IQR=2.7; 95%CI=9.1-19.9, giving the reference cut-off limit of 19.9 for interpreting values for haemoglobinopathy phenotypes, and was significantly lower than the overall for combined haemoglobinopathies, 20.7(8.3), p<0.0001. Those with HbSS had the highest median RDW, 25.4 (5.5), followed by HbAS+HbF, 24.2 (7.4); HbSS+Hbß-thal, 23.3 (7.9); HbSS+HbF, 20.9 (5.5); HbAS+Hbß-thal 21.0 (10.5); ßeta thalassemia, 18.1 (8.3); and, lastly, HbAS, 16.4 (6.5) (*Table 4.3*).

The overall RDW for the study subjects was not significantly different between males, median 17.1(IQR=8.0) and females, median 15.5(IQR=6.7), p=0.317. Similarly, for the control group there was no significant difference in median RDW between males, 14.55(IQR=1.90) and females,14.2(IQR=2.40), p=0.089. Similar trend followed in case group with males having RDW median of 21.2 (7.1) while females had median of 20.1(8.8), p=0.056.). However, there was variation in RDW with age for both the control group and cases. Thus, age ≤5 years had the highest median, 19.0 (IQR=7.1) followed by a median of 15.5(IQR=8.2) of 6–12-year-olds, while age ≥12-years had the lowest median of 14.1 (IQR=2.8), p<0.05**.** For the control group again the median RDW generally varied significantly with age: ≤5-year-olds had the highest, 15.95 (IQR=4.2) compared with 6–12-year-olds, 14.2 (IQR=1.7) and ≥12-year-olds, 13.7 (IQR=1.8), **(p<0.0001).** There was no statistical significance (p=0.347) of RDW median of age group 6-12 years when compared to the age set above 12 years in their respective control groups (*See Table 4.3 for details).*

The case group of age set ≥12 years had a median of 16.5 (IQR=2.75) which was significantly lower ($p<0.0001$) when compared to children's cases ≤5years with a median of 21.2 (IQR=7.10) (*Table 4.3).* The age group of 6-12 years had RDW median of 19.9 (IQR=7.90) in their case group and did not have statistical significance ($p=0.138$) when compared to the case group of children ≤5 years. Similarly, there was no significant difference ($p=0.225$) of RDW in the case group of age set 6-12 years when compared to RDW of case group in those individuals ≥12 years *(Table 4.3*).

Table 4.3. Red Cell Distribution Width by study groups, demographics and haemoglobinopathies

RDW-CV (%)	Overall(N=488)			Control (n=241)		Case (n=247)	
		Median (IQR)	P	Median (IQR)	P	Median (IQR)	p
Demographics				14.5 (2.7)		20.7 (8.3)	**<0.0001**
Gender	Male	17.1(8.0)	0.317	14.55(1.9)	0.089	21.2(7.50)	0.056
	Female	15.5(6.7)		14.2(2.4)		20.1(8.80)	
Age	0-5 Years	19.0 (7.1)	**0.001**	15.95(4.2)	**0.001**	21.2(7.10)	0.138
	6-12 years	15.5(8.2)		14.2(1.7)		19.9(7.90)	
	0-5 Years	19.0 (7.1)	**<0.0001**	15.95(4.2)	**<0.0001**	21.2(7.10)	**<0.0001**
	>12 Years	14.1 (2.8)		13.7(1.8)		16.5(2.75)	
	6-12 Years	15.5(8.2)	**<0.0001**	14.2(1.7)	0.347	19.9(7.90)	0.225
	>12 Years	14.1 (2.8)		13.70(1.8)		16.5(2.75)	
Hb Phenotypes	HbAA			14.5(2.7)			
	HbSS					25.4 (5.5)	**<0.0001**
	Hb SS+HbF					20.9 (5.5)	**<0.0001**
	Hb SS+ß-thalassemia					23.3 (7.9)	**<0.0001**
	HbAS)					16.4 (6.5)	**<0.0001**
	HbAS+HbF					24.2 (7.4)	0.449
	HbAS+ß-thalassemia					20.9 (10.5)	0.791
	ß-thalassemia					18.1 (8.3)	1.00

Note: *P*-values in bold define a statistically significant difference; IQR, interquartile range; SS, sickle cell †, This table shows demographic characteristics of the study participants and red cell distribution width in control and case (haemoglobinopathies) groups with their respective statistical significance

4.4.1 RDW comparison by age between control and case groups for males

The red cell distribution width showed statistical significance (p=0.042**)** in control group of males ≤ 5 years who had RDW median of 16.75 (IQR=4.7) when compared to 6-12 years males control group with a median of 15.1 (IQR=1.6) as shown on *Table 4.4*. Similarly, there was significant difference (p<0.0001**)** in RDW for males control group ≤ 5 years when compared to males control group age group ≥12 years who had a median of 13.7 (IQR=1.8). However, there was no statistical significance (p=0.137) in RDW of males aged group 6-12 years when compared to the same males control group ≥12years. The RDW in the case group males between the three age sets did not have statistical significance (p=0.145). These results are summarized in *table 4.4*

4.4.2 RDW comparison by age between control and case groups for females

The red cell distribution width significance in the three age sets in females followed the same trend as males with female age group ≤5years who had a median of 15.55 (IQR=3.76) showed significant difference (p=0.017) when compared to control group females of age 6-12years with median of 13.8 (IQR=1.81) as tabulated on *table 4.4*. There was also significant difference (p<0.0001) in RDW of female control group ≤5years when compared to age ≥12 years female control group who had a median of 13.75 (IQR=2.22). However, there was no significant difference (p=0.860) in RDW of 6-12 years female control group when compared to the age group ≥12 years female control group. The female case group ≥12 years had the lowest median of 14.75 (IQR=7.34) with statistical significance (p <0.0001**)** when compared to ≤5 years and also when compared to the age set 6-12 years (p=0.024), as shown on *Table 4.4.*

Table 4.4 RDW and demographic characteristics

RDW-CV (%)		Medians (Interquartile Ranges)			
		Control (n=241)	p	Cases(n=247)	P
MALE (n=214)	≤5years	16.75 (4.7)		21.75 (6.54)	
	6-12 years	15.1(1.6)	**0.042**	20.25 (7.96)	
	≤5 years	16.75 (4.7)		21.75 (6.54)	0.145
	≥12 years	13.70(1.8)	**<0.0001**	20.1 (8.36)	
	6-12years	15.1 (1.6)		20.25 (7.96)	
	≥12years	13.7 (1.8)	0.137	20.1 (8.36)	
FEMALES (n=274)	≤5years	15.55 (3.76)		21.0 (7.04)	
	6-12yeas	13.8 (1.81)	**0.017**	19.65 (8.37)	0.732
	≤5 years	15.55(3.76)		21.0 (7.04)	
	≥12 years	13.75 (2.22)	**<0.0001**	14.75 (7.34)	**<0.0001**
	6-12years	13.8 (1.81)		19.65 (8.37)	
	≥12years	13.75 (2.22)	0.860	14.75 (7.34)	**0.024**

Note: *P*-values in bold define a statistically significant difference; RDW-CV; red cell distribution width coefficient of variation

†, This table shows red cell distribution width stratification based on study groups, age and gender

4.5 Sensitivity, specificity, likelihood ratio and accuracy (Youden index) of RDW as a surrogate marker for haemoglobinopathy

The utility of RDW as a surrogate marker for haemoglobinopathy assessed on the basis of the accuracy (in terms of *Youden index*), sensitivity, specificity and predictive values (Positive and Negative) obtained for each phenotype through analysis of respective receiver operating characteristics (ROC) curves for the cut-off (reference) limit, 19.9%, from the control group data. The predicative values, likelihood ratio and odds ratio, accuracy (as represented by the value of the *Youden index*) were then used in the determination ultimately of the RDW(*Table4.5*).

Table 4.5. Receiver operating characteristics (ROC) for the RDW reference limit from the area under curve (AUC) for the various haemoglobinopathies

Hb Phenotypes	Area	Asymptotic Sig.(p)	RDW Cut-off (95% CI Upper Limit)	Sensitivity	Specificity	LR+	LR-
Normal Hb (HbAA)	.178	**.000**	19.1	-	-	-	-
HbSS	.892	**.000**	19.9	88.9	75.2	3.59	0.148
HbSS+HbF	.766	**.000**	19.9	60.0	70.5	2.03	0.57
HbSS+β thal	.805	**.000**	19.9	71	75.1	2.85	0.39
HbAS	.501	.976	19.9	30.1	69.1	0.97	1.01
HbAS+β- thal	.600	.399	19.9	50	69.5	1.64	0.72
β-thalassemia	.542	.662	19.9	44.4	69.6	1.46	0.79
Haemoglobinopathies combined	.746	**.000**	19.9	55.1	94.2	9.5	0.476

Note: Hb; hemoglobin, RDW; red cell distribution width, LR+; positive likelihood ratio, LR-; negative likelihood ratio, AA; normal hemoglobin, SS; sickle cell disease, AS; sickle cell trait, β-thal; beta thalassemia, HbF; hemoglobin F.
†, This table gives a summary of red cell distribution width likelihood ration in terms of Youden index/Accuracy (area under the curve), asymptotic significance (*p*), sensitivity (%) and specifity (%) at a cut-off value of 19.9 in a ROC curve.

4.5.1 Red Cell Distribution Width ROC Curve in Normal Control Group (HbAA)

The proportion of the control group comprised of 49.4% (n=241) and having no significant difference from case group =247 proportion of 50.6% (n), p=0.786 (*See* as shown on *Table 4.1).* Red cell distribution width for control group produced a significantly (P<0.0001**)** low sensitivity and specificity where its ROC curves fell below the diagonal line (*Figure 4.5*) confirmed by reduced Youden index of 0.178 as indicated on *Table 4.5.*

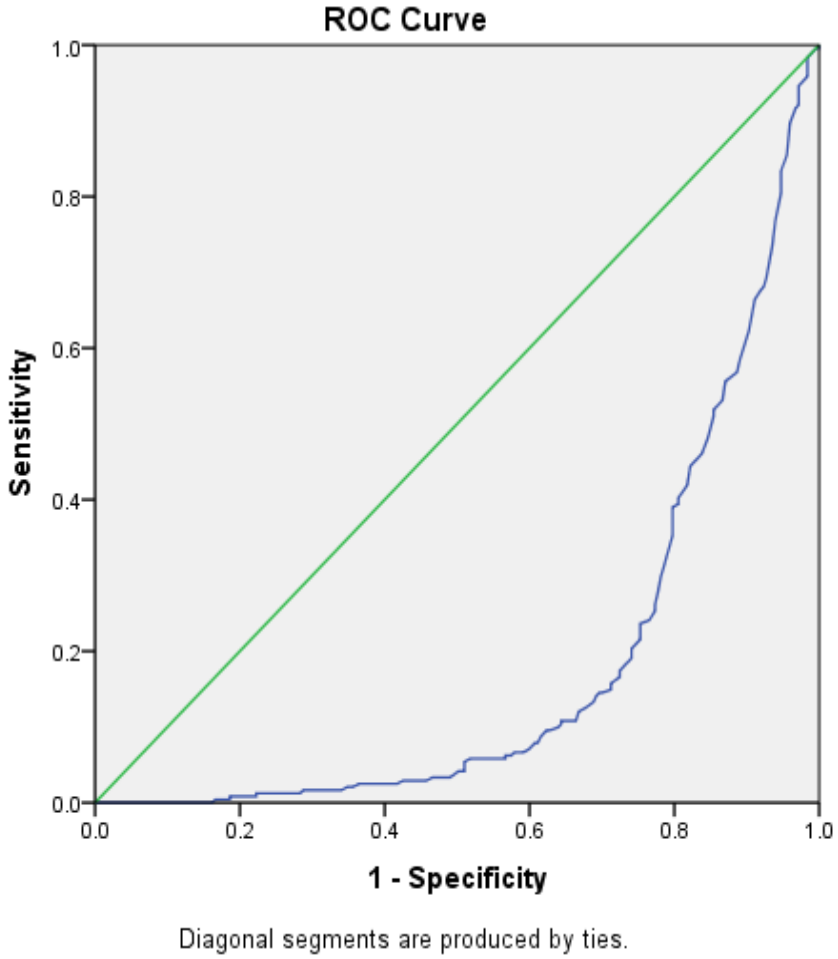


Figure 4.5. Red cell distribution width ROC curve for control group

4.5.2 Red Cell Distribution Width ROC Curves for HbAS

RDW ROC curve flowed along the diagonal line (Figure 4.6) with a Youden Index of 0.501 having no significance in diagnosis of HbAS (p=0.976) as shown on *table 4.5.* At 19.9 standard cut-off value, it had low sensitivity of 30.1% and a specificity of 69.1% whose positive likelihood ratio of 0.97 and negative likelihood ratio of 1.01 was not able to distinguish HbAS diseased from non- diseased population.

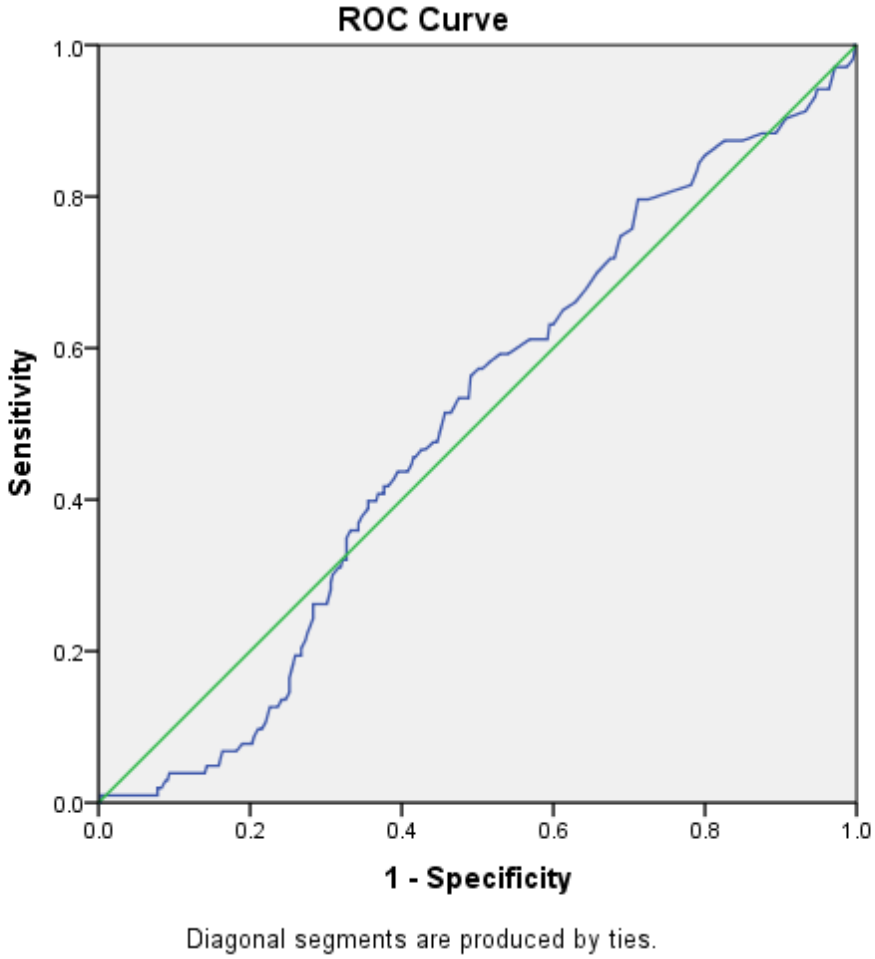


Figure 4.6 Red cell distribution width ROC curve for HbAS

4.5.3 Red Cell Distribution Width in HbAS+Beta Thalassemia

Red cell distribution width ROC curve coordinates at a cut-off point of 19.9, did not have statistical significance (p=0.399) in the diagnosis of HbAS+beta thalassemia phenotype *(Table 4.5).* This was affirmed by a low sensitivity of 50% and a specifity of 69.5%. Its ROC curve flowed along diagonal line (Figure 4.7) with a Youden index of 0.600, positive likelihood ratio of 1.64 and 0.72 negative likelihood ratio as shown on *Table 4.5*.

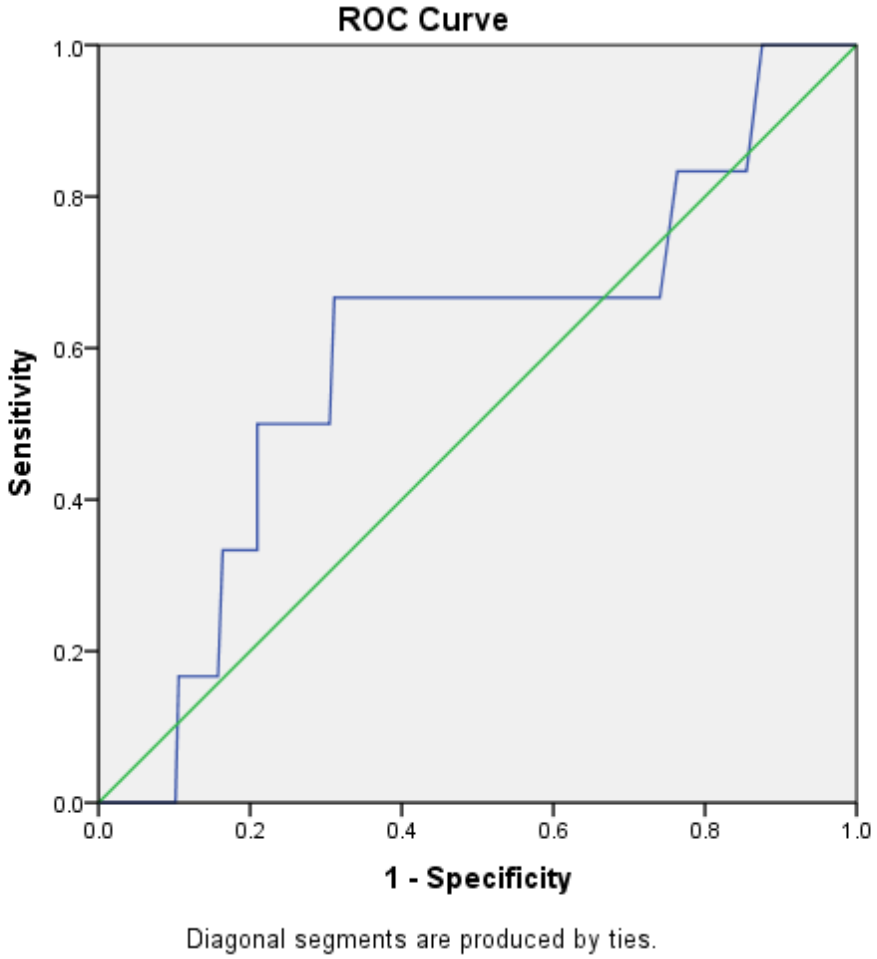


Figure 4.7. Red cell distribution width ROC curve for HbAS+Beta-thalassaemia

4.5.4 Red cell distribution width ROC curve for HbSS

The RDW, at a cut-off limit of 19.9%, had an accuracy (*Youden Index*) of 0.892 (89.2%), sensitivity of 88.9% and a specificity of 75.2%, together with positive and negative likelihood ratio of 3.59 and 0.148, respectively *(Table 4.5 and* Figure 4.8*).* Together, these fairly high values indicate that RDW at the set cut-off limit of 19.9 was potentially a useful biomarker for distinguish individuals with HbSS haemoglobinopathy from those without it (See *Table 4.5* and *Figure 4.8*)

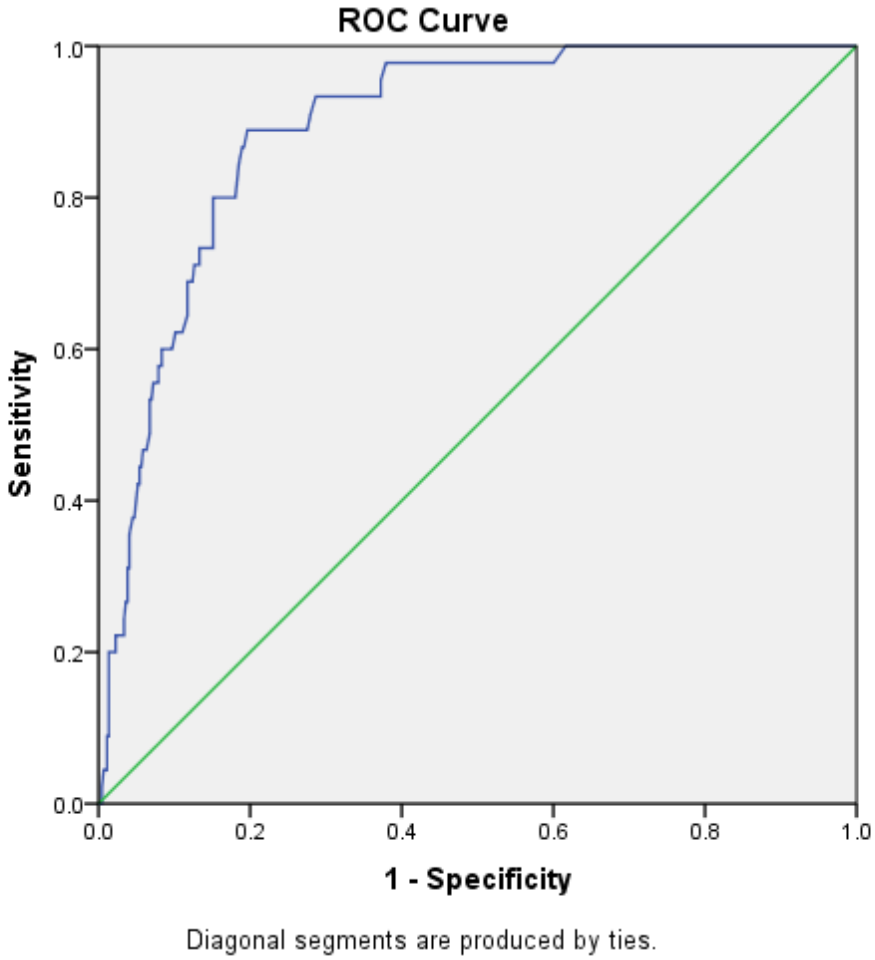


Figure **4.8.** Red cell distribution width ROC curve for HbSS haemoglobinopathy

4.5.5 Red Cell Distribution Width ROC Curve for HbSS+HbF

HbSS+HbF RDW, at cut-off point of 19.9% had a sensitivity, specificity, accuracy of 60%, 70.5% and 0.766 (i.e 76.6%), respectively for HbSS+HbF haemoglobinopathy *(Table 4.5)*. It had a positive likelihood ratio of 2.03 and a negative likelihood ratio of 0.57 (*Figure 4.9*). Considered together, these results suggest modest utility of RDW as a surrogate marker for the detection of HbSS+HbF haemoglobinopathy.

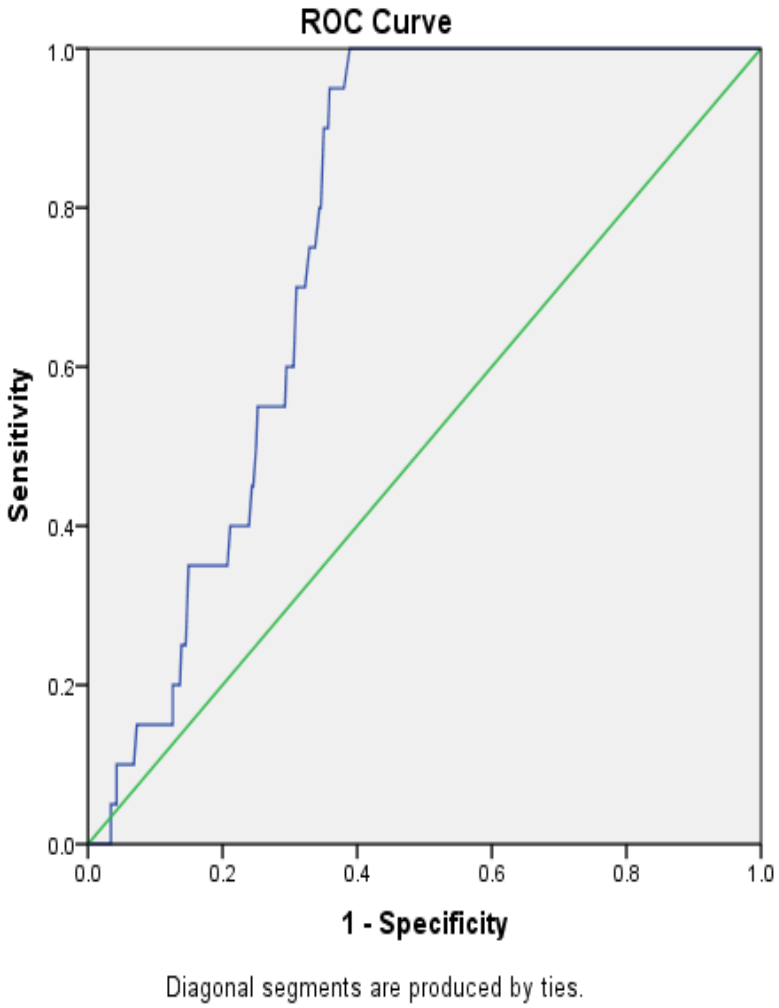


Figure 4.9. Red cell distribution width ROC curve for HbSS+HbF phenotype

4.5.6 Red Cell Distribution Width ROC Curve for HbSS+ßeta-thalassaemia

At a cut-off point of 19.9%, RDW had a sensitivity of 71%, specificity of 75.1 and an accuracy of 0.805 (i.e, 80.5%) for HbSS+ßeta-thalassaemia with a positive likelihood ratio of 2.85, and negative likelihood ratio of 0.39, implying it could be a moderately useful biomarker for both screening and diagnosis of HbSS+ßeta-thalassaemia haemoglobinopathy.

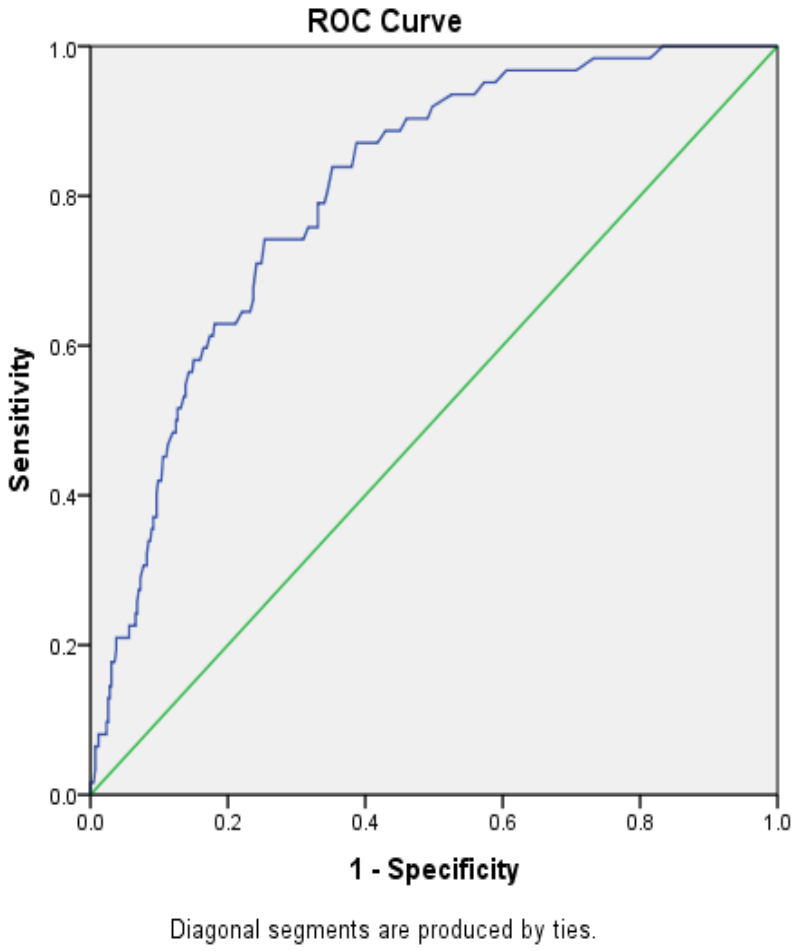


Figure 4.10. Red cell distribution width ROC curve for HbSS+ßeta-thalassaemia

4.5.7 Red Cell Distribution Width ROC curve for Beta-Thalassaemia

According to the relevant ROC, RDW cut-off point of 19.9% showed a sensitivity of 44.4%, specificity of 69.6% and accuracy (Youden index) of 0.542 (i.e, 54.2%) for beta-thalassaemia and a positive likelihood ratio of 1.46 and negative likelihood ratio of 0.79. The results thus demonstrate a modest ability of about 70.0% for RDW as a surrogate biomarker for the diagnosis but poor ability (only 44.4%), screening patients for ß-thalassaemia, **p=0.662 (***Table 4.5 and Figure 4.11).*

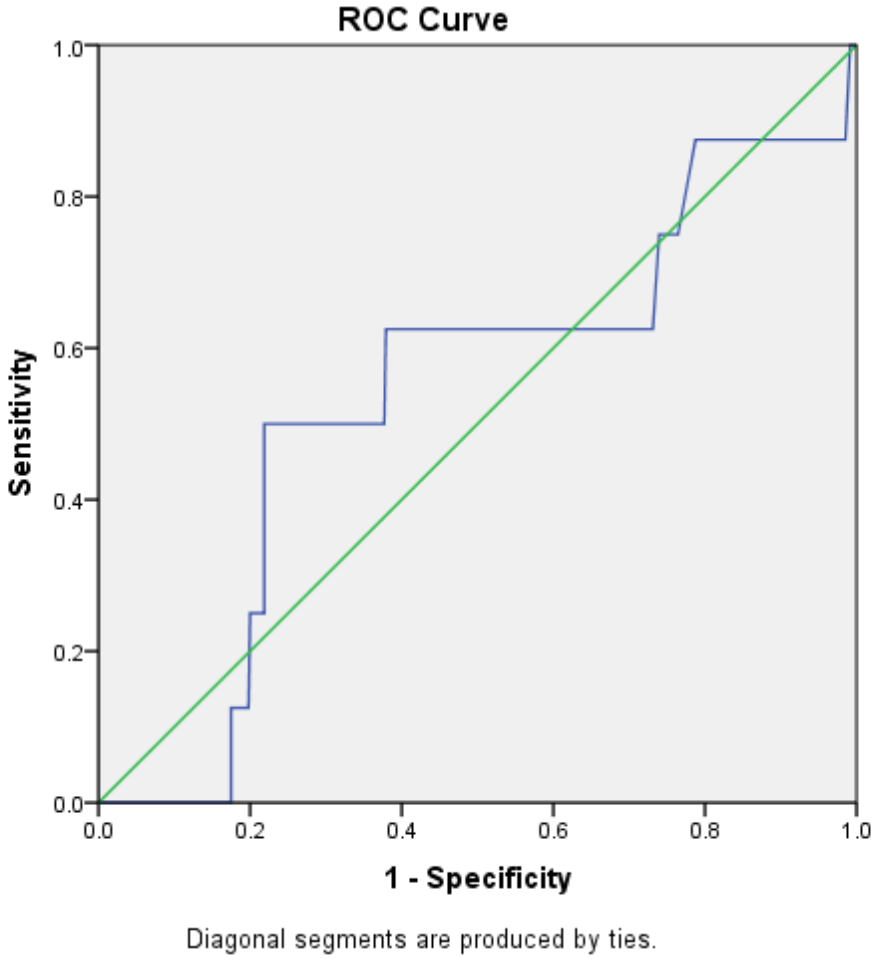


Figure 4.11 Red cell distribution width in beta thalassaemia

4.5.8 Sensitivity, Specificity, Predictive Values, Likelihood and Odds Ratio for RDW Cut-off limit (19.9%) for various haemoglobinopathies among patients at Aga Khan Hospital, Kisumu Western Kenya and its satellites

Values of the sensitivity, specificity, predictive values, likelihood ratio and Youden index (for accuracy), of the set RDW reference cut-off limit, in respect of the various haemoglobinopathies were derived from the ROC curve. These together with odds ratios values computed from sensitivity and specificity were then used to evaluate the utility of RDW as a haemoglobinopathy biomarker, for each haemoglobinopathy. For the

haemoglobinopathies combined together the ROC curve for the RDW cut-off limit of 19.9% gave a sensitivity, specificity and accuracy (Youden index), respectively, of 55.1%, 94.2% and 0.746 (i.e, 74.6%). Results for these parameters for evaluating the utility of the RDW cut-off limit are summarized on *Table 4.6* and depicted on the respective ROC curves (Figure 4.12)

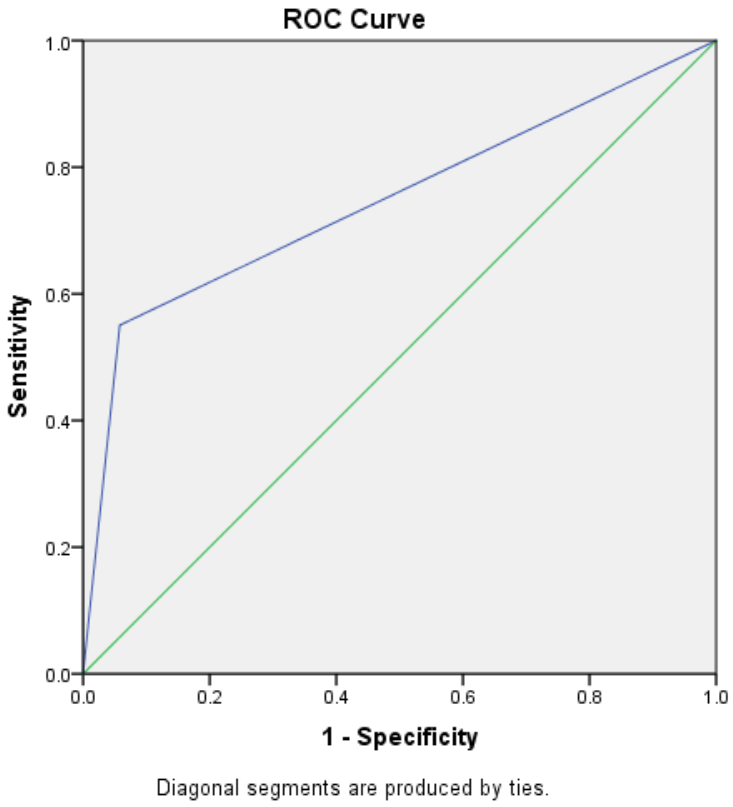


Figure 4.12. Red cell distribution width ROC Curve in Haemoglobinopathies

The related positive and negative predictive for the cut-off limit were, 90.7% and 67.2% respectively, while the positive and negative likelihood ratios were 9.5 and 0.476 (*See Figure 4.12 and Table 4.6*).

Table 4.6. Red cell distribution width ROC values in respective of haemoglobinopathy for 19.9 reference cut-off limit

		Patients &hemoglobinopathy status POSITIVE	NEGATIVE	Total	Predictive Values
RDW	>19.9 POSITIVE	136	14	150	90.7% (+)
	<19.9 NEGATIVE	111	227	338	67.2 % (-)
Total		247	241	488	
Sensitivity & Specificity		55.1%or 0.551	94.2% or 0.942	Odds ratio of 19.86	
Likelihood ratios		9.5	0.476		

Note: RDW; red cell distribution width, Positive predictive value = 90.7%; Negative predictive value = 67.2%, sensitivity=55.1%, specifity=94.2%, Odds ratio=19.86, Positive likelihood ratio = 9.5; Negative likelihood ratio = 0.476.
†, This table gives a summary of red cell distribution width cut-off value of 19.9 predictive ability in terms of sensitivity (%) and specifity (%), likelihood ratio and odds ratio.

To estimate the risk of having an haemoglobinopathy using red cell distribution width at the same cut-off value of 19.9%, the odds ratio of 19.89 (10.96-36.02) was obtained *(table 4.7)*

Table 4.7. Estimate of Odds of having haemoglobinopathy if RDW >19.9

		95% Confidence Interval	
	Value	Lower	Upper
Odds Ratio for RDW (>19.9POSITIVE/<19.9 NEGATIVE)	19.866	10.956	36.022
ForcohortHaemoglobinopathies = HEMOGLOBINOPATHY POSITIVE	2.761	2.351	3.243
ForcohortHaemoglobinopathies = HEMOGLOBINOPATHY NEGATIVE	.139	.084	.230
Number of Cases	488		

Note: RDW; red cell distribution width,
†, This table gives a summary of red cell distribution width cut-off value (>19.9) risk estimate of having haemoglobinopathy using odds ratio.

CHAPTER FIVE

DISCUSSION OF RESULTS

5.1 Distribution of haemoglobinopathy phenotypes in Western Kenya

Overall, this study reports that HbAS had the highest proportion which contrasts somewhat with findings that were reported by other studies done in Western Kenya, which found bearers of α-thalassaemia in the highest number of individuals. For example, Suchdev *et al.*, (2014) reported positivity rates of, respectively, 18.7% for HbS (HbAS 17.1% and HbSS 1.6%) and 48.1% for α-thalassaemia (Hbα/α-thal, 9.6% and 38.5%, Hb-α/α-thal) and none for ß-thalassaemia phenotypes in a population-based study in Bungoma region. Foote *et al*., (2013) reported an identical trend, 17.1% for HbAS and 1.6%, HbSS; 38.5% for Hb -α/α-thal and 9.6%, Hb -α/α-thal; and similarly, no Hbß-thal phenotypes was reported in Western region. Basically, concordant with this, a population study of children in rural areas of Kakamega, Bungoma and Vihiga Counties by Byrd *et al.*, 2019, reported HbAS prevalence of 16.2%; HbSS, 0.2%; Hb -α/αα-thal, 40% and 8.2%, Hb -α/α-thal; HbAS+Hb -α/αα-thal,7.1 and HbAS+Hb-α/α-thal, 1%; and HbSS+Hb -α/α-thal, 0.2%, but similar to the foregoing studies, neither Hbß-thal nor HbF phenotypes were reported. In a population survey of haemoglobinopathy done among children in Kombewa, Kisumu County, Western Kenya reported an overall HbS prevalence of 19.9% (HbAS, 19.0% and HbSS, 0.9%) and 53.2% for α-thalassaemia (homozygous, 8.8% and heterozygous, 44.4%) (Kifude *et al.*, 2007). Another population Kenya survey of children, conducted in Kilifi County, reported prevalence of 8.6% for HbS (HbAS homozygous, 0.8% and HbAS heterozygous,7.8%) plus 65.5% for α-thalassaemia (homozygous, 16.9% and

heterozygous,48.6%) (Uyoga *et al.,* 2019). The study reported, further, a prevalence of 14.2% for HbSS+α-thalassaemia homozygous and 40.2% for HbSS+α-thalassaemia heterozygous. Similarly, in a hospital-based survey done in Jaramogi Odinga referral hospital among children with acute *p. falciparum* recorded a high HbS prevalence of 31.8% (HbSS, 11.1% and HbAS, 20.7%) signifying a future overwhelming burden of haemoglobinopathies in western Kenya region (Kosiyo, *et al.*, 2021).

All the stations, except Kakamega, had at least some form of ß-thalassemia phenotypes, unlike other studies in the region (Suchdev *et al.,* 2014, Foote *et al.,* 2013, Byrd *et al.,* 2019), and Kilifi County in the Indian Ocean Coastal region of Kenya (Uyoga *et al.,*2019. On the other hand, all the stations had no case of α-thalassemia phenotype in any form, in contrast with most previous studies in the region, all of which reported α-thalassemia phenotype at least in some form, but no ß-thalassemia phenotypes of any form at all. In addition, all the stations, except Migori, had some form of HbF phenotype, unlike other studies in the region, which reported no HbF phenotype in any form at all. This differences in distribution of hemoglobinopathy phenotypes could be attributed to differences in method of analysis employed by the present study and those Western Kenya prevalence reports There was a widespread co-inheritance of Hb F with the other Hb phenotypes in western Kenya, the evident from the findings of the present study.

Kisumu station had the largest share of haemoglobinopathies led by HbSS+β-thal with a high proportion of 12.6% (n=31), followed by HbAS with 12.1% (n=30); HbSS, 10.1% (n=25); HbSS+HbF had 2.8% (n=7); HbAS+β-thal 1.2% (n=3); β-thalassemia, 2.4% (n=6) with this last being the highest across all the stations. These results were similar,

on HbS phenotypes, to those of a study done at Jaramogi Odinga Teaching and Referral Hospital in Kisumu (Kosiyo *et al.*, 2021). The study reported a high prevalence of sickle cell haemoglobinopathy in Kisumu, with HbAS having 23.5%, (n=31) while HbSS had 8.3% (n=11) 6.5% % (n=14) were HbAS and 15.3% (n=13) (Kosiyo *et al.*, 2021)(Kosiyo *et al.*, 2021). These reports of high proportions of haemoglobinopathies in Kisumu County more than any other western Kenya region, may be an indicator that haemoglobinopathies may be ethically/ geographically driven disorders as reported by Sawaimul *et al.*, 2018. This could serve as a wakeup call to government to develop strategies to combat these diseases in regions reported to have high prevalence of haemoglobinopathies.

Busia region was the second in high proportions of haemoglobinopathies with literature showing no similar study done in this region. The region is inhabited predominantly by three major ethnic communities as the indigenous population, that include: AbaLuyha, Luo and Teso communities (Hassan 2017) but the high proportions of hemoglobin disorders reported by the present study could not be attributed to any specific ethnic group. However, the findings of this study are in tandem with the observation by (Hamali & Saboor, 2019, Hamali, 2019) that undiagnosed haemoglobinopathy, especially HbS, remains a major threat to public health since unnoticeably creeping in societies when governments and international agencies focus more on other diseases (Hamali & Saboor, 2019, Hamali, 2019). A study to determine the prevalence of sickle cell Hb in Uganda, documented similar findings to the current study on the overall prevalence (14.7% and 2.8%, respectively) of HbAS and HbSS. This as noted by

Hernandez, *et al. (*2019), underscores the high prevalence of sickle cell haemoglobinopathy in Africa.

Homa Bay showed higher proportions of haemoglobinopathies with 9.7% (n=24) of the subjects drawn from there, having HbAS; 2.4%(n=6) as HbSS+β-thal, 1.6%(n=4) as pure HbSS and HbSS+HbF, 0.4% (n=1). On the other hand, 0.8% (n=2) had HbAS+β-thal and 0.4% (n=1) had beta thalassemia. No similar studies are reported for Homa Bay. The subjects drawn from Bungoma had low overall haemoglobinopathy proportions of 5.7% (n=14), with most, 2.8% (n=7), having HbSS+β-thalassemia, followed by 2.0% (n=5) having HbAS, while 0.4% (n=1) HbSS and 0.4% (n=1), HbSS+HbF. None of the subjects had β-thalassemia. Similar to Bungoma, Kakamega, Kitale (in Trans Nzoia County), Kisii and Migori sites recorded significantly low numbers of subjects (p=0.017**)** with overall haemoglobinopathy, 4% (n=10) each. The difference in haemoglobinopathy positivity rates between these counties and Kisumu, Homa Bay and Busia counties can be attributed to the relatively smaller numbers of the subjects obtained from there. This explanation seems plausible, based on a population survey by (Byrd *et al.,* 2019) targeting HbS and Hb α-thal in children of Luhyia ethnic communities' resident in rural areas of these counties plus Vihiga County (also a Western Region County). The survey reported an overall haemoglobinopathy distribution of HbS, 16.4% (HbAS, 16.2% and HbSS, 0.2%); HbAS+Hbα-thal, 8.1% (HbAS+Hb-α/α, 7.1% and HbAS+Hb-α/-α, 1%). Unlike the present study, however, the latter did not dis-aggregate the haemoglobinopathy profile into separate proportions for each of the three counties.

With these reports, that the malaria-holoendemic region of western Kenya has high prevalence of haemoglobinopathies (Kosiyo *et al.*, 2021), it was interesting to observe that regions dominated by Bantu speaking communities including: Bungoma, Kitale, Kakamega, Kisii and Migori recorded significantly ($p=0.017$) low proportions of haemoglobinopathies compared to Kisumu, Busia and Homa Bay counties within western Kenya. According to Hassan (2017) and Alwy & Schech (2004), the Western Region (former Western Province) consisting of present Kakamega, Bungoma, Vihiga and Busia counties), has the Abaluhya, of the Bantu extraction, as the dominant indigenous ethnic community, while Nyanza Region (former Nyanza province), consisting of present Siaya, Kisumu, Homa Bay, Kisii, Nyamira and Migori counties has the Luo, who are of Nilotic extraction, and the Abagusii, of Bantu extraction. The Luos dominate in the counties of Kisumu, Homa Bay, Migori and, Siaya, while the Bantus (Abagusii specifically) dominate in Kisii and Nyamira counties. The counties of Homa Bay, Migori and Siaya have appreciable numbers of Bantu-speaking people as minority communities: Abasuba, Abakuria and Abasonga communities, respectively.

The Luo-dominated counties are the immediate host to Lake Victoria, and experience more directly the climatic influence of the lake, including rainfall, and temperature and hence malaria-endemicity. They are thus likely to have a higher haemoglobinopathy prevalence consequent on these environmental factors (Matsushita *et al.*, 2019, Kosiyo *et al.*, 2020). Ethnicity was not factored in the design of this study, but the highest proportions of haemoglobinopathy were found in the Aga Khan hospital sites situated in the counties immediately around Lake Victoria. This variation in haemoglobinopathy distribution between the sites in two geographical areas might be attributed broadly to

differences in geographical location of habitation or, other, geophysical environment and ethnicity. This underscores the possible role of the geographical environment, through epigenetic influence, and ethnicity on the susceptibility of human populations to haemoglobinopathy. In this case the fresh water nature of Lake Victoria and the temperate climate is likely to favor malaria-endemicity, thereby favoring higher prevalence of haemoglobinopathy among the populations immediately around it. These findings are consistent with a previous study done by Sawaimul *et al. (*2018), demonstrating association between ethnicity and geographical region of habitation on haemoglobin gene expression, capable of causing variation in haemoglobinopathy profiles.

5.2 Red Cell Distribution Width

The upper limit for the 95% CI of the median for the RDW of the control group was used as the reference cut-off limit to evaluate those for the haemoglobinopathy cases. The median was 14.5%, IQR=2.7; 95%CI=9.1-19.9%, giving the upper limit and hence reference cut-off limit as 19.9%. Both the median values and 95% confidence intervals are much higher than those reported by similar previous studies. In Netherlands, Hoffmann *et al.,* (2015) reported for 5216 healthy females aged 1-102 years a median RDW of 12.03% (IQR=11.6-12.54%); for 2873 males aged 1-106years, a median RDW of 12.05% (IQR=11.69-12.55%). Similarly, Alis *et al*., (2015) in Spain reported RDW median values of 13.0% (IQR=12.4-14% for 380 healthy males aged 10-90years; and 13.2% (IQR=12.5-14.2% for 429 healthy females aged 10-90years. Finally, Li *et al* (2020) reported RDW of 13.4% (IQR=12.4-15.4%) in 1280 Taiwanese males aged 66.8-80.2years, while Roberts *et al.,* 1985 reported an RDW mean of 11.3% in children under

the age of 2 years. It can be seen that for these four previous studies the median RDW values were 11.2-13.4%, with the upper limits of their 95% confidence intervals ranging between 12.5 and 15.4%, of which even the highest is considerably lower than the value obtained in this study. The finding of a relatively high RDW in the present study as compared to other previous studies could be attributed to use of indirect approach (patients) rather than the direct approach (healthy population) to derive our reference cut-off limits for patients. This explanation is supported by the observation of lower median RDW values of 11.2% (IQR=2.2) reported by Kosiyo *et al.,* (2020) in Kisumu County, western Kenya for a small number (n=90) of younger subjects (aged one month to 5.5years) free from malaria and with normal Hb phenotype (HbAA), though with other acute febrile illnesses of unspecified etiologies. Other critical factors such as, population age group, assay machines, statistical methods used in computation, geographical environments, ethnicity, malaria co-morbidity, sickle cell and sociocultural backgrounds between our study and that of Kosiyo *et al.*, 2020 could also account in the differences of our RDW compared to other studies. This view is supported by several studies in Africa which attributed the differences in the normal reference ranges for haematological parameters to various factors, including sex, geographical location, race, altitude, and diet (Luganda *et al.,* 2004, Karita *et al.,*2009).

Similarly, a study in Nakuru County, Kenya by Gachie (2018) found lower normal intervals for haematological parameters than those derived outside Africa and attributed the differences to genetic, ethnic, and demographic factors. In another study done at the Kenya Medical Research Institute (KEMRI) in Kisumu by Odhiambo *et al.,*2015, found out that the use of local haematological and biochemical normal reference ranges

categorized fewer study subjects as abnormal, compared to the use of ranges derived in the USA. This is similar to another Kenyan study, done earlier in Kericho County, which compared an Hb reference range derived from black Americans with one done locally and found that the American range would have erroneously classified the entire population of the county as anaemic (Kibaya *et al.,* 2008). All these studies have recommended adoption of locally developed normal ranges for quantitative biomarkers of various diseases (Scherer *et al.,* 2015).

The case group had a significantly higher RDW median of 20.7% (IQR=8.3) than the control group (14.5, IQR=2.7;95% CI=9.1-19.9), $p<0.0001$, indicating marked anisocytosis and we were unable to get RDW previous findings for combined haemoglobinopathies. However, subjects with homozygous sickle cell disease Hb (Hb SS) had the highest RDW median of 25.4% (IQR=5.5) ($p<0.0001$), and similar findings are recorded by Webster & Castro (1986) whose Hb SS had a mean RDW of 22.4±4. Sickle cell disease+β-thalassemia (HbSS+β-thal) had the second highest RDW, median of 23.3% (IQR=7.9) ($p<0.0001$) and sickle cell disease (HbSS+HbF) 20.9 (5.5), which was the lowest among the HbSS phenotypes. A similar order of decreasing magnitude of RDW was documented by Roberts *et al.,* (1985) in anemic patients, with the highest value seen in sickle cell anemia (HbSS), followed by sickle cell thalassaemia (HbSS+β-thal), β-thalassemia major (β-thal), sickle cell trait (HbAS) and lastly iron deficiency anemia (IDA) in decreasing order of magnitude. Qurtom *et al.,* 1989 documented a somewhat contrasting pattern of RDW values for the different hemoglobin disorders where the sickle cell disease (HbSS) had a lower RDW, mean 18.2±3.8, than

HbSS+HbF, 20.3±3.1; followed by HbSS-thal, 20.2±4.5 and Beta-thalassemia, mean of 15.4±1.4. We were not able to figure out an explanation on this phenomenon.

HbSS+HbF had the lowest RDW median among the SCD phenotypes, which could be due to the stability and higher affinity for oxygen of HbF that enables the Hb to reduce anisocytosis, by not entering HbS-polymerization phase, thereby causing haemolysis, or trigger rapid and enhanced erythropoiesis which would lead to reticulocytosis, from tissue hypoxia (Nkya *et a.,l* 2017, Maekwa and Kato 2015, Macharia *et al.,* 2019). This way HbF modulates clinical sickle cell disease severity and its hematological features, such as enhanced and rapid reticulocytosis and elevated RDW (Damanhouri *et al.,* 2015, Sundd *et al* 2019). The survival value and selective advantage of victims with HbF is demonstrated by a prospective study done in Kilifi area of Kenya, that documented reduced morbidity and mortality of SCD Children under the age of 5 years with elevated Hb F (Uyoga *et al.,* 2019). This reduced mortality and morbidity can be explained by the fact that, Hb F has ability to counter the tissue hypoxia caused by the polymerization Hb in HbSS containing RBC and their hemolysis (Carden *et al.,* 2020, Nkya *et al.*, 2017). Indeed, Carden *et al.,* (2020), citing several previous studies reported that HbF concentration together with reticulocyte numbers and erythrocyte deformability (sickling) are among the determinants of cell anaemia pathology and disease severity. Hence, the fact of HbSS+HbF bearers in the current lowest RDW can be attributed to the tendency of HbF to counter the hypoxic effect of HbSS on affected erythrocytes and consequent reticulocytosis normally caused by hypoxia-induced enhanced and rapid erythropoiesis (Nkya *et al*.,2017, Mairbaurl 2018, Zigot *et al.,* 2018, Mandal and Kartthik 2019), thereby minimizing the development of anisocytosis. Minimization of

the anisocytosis, of which the immediate cause is reticulocytosis, is likely to prevent the elevation of RDW (Webster and Castro 1986, Qurton *et al.*, 1989).

The polymerization of Hb molecules, besides distorting structure of RBCs thereby leading to their destruction (Sunnd *et al*.,2019, Carden *et al.,* 2020, Picci *et al.*, 2019), also impairs their ability to take up and transport oxygen to tissue cells, causing tissue hypoxia (Sundd *et al.,* 2019). Hypoxia has been found to be the main trigger for the systemic response of faster-paced and enhanced erythropoiesis, leading to reticulocytosis (Haase 2013, Vlaski *et al.,* 2009, Nkya *et al.,* 2017, Carden *et al*.,2020). Reticulocytosis has been found to be a significant determinant of RDW (Castro and Webster1986, Tariq *et al.,* 2019), demonstrating the link of this haematological parameter with haemoglobinopathy.

Sickle cell trait (HbAS) produced a median of 16.4% (IQR=6.5) which was the lowest among hemoglobin disorders investigated by the present study. Though significantly elevated ($p<0.0001$), RDW would generate many false positive results for HbAS if used as a screening tool because its median was within the standard cut-off value of 17.2. Similar findings are documented by Olujohungbe *et al.,* (1993), that HbAS bearers have normal RDW, making it difficult to use this parameter to distinguish it from normal population. This is understandable, considering that in HbAS, the Hb molecule is partially normal structurally and thus functionally. This co-existence of normal Hb with Hb S is likely to mitigate the hypoxic-erythropoietic effect of the latter, thereby minimizing the RDW elevation common with the simple HbSS phenotypes. At the same time, the RDW median values for HbAS+β-thal and Hbβ-thal had values of 24.2 **(7.4)**, 20.9 (10.5) and 18.1 (8.3) and these showed no statistically significant difference from

the reference cut-off value (p=0.449, p=0.791 and p=1.0 respectively). This implies that RDW cannot help in discriminating these haemoglobin phenotypes from each other, a conclusion that concurs with a similar observation by Qurtom *et al., (*1989). Though their RDW medians seemed elevated, but when subjected under statistical test, did not have statistical significance. This might be due to the small numbers of suggesting small sample size could have resulted into an abnormally high RDW.

RDW was higher in males with, a median of 17.1% (IQR=8.0) than females who had a median of 15.5% (IQR=6.7) but did not have statistical significance (p=0.317) between the two sexes as shown on Table 4.3. Their respective control group showed no statistical significance (p=0.89) with males having a median of 14.55% (IQR=1.90) while female had a median of 14.2% (IQR=2.40) respectively, suggesting that a single RDW normal reference range can be used in both males and females. In a study done by Qurtom *et al.*, (1989), similarly, found no statistically significant RDW gender difference in RDW across the various age groups. This is supported by Hoffmann *et al.* 2015, who reported no association between RDW and gender. Thus, the same normal reference range for RDW can be used for both sexes.

Concerning age and RDW, children under the age of 5 years had the highest overall RDW which gradually and significantly decreased with increasing age of the study participants with those aged above 12 years having the lowest RDW median **(**$p<0.05$) (Table 4.3). However, the RDW of study participants aged 6-12 years when compared to the age group above 12 years, did not show a statistically significant difference from the reference cut-off limit, irrespective of gender (p=0.347) (See Table 4.4). This suggests that individuals aged above 6 years of age to adulthood can have a common RDW

reference range but children under the age of 5 years need to have their own. Save for Qurton *et al.,* 1989, these findings in relationship between age and RDW appear contrary to those of several previous studies except those by Qurton *et al.,* (1989). The latter reported similar findings with children under the age of 2 years having a higher mean RDW than the adult population. This can be due to high erythropoietic activity in children causing release of reticulocytes that are larger than normal rbc's leading to anisocytosis (elevated RDW). The authors noted that, RDW becomes abnormal earlier than Hb or even MCV, making it a good biomarker for anisocytosis in children under the age of 5 years with anaemia related disorders. In contrast, studies by Hoffmann *et al.,* (2015); Alis and Badrick, (2014) and Lippi *et al*., (2014) reported a strong positive association between RDW and age. However, these studies employed subjects of age sets ≥ 10years. Specifically, Hoffmann *et al., (*2015) employed subjects aged above 18 years and grouped those less than 18 years as a single age, the RDW in their study increased with age, most especially those above 18 years. On the other hand, Alis and Badrick, 2014; and Lippi *et al.,* 2014) used individuals aged ≥ 10years and ≥2 0years, respectively. The reason behind high RDW in children under the age of 5 years could be due to vigorous erythropoietic activity of bone marrow, releasing immature RBC (reticulocytes) that are usually larger than normal red cells. Release of these reticulocytes into the peripheral blood circulation, having also the mature, normocytic erythrocytes brings about this variation of the sizes of red cells in children, and for under-fives given the enhanced pace and magnitude of erythropoietic activity results into heightened reticulocytosis. Thus, the observed trend in the older people, might be as a result of adverse events in a multitude of clinical conditions that comes with advancing age (Hoffmann *et al.,* 2015).

5.3 Sensitivity, specificity, Youden index (overall accuracy), predictive validity and likelihood ratio of RDW as a biomarker for haemoglobinopathy in Western Kenya.

The evaluation of utility of RDW as a surrogate biomarker for laboratory detection of haemoglobinopathies was based on the observed values for the above laboratory test or disease biomarker evaluation parameters, using the reference cut-off limit, set from values for the control group. The values were derived through the analysis of the ROC curves of the RDW reference cut-off limit, which was 19.9% in the manner described by Ruopp, *et al. (*2008), for all the haemoglobinopathies combined, and for the individual haemoglobinopathy phenotypes as well. The acceptable sensitivity and specificity, of a diagnostic test or biomarker, respectively, are at the optimal point of the ROC curve while the threshold for Youden index is obtained as the area under the ROC curve (AUROC) which is the optimal accuracy of a biomarker (Ruopp *et al.,* 2008). The ROC curve parameters used to evaluate the utility of the chosen RDW reference cut-off limit of 19.9% include sensitivity, specificity and Youden Index (for accuracy of the set RDW reference cut-off limit). According to Parikh *et al.*, (2004), a test or biomarker with a sensitivity and specificity of 90% (i.e. 0.9) and 95% (i.e. 0.95) respectively, should be considered of excellent diagnostic and screening utility. For the Youden Index, values of ≥50% (i.e, ≥0.5) reflect acceptable overall accuracy of test method or, other, biomarker analyzed by a test method, in detecting a disease.

For the combined haemoglobinopathies, this study found sensitivity and specificity of 55.1%., 94.2%; Youden index of 0.746 (accuracy of 74.6%); likelihood ratios of 9.5(positive) and 0.476(negative); positive predictive value of 90.7% and negative predictive value of 67.2%; and an odds ratio of 19.86 (10.96-36.02). These results show

satisfactory clinical specificity, but low sensitivity of the red cell distribution width in haemoglobinopathy diagnosis generally or, other, confirmation of presence of the HbSS phenotype ($p< 0.0001$). However, the low value for sensitivity shows poorer utility of RDW as a biomarker for haemoglobinopathy screening generally. Studies determining sensitivity, specifity, predictive, likelihood ratio of RDW cut-off value as biomarker for haemoglobinopathies does not exist currently, making the present study the first ever attempt to evaluate RDW rigorously as a potential biomarker for hemoglobin disorders.

For HbAS, a borderline Youden index value of 0.501 vs the normal threshold of 0.5 was observed, together with a low sensitivity of 30.1% and a modest specificity of 69.1%. In addition, negative and positive likelihood ratios of 0.97 and 1.01, respectively were observed. Together with a Youden index of 0.501 (reflecting borderline overall accuracy), these results indicate modest potential utility of RDW as a diagnostic or confirmatory testing biomarker, but not as a screening biomarker for HbAS haemoglobinopathy. A similar trend as for HbAS was observed in respect of HbAS+β thalassemia, whereby there was low sensitivity (50%) against modest specificity (69.5%) and above borderline accuracy (a Youden index of 0.6), representing modest diagnostic but poor screening utility for HbAS+Hbβ-thalassemia.

For HbSS, in contrast with HbAS and HbAS+Hbβ-thalassemia, RDW had the significantly high sensitivity of 88.9%, specificity of 75.2%, overall accuracy (Youden index) of 0.892 and positive likelihood ratio of 3.59 for HbSS ($p<0.0001$). This shows an above-borderline potential utility as both a screening and diagnostic biomarker for HbSS or other sickle cell disease, compared to sickle cell trait or sickle cell trait with

beta thalassaemia Hb phenotypes. Webster & Castro (1986) documented similar findings of increased RDW in HbSS, followed by HbSS+β-thalassemia and HbSS+Hb F.

The RDW was also proved to be a good diagnostic but poor screening biomarker for sickle cell disease with fetal Hb (HbSS+HbF) having recorded a sensitivity of 60%, specificity of 70.5 %, Youden index of 0.766 and a positive likelihood ratio of 2.03. The reduced sensitivity of RDW for HbSS+HbF could be attributed to the fact that HbF is more stable than HbSS and resists polymerization under reduced O_2 thus reducing red cell lysis and consequent variation in red cell size and shape variation (Damanhouri *et al.,* 2015).

For the HbSS+ß-thalassaemia Hb phenotype, (HbSS+β-thal), the sensitivity of 71%, specificity of 75.1%, an accuracy of 0.805 and 2.85 positive likelihood suggesting efficiency in detecting this haemoglobinopathy. Beta-thalassemia (p=0.662) had a sensitivity of 44.4%, specificity of 69.6% and a Youden index of 0.542 in detecting beta-thalassemia as shown on Table 4.5. Qurtom *et al.,* 1989 documented similar findings where beta-thalassemia had a normal or mildly elevated at a mean RDW of 15.4±1.21 which indicated that red cell distribution width could not discriminate beta-thalassemia from beta thalassemia-free patients.

At the reference cut-off limit of 19.9%, varying utility as a screening and diagnostic biomarker for all the haemoglobinopathies combined, as shown in *Table 4.5*. It gave a sensitivity of 55.1%, specificity of 94.2% and a Youden index of 0.746. This means that at this cut-off limit, RDW can correctly detect the presence of 55.1% of the patients having a haemoglobinopathies and correctly identify as haemoglobinopathy-free 94.2% of patients actually not having a haemoglobinopathy. At the same time, it showed

positive and negative predictive values of 90.7% and 67.2%, respectively. This means that there was a probability of 90.7 % that having an RDW value above 19.9% actually represents presence of haemoglobinopathy, while having probability of 67.2% of having an RDW below 19.9% when a patient actually has no haemoglobinopathy.

Likelihood and odds ratio were also derived as part of the evaluation of the utility as a diagnostic marker for combined haemoglobinopathies. Likelihood ratios are, clinically more useful than sensitivity and specificity in providing usefulness of a diagnostic tests (Akobeng, 2007). The Positive likelihood ratio expresses how likely a test is going to correctly diagnose the presence of the condition where the greater the positive likelihood ratio (LR+), the more likely the test is going to help us come up with a true positive diagnosis (Akobeng, 2007). Negative likelihood ratio (LR-) is the ratio of how likely a test will correctly diagnose the absence of a condition whose value is usually less than one (<1) and the smaller the value going to zero, the better the test is correctly indicating the absence of the condition. When a test in positive or negative likelihood ratio is close to one (1), indicates that such a test has little influence to predict the presence or absence of a disease and therefore such a test is worthless in clinical practice (Akobeng, 2007). At the cut-off limit of 19.9%. the LR+ and LR- were respectively, 9.5 and 0.476 respectively, meaning any individual having RDW above 19.9% was 9.5 times more likely to have a hemoglobinopathy, while having RDW below 19.9% was about 0.476 times likely to have haemoglobinopathy. On the other hand, the odds ratio (OR) means the odds or risk of having a haemoglobinopathy when RDW was higher than the cut-off of 19.9%. In this study, an odds ratio of 19.86 was obtained, meaning that individuals with RDW of more than 19.9% were at 19.86 times at risk of hemoglobinopathies

compared to those with a RDW of less than 19.9% (Table 4.8.). Finally, value of Youden Index, J, was 0.746. This value compared to the minimum of just over 0.5 (or 50%) indicates that RDW can enable the diagnosis of haemoglobinopathies with an overall accuracy of 74.6%. The values of all these parameters obtained for the RDW cut-off limit (i.e 19.9%), indicate good potential of the haematological index (RDW) to serve as a haemoglobinopathy diagnostic biomarker but the low sensitivity indicates poor potential of the index as a haemoglobinopathy screening biomarker.

Despite relatively low sensitivity compared its own specifity, this test is still diagnostically higher than of the aforementioned traditional tests used easy-to-perform tests that are far cheaper than cellulose acetate paper electrophoresis and HPLC (Okwi *et al.,* 2010). Its noteworthy, that even though later studies evaluating the above traditional SCD testing methods have given higher ratings of performance in terms of sensitivity, specificity and overall accuracy (Arish *et al.,* 2021, Aleluia *et al.,* 2017, Piety *et al.,* 2016, Mkocha *et al.,*2018), they have not been subjected to the same rigorous evaluation of their utility as the RDW in the current study, which has also incorporated the Youden index and the area under the curve (AUC) of the receiver operating Characteristics (ROC). These findings, considered together, indicate that RDW has a significant utility as a diagnostic test for patients with haemoglobinopathies and it is still superior to the latter as a screening test for haemoglobinopathies. On the same note, RDW showed a significant overall ability as a diagnostic marker for both screening and diagnosis of sickle cell disease and is therefore very suitable for a resource poor setting such as Western Kenya and, indeed the whole country and other countries with similar socioeconomic conditions.

CHAPTER SIX

CONCLUSIONS AND RECOMMENDATIONS

6.1 Introduction

The present study aimed at assessing the validity of RDW as an accurate but cost-effective biomarker in laboratory detection of haemoglobinopathies thus, its findings is expected to improve the management of these disorders in Kenya and other resource limited settings. This was a cross-sectional retrospective comparative study done in Aga Khan Hospital, Kisumu and its satellite centers located in the wider western Kenya. Since RDW is routinely, widely used, easy and affordable, its established cut-off value of 19.9, gave a sensitivity of 55.1%, specificity of 94.2% and accuracy of 74.6% is expected to serve as an effective biomarker to accurately and timely diagnose haemoglobin disorders, therefore improve chances of survival (reduce mortality) of infants and children with severe haemoglobinopathies born in poorer countries with financial difficulties most especially sub-Saharan African countries including Kenya.

6.2 Conclusions

The study represents the burden of haemoglobinopathies in a malaria-holoendemic region of Western Kenya and even though the present study did not include ethnicity in data collection, stations from where the data was collected are predominated by different communities therefore, there may be an ethnic correlation in variation of haemoglobinopathies in western Kenya. The communities juxtaposed to the lake seems the most affected ethnic group along Lake Victoria Economic Block region thus, it may be erroneous to assume that the entire malaria-holoendemic region of western Kenya has high prevalence of hemoglobin disorders without factoring ethnicity and geographical

location in a properly conducted population-based prevalence study in wider western Kenya

The overall RDW median of haemoglobinopathy cases was 20.7 (IQR=8.3) while the control group was 14.5, IQR=2.7; 95% CI=9.1-19.9 which was used to derive 19.9 cut-off point at 95% C.I. Above the cut-off point, RDW is a promising diagnostic biomarker for SCD phenotypes (Hb SS, Hb SS+Hb F & Hb SS+β Thal) but not SCT phenotypes (Hb AS, Hb AS+Hb F & Hb AS+ β-thal) and beta thalassemia

Red cell distribution width at 95% C.I was 19.9 [14.5+(2.7×2=19.9)] the cut-off point of 19.9, showed a low sensitivity, a good specificity and predictive values plus likelihood ratio of 55.1%., 94.2%; Youden index of 0.746 (accuracy of 74.6%); positive predictive value of 90.7% and negative predictive value of 67.2%; likelihood ratios of 9.5 (positive) and 0.476 (negative) and an odds ratio of 19.86 (10.96-36.02) in respect of haemoglobinopathy in general. Accordingly, the study suggests that RDW potentially has high utility as a biomarker for diagnosis or confirmatory testing, but not screening, in regard to combined haemoglobinopathy as well as HbAS phenotypes. On the other hand, it showed same pattern of potentially good performance as a diagnostic as opposed to a screening biomarker for HbSS phenotypes, in contrast with HbAS phenotypes and beta thalassemia.

6.3 Recommendations

(i) Premarital and newborn screening of haemoglobinopathies should be implemented as the initial diagnostic strategy to prevention and control of these disorders in the Western Kenya future generation.

(ii) Even though RDW proved to be a poor screening tool for beta thalassaemia, HbAS+HbF and HbAS+β-thalassemia, other CBC parameters such as MCV and red cell count can be used to identify thalassemia syndromes as well as iron deficiency. Though out of the scope of this work, highlighting the significance of these parameters in addition to the RDW would improve its feasibility as a screening tool for all haemoglobinopathies.

(iii) The RDW high accuracy at 19.9% cut-off point in detecting hemoglobinopathies, is expected to improve the clinical outcome in early diagnosis of haemoglobinopathies especially SCD phenotype, if properly combined with other hemoglobin parameters that are known to be interfered by haemoglobinopathies mostly Hb, HCT, MCV and MCHC (Valavi *et al.*, 2010). This will be a quick guide for clinicians to make accurate clinical decision to arrive at accurate diagnosis reducing waste of money and improve patient prognosis (cost-effective). However, it would be necessary to confirm the tests thus innovation of solar powered mini-electrophoretic machines would be appropriate in resource limited settings.

6.3.1 Recommendations for action practice

(i) The high prevalence of HbS carriers in Western Kenya needs urgent government intervention to develop strategies to combat haemoglobinopathies through routine screening, counselling, education, and diagnosis of haemoglobinopathies

(ii) Children under the age of 5 years needs their own RDW normal reference range for precise marking of disorders that bring about variation of red cell shape and sizes not limited to hemoglobinopathies. RDW normal reference range can be used for both males and females provided they are under the same age group. i.e. under 5 years or above 5 years.

(iii) These findings indicate that RDW is significantly promising as a diagnostic marker for haemoglobinopathy thus 19.9% (95% C.I) cut-off point should be included in hematology policy screening algorithm as a critical value above which the unknown cases qualify to be investigated for haemoglobinopathy subject after validation of the present study.

6.3.2 Recommendations for further research

(i) A proper conducted prospective population-based prevalence study on haemoglobinopathies that factors ethnicity and geographical location needs to be conducted in wider western Kenya rather than assuming the entire malaria-holo-endemic region has high prevalence of hemoglobinopathies. This will inform variation of gene polymorphism of different ethnic groups living in wider Western Kenya.

(ii) A study on interference of the entire hematological indices by haemoglobinopathies needs to be done to inform the variation of these parameters in individual haemoglobinopathies. This will improve sensitivity, specifity and accuracy than RDW alone. In addition, prospective study would be necessary as we were unable to control of the influence of some factors, for instance, inter-instrument variations especially in the values of RDW generated over the long period of 5 years,

(iii) The data used were retrospective and hence the diagnostic utility of this haematological index for haemoglobinopathy should be explored further using prospective data since the present study relied on hospital patients rather than a healthy population as a source of our reference cut-off RDW. This calls for empirical verification (validation) of the results to establish their generalizability to enable necessary direction for practice policy

REFERENCES

Adu, P., Simpong, N., Kontor, K., & Ephraim, R. K. (2017). Misleading Presentation of Haemoglobin Electrophoresis data. *Ghana medical journal*, *51*(1), 36-38.

Akobeng, A. K. (2007). Understanding diagnostic tests 2: likelihood ratios, pre-and post- test probabilities and their use in clinical practice. *Acta paediatrica*, *96*(4), 487-491.

Alapan, Y., Fraiwan, A., Kucukal, E., Hasan, M. N., Ung, R., Kim, M., ... & Gurkan, U. A. (2016). Emerging point-of-care technologies for sickle cell disease screening and monitoring. *Expert review of medical devices*, *13*(12), 1073-1093.

Aleluia, M. M., Fonseca, T. C. C., Souza, R. Q., Neves, F. I., da Guarda, C. C., Santiago, R. P., ... & de Souza Gonçalves, M. (2017). Comparative study of sickle cell anemia and hemoglobin SC disease: clinical characterization, laboratory biomarkers and genetic profiles. *BMC hematology*, *17*(1), 1-10.

Alis, R., Fuster, O., Rivera, L., Romagnoli, M., & Vaya, A. (2015). Influence of age and gender on red blood cell distribution width. *Clinical Chemistry and Laboratory Medicine (CCLM)*, *53*(2), e25-e28.

Al-Numan, A. H., & Al-Obeidi, R. A. (2021). The value of the red cell distribution width index versus other parameters in the differentiation between iron deficiency anemia and beta thalassemia trait among children in Mosul, Iraq. *Expert Review of Hematology*, *14*(4), 401-406

Alwy, A., & Schech, S. (2004). Ethnic Inequalities in Education in Kenya. *International Education Journal*, 5(2), 266-274.

Arishi, W. A., Al-Hadrami, H. A., & Zourob, M. (2021). Techniques for the Detection of Sickle Cell Disease: A Review. *Micromachines*, *12*(5), 519.

Arzideh, F., Özcürümez, M., Albers, E., Haeckel, R., & Streichert, T. (2021). Indirect estimation of reference intervals using first or last results and results from patients without repeated measurements. *Journal of Laboratory Medicine*, *45*(2), 103-109.

Aulakh, R., Sohi, I., Singh, T., & Kakkar, N. (2009). Red cell distribution width (RDW) in the diagnosis of iron deficiency with microcytic hypochromic anemia. *The Indian Journal of Pediatrics*, 76(3), 265-268.

Mutua, B. M., Sowayi, G, Okoth, P (2022). *Red cell distribution width as a surrogate marker of haemoglobinopathies in western Kenya*.

Bakan, E., Polat, H., Ozarda, Y., Ozturk, N., Baygutalp, N. K., Umudum, F. Z., & Bakan, N. (2016). A reference interval study for common biochemical analytes in Eastern Turkey: a comparison of a reference population with laboratory data mining. *Biochemia medica*, *26*(2), 210-223.

Battacharya, P. K. (1967). Some general properties of potential fields in space and frequency domain: A review: Geoexploration, 5, 127–143, doi: 10.1016/0016-7142 (67) 90021-X.

Brancaleoni, V., Di Pierro, E., Motta, I., & Cappellini, M. D. (2016). Laboratory diagnosis of thalassemia. *International Journal of laboratory hematology*, *38*, 32-40.

Boyce, W. T., Sokolowski, M. B., & Robinson, G. E. (2020). Genes and environments, development and time. *Proceedings of the National Academy of Sciences*, *117*(38), 23235-23241.

Byrd, K. A., Williams, T. N., Lin, A., Pickering, A. J., Arnold, B. F., Arnold, C. D.,& Stewart, C. P. (2018). Sickle cell and α+-thalassemia traits influence the association between ferritin and hepcidin in rural Kenyan children aged 14–26 months. *The Journal of nutrition*, *148*(12), 1903-1910.

Carden, M. A., Fasano, R. M., & Meier, E. R. (2020). Not all red cells sickle the same: contributions of the reticulocyte to disease pathology in sickle cell anemia. *Blood reviews*, *40*, 100637.

Chakravorty, S., & Williams, T. N. (2015). Sickle cell disease: a neglected chronic disease of increasing global health importance. Archives of disease in childhood, 100(1), 48-53.

Damanhouri, G. A., Jarullah, J., Marouf, S., Hindawi, S. I., Mushtaq, G., & Kamal, M. A. (2015). Clinical biomarkers in sickle cell disease. *Saudi journal of biological sciences*, *22*(1), 24-31.

DeBaun, M. R., Galadanci, N. A., & Vichinsky, E. P. (2019). Sickle cell disease in sub-Saharan Africa.

Dufu, K., Lehrer-Graiwer, J., Ramos, E., & Oksenberg, D. (2016). GBT440 inhibits sickling of sickle cell trait blood under in vitro conditions mimicking strenuous exercise. *Hematology reports*, *8*(3).

Dugdale, A. E., & Badrick, T. (2018). Red blood cell distribution width (RDW)—a mechanism for normal variation and changes in pathological states. *J Lab Precis Med*, *3*, 73.

Eldibany, M. M., Totonchi, K. F., Joseph, N. J., & Rhone, D. (1999). Usefulness of certain red blood cell indices in diagnosing and differentiating thalassemia trait from iron- deficiency anemia. *American journal of clinical pathology*, 111(5), 676-682.

Farashi, S., & Harteveld, C. L. (2018). Molecular basis of α-thalassemia. *Blood Cells, Molecules, and Diseases, 70*, 43-53.

Farrell, C. J. L., & Nguyen, L. (2019). Indirect reference intervals: harnessing the power of stored laboratory data. *The Clinical biochemist Reviews*, *40*(2), 99.

Fibach, E., & Rachmilewitz, E. A. (2017). Pathophysiology and treatment of patients with beta-thalassemia–an update. *F1000Research, 6*.

Forget, B. G., & Bunn, H. F. (2013). Classification of the disorders of hemoglobin. *Cold Spring Harbor perspectives in medicine*, 3(2), a011684.

Foote, E. M., Sullivan, K. M., Ruth, L. J., Oremo, J., Sadumah, I., Williams, T. N., & Suchdev, P. S. (2013). Determinants of anemia among preschool children in rural, western Kenya. *The American journal of tropical medicine and hygiene*, *88*(4), 757.

Fucharoen, S., & Weatherall, D. J. (2012). The hemoglobin E thalassemias. *Cold Spring Harbor perspectives in medicine*, 2(8), a011734.

Galanello, R., & Origa, R. (2010). Beta-thalassemia. *Orphanet journal of rare diseases*, 5(1), 11

Galanello, R., & Cao, A. (2011). Alpha thalassemia. *Genetics in medicine*, 13(2), 83-88.

Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: a guide for non-statisticians. *International journal of endocrinology and metabolism*, 10(2), 486.

Gonçalves, B. P., Gupta, S., & Penman, B. S. (2016). Sickle haemoglobin, haemoglobin C and malaria mortality feedbacks. *Malaria journal*, 15(1), 26.

Gachie, R.N.M (2018). Haematological Reference Intervals for Adolescents and Adults in Nakuru County, Kenya (doctoral dissertation, department of pathology a thesis submitted for the award of the degree of Doctor of Philosophy (Cell haematology) in the school of medicine, Kenyatta university).

Grosse, S. D., Odame, I., Atrash, H. K., Amendah, D. D., Piel, F. B., & Williams, T. N. (2011). Sickle cell disease in Africa: a neglected cause of early childhood mortality. *American journal of preventive medicine*, 41(6), S398-S405.

Haase, V. H. (2013). Regulation of erythropoiesis by hypoxia-inducible factors. *Blood reviews*, *27*(1), 41-53.

Habibzadeh, F., Habibzadeh, P., & Yadollahie, M. (2016). On determining the most appropriate test cut-off value: the case of tests with continuous results. *Biochemia medica*, *26*(3), 297-307.

Hamali, H. A., & Saboor, M. (2019). Undiagnosed Hemoglobinopathies: A potential threat to the premarital screening program. *Pakistan Journal of Medical Sciences*, *35*(6), 1611.

Hamali, H. A. (2019). *Undiagnosed Hemoglobinopathies : A potential threat to the premarital screening program*. *35*(6), 1611–1615.

Hassan, M. (2017). The strategic shuffle: Ethnic geography, the internal security apparatus, and elections in Kenya. American Journal of Political Science, 61(2), 382-395.

Harewood, J., & Azevedo, A. M. (2021). Alpha Thalassemia. *StatPearls [Internet]*.

Hernandez, A. G., Kiyaga, C., Howard, T. A., Ssewanyana, I., Ndeezi, G., Aceng, J. R., & Ware, R. E. (2021). *Trends in sickle cell trait and disease screening in the Republic of Uganda , 2014 – 2019*. *26*(1), 23–32.

Hoffmann, J. J., Nabbe, K. C., & van den Broek, N. M. (2015). Effect of age and gender on reference intervals of red blood cell distribution width (RDW) and mean red cell volume (MCV). *Clinical Chemistry and Laboratory Medicine (CCLM)*, *53*(12).

Hoo, Z. H., Candlish, J., & Teare, D. (2017). What is an ROC curve? *Emergency Medicine Journal*, 34(6), 357-359.

Ilesanmi, O. O. (2010). Pathological basis of symptoms and crises in sickle cell disorder: implications for counseling and psychotherapy. Hematology reports, 2(1).

Jahn, W. T. (2011). The 4 basic ethical principles that apply to forensic activities are respect for autonomy, beneficence, nonmaleficence, and justice. *Journal of chiropractic medicine*, 10(3), 225.

Jahangiri, M., Rahim, F., & Malehi, A. S. (2019). Diagnostic performance of hematological discrimination indices to discriminate between βeta thalassemia trait and iron deficiency anemia and using cluster analysis: Introducing two new indices tested in Iranian population. *Scientific reports*, *9*(1), 1-13.

Jandial, A., Kumar, S., Bhalla, A., Sharma, N., Varma, N., & Varma, S. (2017). Elevated red cell distribution width as a prognostic marker in severe sepsis: a prospective observational study. *Indian journal of critical care medicine: peer-reviewed, official publication of Indian Society of Critical Care Medicine*, *21*(9), 552.

Kallner, A. (2018). Bayes' theorem, the ROC diagram and reference values: Definition and use in clinical diagnosis. *Biochemia medica*, *28*(1), 16-25.

Kaori, M., Litao, S., & Kamat, D. Back to Basics Red Blood Cell Distribution Width.

Kibaya, R. S., Bautista, C. T., Sawe, F. K., Shaffer, D. N., Sateren, W. B., Scott, P. T., ... & de Souza, M. S. (2008). Reference ranges for the clinical laboratory derived from a rural population in Kericho, Kenya. *PloS one*, *3*(10), e3327.

Kosiyo, P., Otieno, W., Gitaka, J., Munde, E. O., & Ouma, C. (2020). Association between haematological parameters and sickle cell genotypes in children with Plasmodium falciparum malaria resident in Kisumu County in Western Kenya. *BMC Infectious Diseases*, *20*(1), 1-11.

Karimi, M., Haghpanah, S., Azarkeivan, A., Matin, S., Safaei, A., & De Sanctis, V. (2021). Coronavirus disease 2019 (COVID-19) severity in patients with thalassemias: a nationwide Iranian experience. *Mediterranean Journal of Hematology and Infectious Diseases*, *13*(1).

Karita, E., Ketter, N., Price, M. A., Kayitenkore, K., Kaleebu, P., Nanvubya, A., ... & Kamali, A. (2009). CLSI-derived hematology and biochemistry reference intervals for healthy adults in eastern and southern Africa. *PloS one*, *4*(2), e4401

Kato, G. J., Piel, F. B., Reid, C. D., Gaston, M. H., Ohene-Frempong, K., Krishnamurti, L.,& Vichinsky, E. P. (2018). Sickle cell disease. *Nature Reviews Disease Primers*, 4(1), 1-22.

Khanam, B., Maurya, S., Gupta, P., & Shaikh, S. (2018). A Study on Hemoglobin Electrophoresis Profile in Patients with Moderate to Severe Anemia. *Journal of Contemporary Medical Research*, *5*(3), C21-C24.

Kifude, C. M. (2007). Prevalence of Sickle Cell and α-Thalassemia Traits in Children Enrolled in a Malaria Vaccine Clinical Trial, in Kombewa, Western Kenya (Doctoral dissertation, Kenyatta University).

Kim, M., Odame, I., Little, J. A., & Gurkan, U. A. (2017). *HHS Public Access*. *13*(12), 1073–1093. https://doi.org/10.1080/17434440.2016.1254038.Emerging

Kosiyo, P., Otieno, W., Gitaka, J., Munde, E. O., & Ouma, C. (2021). *Haematological abnormalities in children with sickle cell disease and non-severe malaria infection in western Kenya*. 1–11.

Koury, M. J., & Ponka, P. (2004). New insights into erythropoiesis: the roles. *Annu. Rev. Nutr*, *24*, 105-131.

Lalkhen, A. G., & McCluskey, A. (2008). Clinical tests: sensitivity and specificity. *Continuing education in anaesthesia critical care & pain*, *8*(6), 221-223.

Lee, D. K., In, J., & Lee, S. (2015). Standard deviation and standard error of the mean. *Korean journal of anesthesiology*, *68*(3), 220.

Li, C. M., Chao, C. T., Chen, S. I., Han, D. S., & Huang, K. C. (2020). Elevated red cell distribution width is independently associated with a higher frailty risk among 2,932 community-dwelling older adults. *Frontiers in Medicine*, *7*, 470.

Lippi, G., & Plebani, M. (2014). Red blood cell distribution width (RDW) and human pathology. One size fits all. *Clinical Chemistry and Laboratory Medicine (CCLM)*, *52*(9), 1247-1249.

Li, N., Zhou, H., & Tang, Q. (2017). Red blood cell distribution width: a novel predictive indicator for cardiovascular and cerebrovascular diseases. *Disease markers*, *2017*.

Luganda, E. S., Mermin, J., Kaharuza, F., Ulvestad, E., Were, W., Langeland, N., ... & Downing, R. (2004). Population-based hematologic and immunologic reference values for a healthy Ugandan population. *Clinical and diagnostic laboratory immunology*, *11*(1), 29-34.

Machogu, E. M., & Machado, R. F. (2018). How I treat hypoxia in adults with hemoglobinopathies and hemolytic disorders. *Blood, The Journal of the American Society of Hematology*, *132*(17), 1770-1780.

Macharia, A. W., Uyoga, S., Ndila, C., Nyutu, G., Makale, J., Tendwa, M., ... & Williams, T. N. (2019). The population dynamics of hemoglobins A, A2, F and S in the context of the hemoglobinopathies HbS and α-thalassemia in Kenyan infants. *haematologica*, *104*(5), e184.

Maekawa, S., & Kato, T. (2015). Diverse of erythropoiesis responding to hypoxia and low environmental temperature in vertebrates. *BioMed Research International*, *2015*.

Mairbäurl, H. (2018). Neocytolysis: how to get rid of the extra erythrocytes formed by stress erythropoiesis upon descent from high altitude. *Frontiers in Physiology*, *9*, 345.

Makani, J., Cox, S. E., Soka, D., Komba, A. N., Oruo, J., Mwamtemi, H., ... & Lowe, B. (2011). Mortality in sickle cell anemia in Africa: a prospective cohort study in Tanzania. PloS one, 6(2).

Mandal, P. K., & Kartthik, S. (2019). Autoimmune hemolytic anemia: an uncommon cause of elevated fetal hemoglobin. *Journal of Hematopathology*, *12*(1), 19-20.

Mansilla-Soto, J., Riviere, I., Boulad, F., & Sadelain, M. (2016). Cell and gene therapy for the beta-thalassemias: advances and prospects. *Human gene therapy*, *27*(4), 295-304.

Masinde, A. A., Kwambai, K. T., & Wambani, N. H. (2011). Evaluation of tomato (Lycopersicon esculentum L.) variety tolerance to foliar diseases at Kenya Agricultural Research Institute Centre-Kitale in Northwest Kenya. *African Journal of Plant Science*, *5*(11), 676-681.

Matos, J. F., Dusse, L., Borges, K. B., de Castro, R. L., Coura-Vital, W., & Carvalho, M. D. G. (2016). A new index to discriminate between iron deficiency anemia and thalassemia trait. *Revista brasileira de hematologiae hemoterapia*, 38(3), 214-219.

Matsushita, N., Kim, Y., Fook, C., Ng, S., Moriyama, M., Igarashi, T., Yamamoto, K., Otieno, W., & Minakawa, N. (2019). *Di ff erences of Rainfall – Malaria Associations in Lowland and Highland in Western Kenya*. 1–13.

Matsushita, N., Kim, Y., Ng, C. F. S., Moriyama, M., Igarashi, T., Yamamoto, K., ... & Hashizume, M. (2019). Differences of rainfall–malaria associations in lowland and highland in western Kenya. *International journal of environmental research and public health*, *16*(19), 3693.

Mkocha, N. N., Tebuka, E., Ambrose, E. E., Msemwa, B., & Silago, V. (2018). Utility of paper-based sickle cell test compared to sodium metabisulfite sickling test using hemoglobin electrophoresis as a gold standard at Bugando Medical Center, Mwanza. *Journal of Applied Hematology*, *9*(3), 91.

Miri-Moghaddam, E., & Sargolzaie, N. (2014). Cut off determination of discrimination indices in differential diagnosis between iron deficiency anemia and β-thalassemia minor. *International journal of hematology-oncology and stem cell research*, 8(2), 27.

Modell, B., & Darlison, M. (2008). Global epidemiology of haemoglobin disorders and derived service indicators. Bulletin of the World Health Organization, 86, 480-487.

Morris, J., Dunn, D., Beckford, M., Grandison, Y., Mason, K., Higgs, D., ... & Serjeant, G. (1991). The haematology of homozygous sickle cell disease after the age of 40 years. *British journal of haematology*, *77*(3), 382-385.

Muncie Jr, H. L., & Campbell, J. S. (2009). Alpha and beta thalassemia. *American family physician*, 80(4), 339-344.

Mutua, B. M., Sowayi, G., & Okoth, P. (2022). Red cell distribution width as a surrogate marker of haemoglobinopathies in western Kenya. *African Journal of Laboratory Medicine*, *11*(1).

Nahm, F. S. (2016). Nonparametric statistical tests for the continuous data: the basic concept and the practical use. *Korean journal of anesthesiology*, 69(1), 8

Nam, J. M. (1992). Sample size determination for case-control studies and the comparison of stratified and unstratified analyses. Biometrics, 389-395.

Nantume, A. S. (2019). Hydroxyurea use and Plasmodium falciparum prevalence among children with sickle cell anemia in Homa Bay, Kenya; a cross sectional study.

Needs, T., Gonzalez-Mosquera, L. F., & Lynch, D. T. (2020). Beta thalassemia. *StatPearls [Internet]*.

Ndeezi, G., Kiyaga, C., Hernandez, A. G., Munube, D., Howard, T. A., Ssewanyana, I., ... & Aceng, J. R. (2016). Burden of sickle cell trait and disease in the Uganda Sickle Surveillance Study (US3): a cross-sectional study. *The Lancet Global Health*, 4(3), e195-e200.

Nienhuis, A. W., & Nathan, D. G. (2012). Pathophysiology and clinical manifestations of the β-thalassemias. *Cold Spring Harbor perspectives in medicine*, *2*(12), a011726.

Nishad, A. N., de Silva, I. S., Perera, H. L., Pathmeswaran, A., Kastutiratne, K. A., & Premawardhena, A. P. (2014). Role of Red Cell Distribution Width in Screening for Hb E Trait in Population Screening for Hemoglobin Disorders. *Journal of pediatric hematology/oncology*, 36(8), e490-e492

Nkya, S., Mgaya, J., Urio, F., Makubi, A., Thein, S. L., Menzel, S., ... & Makani, J. (2017). Fetal hemoglobin is associated with peripheral oxygen saturation in sickle cell disease in Tanzania. *EBioMedicine*, *23*, 146-149.

Njoroge, J., Shiluli, M., Nyaga, J., Kinuthia, R., Mogaka, J., Muoni, T., ... & Duncan, A. J. (2020). Using the LegumeCHOICE tool to support legume use on smallholder farms in Kisii and Migori counties in Kenya. *ILRI Project Report*.

Odhiambo, C., Oyaro, B., Odipo, R., Otieno, F., Alemnji, G., Williamson, J., & Zeh, C. (2015). Evaluation of locally established reference intervals for hematology and biochemistry parameters in Western Kenya. *PloS one*, *10*(4), e0123140.

Ojowi, M. O., Ogidi, R. O., Obanyi, J. N., & Owango, M. O. (2001). Smallholder dairy production and marketing in Kisii, Nyamira and Rachuonyo Districts: a review of literature.

Okwi, A. L., Byarugaba, W., Parkes, A., & Ocaido, M. (2010). The reliability of sickling and solubility tests and peripheral blood film method for sickle cell disease screening at district health centers in Uganda. *Clinics in Mother and Child Health*, *7*(1).

Olujohungbe, A., Haem, D., & Garvey, B. (1993). Red cell distribution width (RDW) in sickle cell disease. *Clinical & Laboratory Haematology*, *15*(1), 72-73.

Omukunda, E., Githeko, A., Ndong'a, M. F., Mushinzimana, E., & Yan, G. (2012). Effect of swamp cultivation on distribution of anopheline larval habitats in Western Kenya. *Journal of vector borne diseases*, *49*(2), 61.

Ozarda, Y., Ichihara, K., Jones, G., Streichert, T., & Ahmadian, R. (2021). Comparison of reference intervals derived by direct and indirect methods based on compatible datasets obtained in Turkey. *Clinica Chimica Acta*, *520*, 186-195.

Parikh, R., Mathai, A., Parikh, S., Sekhar, G. C., & Thomas, R. (2008). Understanding and using sensitivity, specificity and predictive values. *Indian journal of ophthalmology*, *56*(1), 45.

Parikh, N. I., & Vasan, R. S. (2007). Assessing the clinical utility of biomarkers in medicine.

Patel, K. V., Ferrucci, L., Ershler, W. B., Longo, D. L., & Guralnik, J. M. (2009). Red blood cell distribution width and the risk of death in middle-aged and older adults. *Archives of internal medicine*, *169*(5), 515-523.

Płaczkowska, S., Terpińska, M., & Piwowar, A. (2022). Establishing laboratory-specific reference intervals for TSH and fT4 by use of the indirect Hoffman method. *Plos one*, *17*(1), e0261715.

Piccin, A., Murphy, C., Eakins, E., Rondinelli, M. B., Daves, M., Vecchiato, C., ... & Smith, O. P. (2019). Insight into the complex pathophysiology of sickle cell anaemia and possible treatment. *European journal of haematology*, *102*(4), 319-330.

Piel, F. B., & Weatherall, D. J. (2015). Sickle-cell disease: a call to action.

Piety, N. Z., Yang, X., Kanter, J., Vignes, S. M., George, A., & Shevkoplyas, S. S. (2016). Validation of a low-cost paper-based screening test for sickle cell anemia. *PLoS One*, *11*(1), e0144901).

Piriyakhuntorn, P., Tantiworawit, A., Rattanathammethee, T., Chai-Adisaksopha, C., Rattarittamrong, E., & Norasetthada, L. (2018). The role of red cell distribution width in the differential diagnosis of iron deficiency anemia and non-transfusion dependent thalassemia patients. *Hematology Reports*, *10*(3), 7605.

Qurtom, H. A., Al-Saleh, Q. A., Lubani, M. M., Hassanein, A., Kaddoorah, N., Qurtom, M. A., & Al-Sheikh, T. (1989). The value of red cell distribution width in the diagnosis of anaemia in children. *European journal of pediatrics*, 148(8), 745-748.

Rakama, S. O., Obiri, J. F., & Mugalavai, E. M. (2017). Evaluation of land use change pattern of Kajulu-Riat hill peri-urban area near Kisumu City, Kenya. *IJSRIT*, *4*(7), 42-52.

Ramby, A. L., Goodman, D. M., Wald, E. L., & Weiss, S. L. (2015). Red blood cell distribution width as a pragmatic marker for outcome in pediatric critical illness. PloS one, 10(6).

Ribeil, J. A., Hacein-Bey-Abina, S., Payen, E., Magnani, A., Semeraro, M., Magrin, E., ... & Cavazzana, M. (2017). Gene therapy in a patient with sickle cell disease. *New England Journal of Medicine*, *376*(9), 848-855.

Roberts, G. T., & EL Badawi, S. B. (1985). Red blood cell distribution width index in some hematologic diseases. *American journal of clinical pathology*, *83*(2), 222-226.

Ropero, P., González, F. A., Nieto, J. M., Torres-Jiménez, W. M., & Benavente, C. (2022). β-Thalassemia Intermedia: Interaction of α-Globin Gene Triplication with β-thalassemia heterozygous in Spain. *Frontiers in Medicine*, 756.

Roy, Y., Banville, H., Albuquerque, I., Gramfort, A., Falk, T. H., & Faubert, J. (2019). Deep learning-based electroencephalography analysis: a systematic review. *Journal of neural engineering*, *16*(5), 051001.

Ruopp, M. D., Perkins, N. J., Whitcomb, B. W., & Schisterman, E. F. (2008). Youden Index and optimal cut-point estimated from observations affected by a lower limit of detection. Biometrical Journal: *Journal of Mathematical Methods in Biosciences*, 50(3), 419-430.

Sadiq, S., Khalid, M. U., Ullah, S., Aslam, W., Mehmood, A., Choi, G. S., & On, B. W. (2021). Classification of β-Thalassemia Carriers From Red Blood Cell Indices Using Ensemble Classifier. *IEEE Access*, *9*, 45528-45538.

Sarah, B., Sheikh, K., & Shah, T. (2018). Red Cell Distribution Width is Early Marker for Detection of Iron Deficiency Anemia during Pregnancy. Journal of Liaquat University of Medical & Health Sciences, 17(03), 165-169.

Saulnier, K. M., & Dupras, C. (2017). Race in the postgenomic era: Social epigenetics calling for interdisciplinary ethical safeguards. *The American Journal of Bioethics*, *17*(9), 58-60.

Santhosh, S R & Sampath. (2011). Hemoglobin Electrophoresis as a Diagnostic Tool in Anaemias. *Indian Journal of Aerospace Medicine* 55(2), 2011. IJASM 2011; 55(2): 16-20

Sawaimul, K. D., Iqbal, M. B., Sawaimul, V. D., Kambale, T., & Hanmante, R. (2018). Study to identify the role of high-performance liquid chromatography in detecting haemoglobinopathies in antenatal patients. *Indian Journal of Pathology and Oncology*, 5(1), 6-11.

Shah, N., Khonglah, Y., Raphael, V., Swer, B., Nath, C., & Singh, A. S. (2021). *Antenatal Screening for Hemoglobinopathies with HPLC*. *October 2018*. https://doi.org/10.24321/2454.8642.201809

Scherer, P. S., Moraes, D., Munhoz, T. P., & Sgnaolin, V. (2015). New red blood cell and reticulocyte parameters and reference values for healthy individuals and in

chronic kidney disease. *Jornal Brasileiro de Patologia e Medicina Laboratorial*, *51*(2), 77-84.

Sharma, A., Sharma, M., & Sharma, V. (2016). Evaluation of red cell distribution width in the diagnosis of iron deficiency anemia. Int J Res Med Sci, 4(9), 3733-3736.

Šimundić, A. M. (2009). Measures of diagnostic accuracy: basic definitions. *ejifcc*, *19*(4), 203.

Song, B., Shi, P., Xiao, J., Song, Y., Zeng, M., Cao, Y., & Zhu, X. (2020). Utility of red cell distribution width as a diagnostic and prognostic marker in non-small cell lung cancer. *Scientific Reports*, *10*(1), 1-7.

Subhashree, A. R., Parameaswari, P. J., Shanthi, B., Revathy, C., & Parijatham, B. O. (2012). The reference intervals for the haematological parameters in healthy adult population of chennai, southern India. *Journal of clinical and diagnostic research:* JCDR, 6(10), 1675

Suchdev, P. S., Ruth, L. J., Earley, M., Macharia, A., & Williams, T. N. (2014). The burden and consequences of inherited blood disorders among young children in western Kenya. Maternal & child nutrition, 10(1), 135-144.

Sundd, P., Gladwin, M. T., & Novelli, E. M. (2019). Pathophysiology of sickle cell disease. *Annual review of pathology*, *14*, 263.

Taher, A. T., Musallam, K. M., & Cappellini, M. D. (2021). β-Thalassemias. *New England Journal of Medicine*, *384*(8), 727-743.

Tariq, M., Zafar, Q., Waheed, W., Shuaib, A., Qaiser, J., & Malik, M. N. (2019). Haemoglobinopathies in Referred Cases for Haemoglobin Electrophoresis in a Tertiary Care Teaching Hospital. *Journal of Rawalpindi Medical College*, *23*(S-2).

Thame, M., GRANDISON, Y., MASON, K., Thompson, M., HIGGS, D., MORRIS, J., ... & SERJEANT, G. (1991). The red cell distribution width in sickle cell disease— is it of clinical value? Clinical & Laboratory Haematology, 13(3), 229-237.

Titcomb, C. P. (2017). Red cell distribution width (RDW): An underappreciated marker for increased mortality. *On the risk*, *33*(1), 30-4

Tluway, F., & Makani, J. (2017). Sickle cell disease in Africa: an overview of the integrated approach to health, research, education and advocacy in Tanzania, 2004–2016. *British journal of haematology*, 177(6), 919-929.

Tonbul, A., Tayman, C., Catal, F., Kara, S., & Tatli, M. M. (2011). Red cell distribution width (RDW) in the newborn: normative data. J*ournal of clinical laboratory analysis*, 25(6), 422-425.

Trent, R. J. (2006). Diagnosis of the haemoglobinopathies. *Clinical Biochemist Reviews*, 27(1), 27.

Uyoga, S., Macharia, A. W., Mochamah, G., Ndila, C. M., Nyutu, G., Makale, J., ... & Williams, T. N. (2019). The epidemiology of sickle cell disease in children recruited in infancy in Kilifi, Kenya: a prospective cohort study. *The Lancet Global Health*, *7*(10), e1458-e1466.

Valavi, E., Ansari, M. J. A., & Zandian, K. (2010). How to reach rapid diagnosis in sickle cell disease?. *Iranian journal of pediatrics*, *20*(1), 69.

Vehapoglu, A., Ozgurhan, G., Demir, A. D., Uzuner, S., Nursoy, M. A., Turkmen, S., & Kacan, A. (2014). Hematological indices for differential diagnosis of Beta thalassemia trait and iron deficiency anemia. *Anemia*, *2014*.

Vinjamur, D. S., Bauer, D. E., & Orkin, S. H. (2018). Recent progress in understanding and manipulating haemoglobin switching for the haemoglobinopathies. *British journal of haematology*, *180*(5), 630-643.

Viswanath, D., Hegde, R., Murthy, V., Nagashree, S., & Shah, R. (2001). Red cell distribution width in the diagnosis of iron deficiency anemia. *The Indian Journal of Pediatrics*, 68(12), 1117-1119.

Vlaski, M., Lafarge, X., Chevaleyre, J., Duchez, P., Boiron, J. M., & Ivanovic, Z. (2009). Low oxygen concentration as a general physiologic regulator of erythropoiesis beyond the EPO-related downstream tuning and a tool for the optimization of red blood cell production ex vivo. *Experimental Hematology*, *37*(5), 573-584.

Wacholder, S., McLaughlin, J. K., Silverman, D. T., & Mandel, J. S. (1992). Selection of controls in case-control studies: I. Principles. *American journal of epidemiology*, 135(9), 1019-1028.

Watenga, E. K. (2020). *The prevalence of sickle cell trait among adolescents in Bungoma County, Kenya* (Doctoral dissertation, Moi University).

Weatherall, D. J., & Clegg, J. B. (2001). Inherited haemoglobin disorders: an increasing global health problem. *Bulletin of the World Health Organization*, 79, 704-712.

Weatherall, D. (2011). The inherited disorders of haemoglobin: an increasingly neglected global health burden. *The Indian journal of medical research*, 134(4), 493

Webster, P. A. U. L., & Castro, O. S. W. A. L. D. O. (1986). Red cell distribution width in sickle cell disease. *Annals of Clinical & Laboratory Science*, 16(4), 274-277

Wickramaratne, K. A. C., & Wijewickrama, D. C. (2021). Screening for beta-thalassemia trait; applicability of red cell indices and parameters–A study in Sri Lanka. *International Journal of Health Sciences*, *15*(1), 29.

Wilson, C. I., Hopkins, P. L., Cabello-Inchausti, B., Melnick, S. J., & Robinson, M. J. (2000). The peripheral blood smear in patients with sickle cell trait: a morphologic observation. Laboratory Medicine, 31(8), 445-447

Yan, R., Li, K., Lv, Y., Peng, Y., Halm-Lutterodt, V., Song, W., ... & Ni, X. (2022). Comparison of reference distributions acquired by direct and indirect sampling techniques: exemplified with the Pediatric Reference Interval in China (PRINCE) study. *BMC medical research methodology*, *22*(1), 1-10.

Zivot, A., Lipton, J. M., Narla, A., & Blanc, L. (2018). Erythropoiesis: insights into pathophysiology and treatments in 2017. *Molecular Medicine*, *24*(1), 1-15.

APPENDICES

Appendix I: Western Kenya Map

Appendix II: Study Approval from Postgraduate Studies

MASINDE MULIRO UNIVERSITY OF SCIENCE AND TECHNOLOGY (MMUST)

Tel: 056-30870
Fax: 056-30153
E-mail: directord s mmust.ac.ke
Website: www.mmust.ac.ke

P.O Box 190
Kakamega – 50100
Kenya

Directorate of Postgraduate Studies

Ref: MMU/COR: 509099

Date: 14th September, 2020

Benard Munguti Mutua,
HML/G/02/2016,
P.O. Box 190-50100,
KAKAMEGA.

Dear Mr. Mutua,

RE: APPROVAL OF PROPOSAL

I am pleased to inform you that the Directorate of Postgraduate Studies has considered and approved your masters proposal entitled: ***"Red Cell Distribution Width as a Surrogate Marker of Haemoglobinopathies among Patients Attending Aga Khan Hospital Kisumu, Western Kenya"*** and appointed the following as supervisors:

1. Mr. George Sowayi - SPHBST, MMUST
2. Dr. Okoth Patrick - SONAS, MMUST

You are required to submit through your supervisor(s) progress reports every three months to the Director Postgraduate Studies. Such reports should be copied to the following: Chairman, School of Public Health, Biomedical Sciences and Technology Graduate Studies Committee and Chairman, Medical Laboratory Sciences Department. Kindly adhere to research ethics consideration in conducting research.

It is the policy and regulations of the University that you observe a deadline of two years from the date of registration to complete your master's thesis. Do not hesitate to consult this office in case of any problem encountered in the course of your work.

We wish you the best in your research and hope the study will make original contribution to knowledge.

Yours S

DEAN
SCHOOL OF GRADUATE STUDIES
MASINDE MULIRO UNIVERSITY OF
SCIENCE & TECHNOLOGY
Date: .....Sign:.....

Prof. John Obiri
DIRECTOR, DIRECTORATE OF POSTGRADUATE STUDIES

Appendix III: Institutional Ethics Review Committee (IERC)

MASINDE MULIRO UNIVERSITY OF SCIENCE AND TECHNOLOGY
Tel: 056-31375
Fax: 056-30153
E-mail: ierc@mmust.ac.ke
Website: www.mmust.ac.ke
P. O. Box 190-50100
Kakamega, Kenya

Institutional Ethics Review Committee (IERC)

Ref: MMU/COR: 403012 Vol 3 (03) **Date:** 14th October, 2020

Benard Mutua
Masinde Muliro University of Science and Technology,
P.O. Box 190-50100.
KAKAMEGA.

Dear Mr. Mutua,

RE: Red cell distribution width as a surrogate marker of haemoglobinopathies among patients attending Aga Khan hospital Kisumu, Western Kenya - MMUST/IERC/128/2020

Thank you for submitting your proposal entitled as above for initial review. This is to inform you that the committee conducted the initial review and approved (with no further revisions) the above Referenced application for one year.
This approval is valid from **14th October 2020** through to **14th October 2021**. Please note that authorization to conduct this study will automatically expire on **14th September 2021**. If you plan to continue with data collection or analysis beyond this date please submit an application for continuing approval to the MMUST IERC by **14th September 2021.**

Approval for continuation of the study will be subject to submission and review of an annual report that must reach the MMUST IERC Secretariat by **14th September 2021.** You are required to submit any amendments to this protocol and any other information pertinent to human participation in this study to MMUST IERC prior to implementation.

Please note that any unanticipated problems or adverse effects/event resulting from the conduct of this study must be reported to MMUST IERC. Also note that you are required to seek for research permit from NACOSTI prior to the initiation of the study.

.

Yours faithfully,

Dr. Gordon Nguka (PhD)
Chairman, Institutional Ethics Review Committee

Copy to:
- The Secretary, National Bio-Ethics Committee
- Vice Chancellor
- DVC (PR&I)

Appendix IV: Research License from NACOSTI

REPUBLIC OF KENYA

NATIONAL COMMISSION FOR SCIENCE,TECHNOLOGY & INNOVATION

Ref No: **407653**

Date of Issue: **09/November/2020**

RESEARCH LICENSE

This is to Certify that Mr.. BENARD MUNGUTI MUTUA of Masinde Muliro University of Science and Technology, has been licensed to conduct research in Bungoma, Busia, Homabay, Kakamega, Kisii, Kisumu, Migori on the topic: RED CELL DISTRIBUTION WIDTH AS A SURROGATE MARKER OF HAEMOGLOBINOPATHIES AMONG PATIENTS ATTENDING AGA KHAN HOSPITAL KISUMU, WESTERN KENYA for the period ending : 09/November/2021.

License No: **BAHAMAS ABS/P/20/7511**

407653

Applicant Identification Number

Director General
NATIONAL COMMISSION FOR SCIENCE,TECHNOLOGY & INNOVATION

Verification QR Code

NOTE: This is a computer generated License. To verify the authenticity of this document, Scan the QR Code using QR scanner application.

THE SCIENCE, TECHNOLOGY AND INNOVATION ACT, 2013

The Grant of Research Licenses is Guided by the Science, Technology and Innovation (Research Licensing) Regulations, 2014

CONDITIONS

1. The License is valid for the proposed research, location and specified period
2. The License any rights thereunder are non-transferable
3. The Licensee shall inform the relevant County Director of Education, County Commissioner and County Governor before commencement of the research
4. Excavation, filming and collection of specimens are subject to further necessary clearence from relevant Government Agencies
5. The License does not give authority to tranfer research materials
6. NACOSTI may monitor and evaluate the licensed research project
7. The Licensee shall submit one hard copy and upload a soft copy of their final report (thesis) within one year of completion of the research
8. NACOSTI reserves the right to modify the conditions of the License including cancellation without prior notice

National Commission for Science, Technology and Innovation
off Waiyaki Way, Upper Kabete,
P. O. Box 30623, 00100 Nairobi, KENYA
Land line: 020 4007000, 020 2241349, 020 3310571, 020 8001077
Mobile: 0713 788 787 / 0735 404 245
E-mail: dg@nacosti.go.ke / registry@nacosti.go.ke
Website: www.nacosti.go.ke

Appendix V: Aga Khan Hospital, Kisumu Research Permit

The Aga Khan Hospital, Kisumu

An institution of the Aga Khan Health Service

PO. Box 530, Kisumu, Kenya

Telephone: +254 57-2024374 ,0722 203 622

Fax: 057 2024 412

E-mail: ksm.admin@akhskenya.org

ADM/007/089

12th April 2021

TO WHOM IT MAY CONCERN

Dear Sir/ Madam

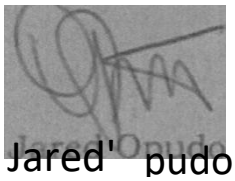
RE: PERMISSION TO COLLECT DATA AT AGA KHAN HOSPITAL KISUMU FOR BENARD MUNGUTI MUTUA

The above mentioned who is pursuing master's degree in medical Laboratory Technology' Hematology and Transfusion Science) at Masinde Muliro University of Science and Technology has been granted permission to collect data at our institution in partial fulfilment of his academic requirement.

His research topic is: 'Red Cell Distribution Width as A Surrogate Marker of Haemoglobinopathies Among Patients Attending Aga Khan Hospital Kisumu' He is expected to maintain high standards of professionalism and strictly adhere to research ethics guidelines in the conduct of his study.

Kindly accord him the support he may need.

Yours sincerelv,



Jared' pudo

(Secretary, Ethics Committee-AKHK)

Mobile: 0721 749848

Appendix VI Testing Data Normality

Testing of the normality of the data was done using visual methods of inspection (Q-Q plots and plot boxes) and two normality tests that was Kolmogorov-Smirnov (K-S) test and Shapiro-Wilk test which are normality tests supplementary to the graphical assessment of normality. According to previous findings, Kolmogorov-Smirnov (K-S) test and Shapiro-Wilk test are the accurate normality tests for data that can be used to samples more than 50. The use of the two normality tests was very important to compare their precision of p-value where p-value of < 0.05 indicates none normally distributed data and a $p>0.05$ indicates normally distributed data however, Shapiro-Wilk is more superior as compared to K-S (Ghasemi & Zahediasl 2012). The study variables in the present study had statistical significance ($p<0.0001$) indicating non-normally distributed *(refer to appendix VII)* thus validating the use of non-parametric tool of measurement. Therefore, the present study used Kruskal-Wallis and Dunn test post hock as the analytical tools.

Appendix VII: Kolmogorov-Smirnov and Shapiro-Wilk test analysis for normality test

	STUDY GROUPS	Kolmogorov-Smirnov			Shapiro-Wilk		
		Statistic	Df	Sig.	Statistic	df	Sig.
RDW- CV(%)	CONTROL	.122	205	.000	.881	205	.000
	CASE	.072	207	.010	.953	207	.000
HBA1(%)	CONTROL	.152	205	.000	.897	205	.000
	CASE	.287	207	.000	.787	207	.000
HBA2(%)	CONTROL	.165	205	.000	.903	205	.000
	CASE	.204	207	.000	.478	207	.000
HB S(%)	CONTROL	.516	205	.000	.049	205	.000
	CASE	.166	207	.000	.919	207	.000
HB F(%)	CONTROL	.273	205	.000	.582	205	.000
	CASE	.192	207	.000	.831	207	.000

Appendix VIII: calculation of sensitivity, specifity, predictive values and likelihood ratios

$$\text{Sensitivity} = \frac{\text{True Positive (RDW} > 19.9)}{\text{True Positive (RDW} > 19.9) + \text{False Positive (RDW} < 19.9)}$$

$$= \frac{136}{136 + 111}$$

$$= 55.1\% \text{ or } 0.551$$

$$\text{Specivity} = \frac{\text{True Negative (RDW} < 19.9)}{\text{True Negative (RDW} < 19.9) + \text{False Negative (RDW} > 19.9)}$$

$$= \frac{227}{227 + 14}$$

$$= 94.2\% \text{ or } 0.942$$

$$\text{PPV} = \frac{True\ positive\ (RDW > 19.9)}{True\ positives\ (RDW > 19.9) + False\ negative\ (RDW > 19.9)}$$

$$PPV = \frac{136}{136 + 14}$$

$$=90.7\%$$

$$NPV = \frac{\text{True Negative (RDW} < 19.9)}{\text{True Negative for SCD (RDW} < 19.9) + \text{False Positive for SCD (RDW} < 19.9)}$$

$$= \frac{227}{227 + 111}$$

$$= 67.2\%$$

$$Positive\ likelihood\ ratio\ = \frac{Sensitivity}{1 - Specifity}$$

$$= \frac{0.551}{0.551 - 0.942}$$

$$= 9.5$$

$$Negative\ likelihood\ ratio = \frac{(1 - Sensitivity)}{Specivity}$$

$$= \frac{(1 - 0.551)}{0.942}$$

$$= 0.476$$

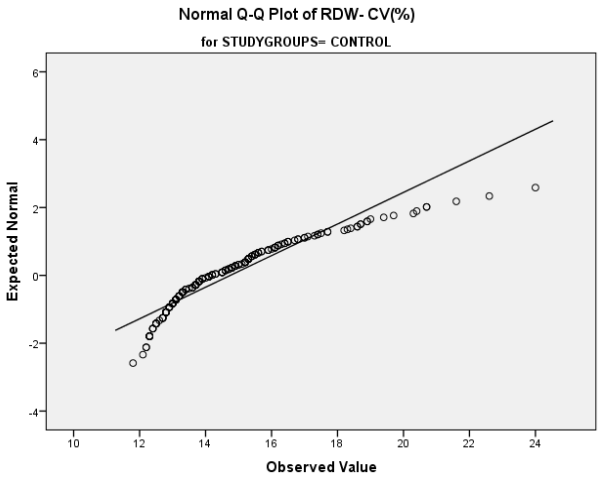
Appendix IX: RDW & Abnormal Hb phenotype 2×2 contingency table

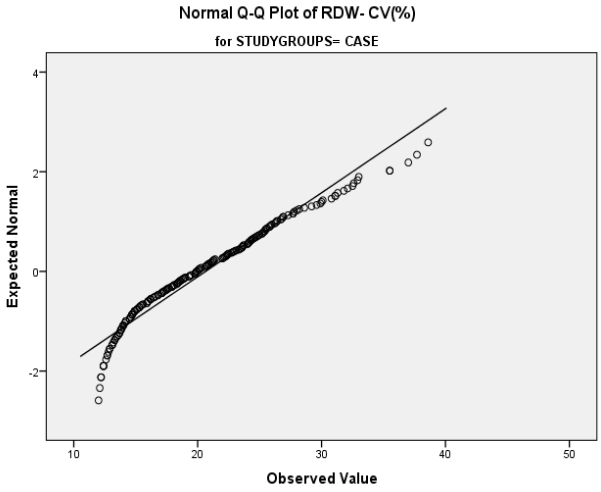
			Haemoglobinopathies		
			hemoglobinopathy Positive	hemoglobinopathy Negative	Total
RDW-CV (%)	>19.9 POSITIVE	Count	136	14	150
		% within RDW	90.7%	9.3%	100.0%
	<19.9 NEGATIVE	Count	111	227	338
		% within RDW	32.8%	67.2%	100.0%
Total		Count	247	241	488
		% within RDW	50.6%	49.4%	100.0%

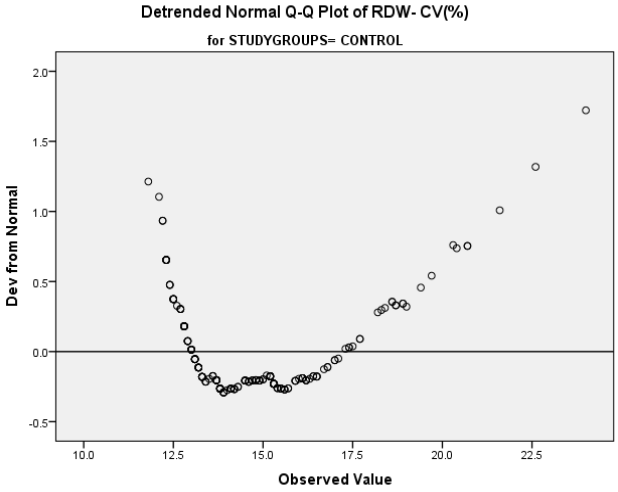
Appendix X: Visual normality plots

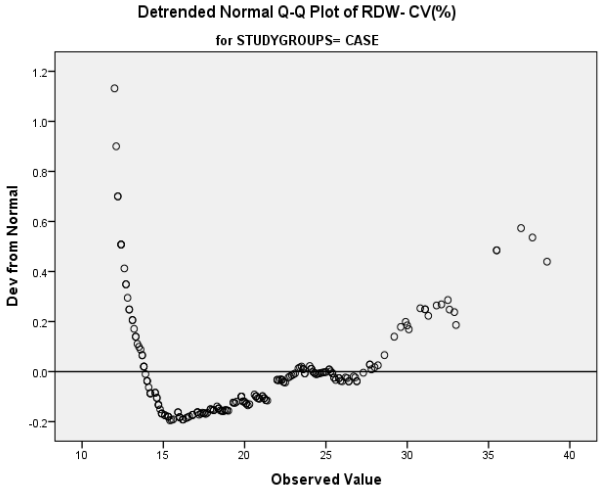
Q-Q plots were generated by SPSS which are observed values with plots quantiles of data sets instead of every individual score and when the data was normally distributed the results were on a straight diagonal line which was not observed on the present study. The boxplot showed the median as a horizontal line inside the box and the inter-quartile range as the length of the box. The whiskers line demonstrated 1.5 times inter-quartile range from either ends of the box. The more the whisker line was located at the center the more the data was approximately normally distributed, and the more the line was asymmetrical the more outliers representing none normally distributed data. Only Hb A_2 in the present study that demonstrated symmetrically whisker line in boxplots.

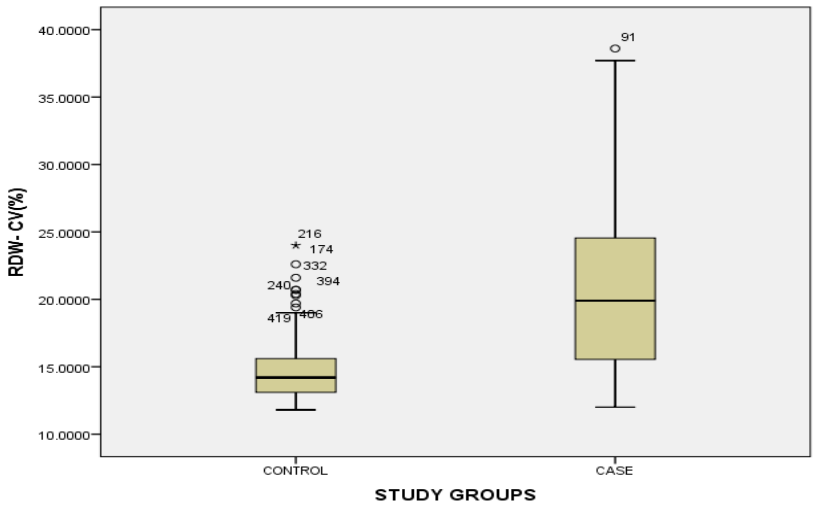
RDW Q-Q Plots and Box Plots for the Study Groups



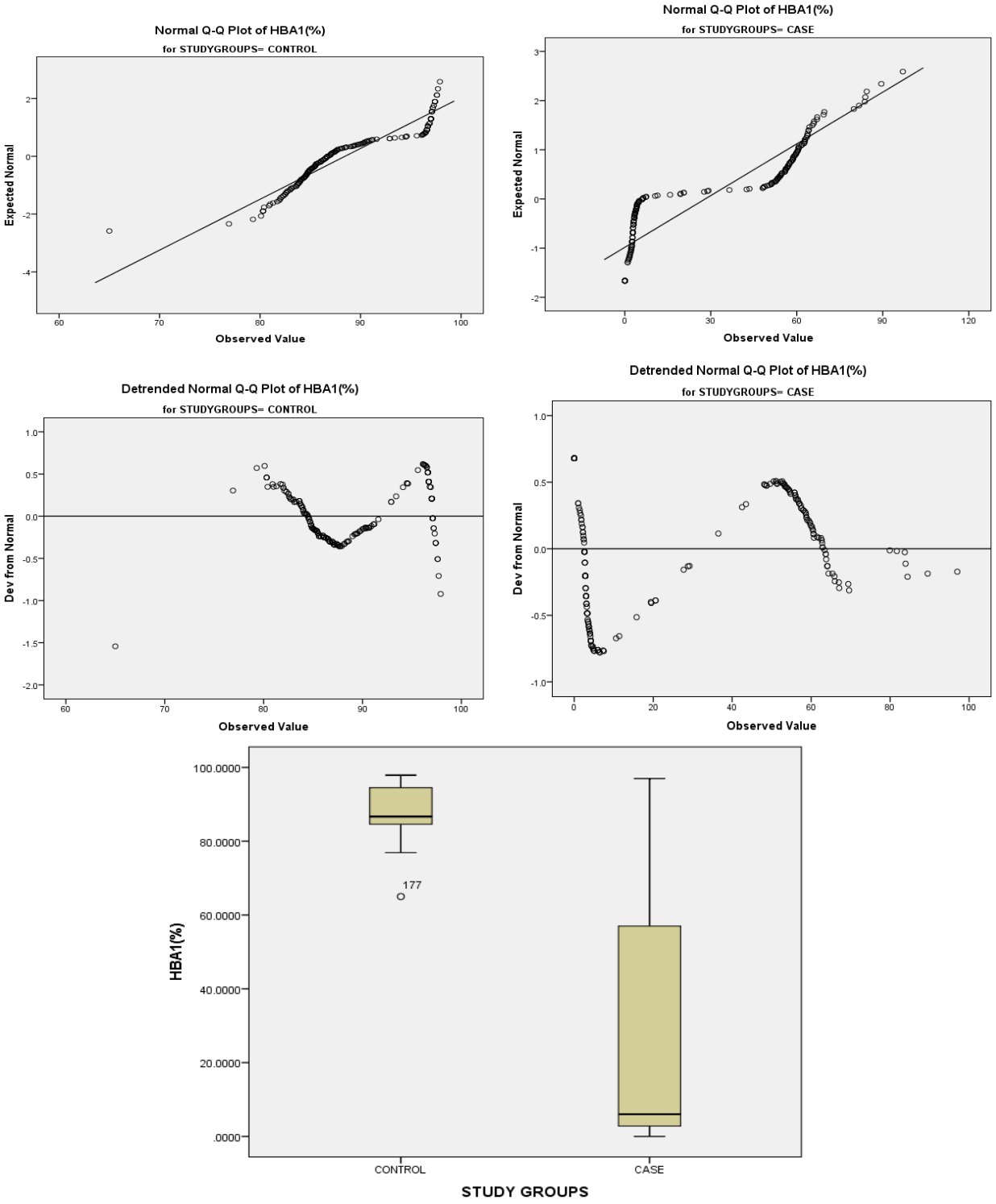




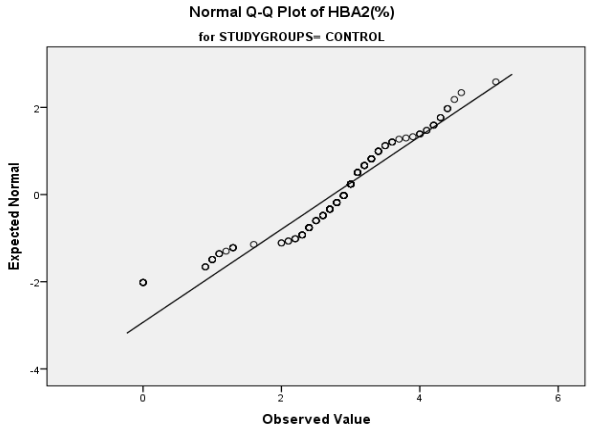


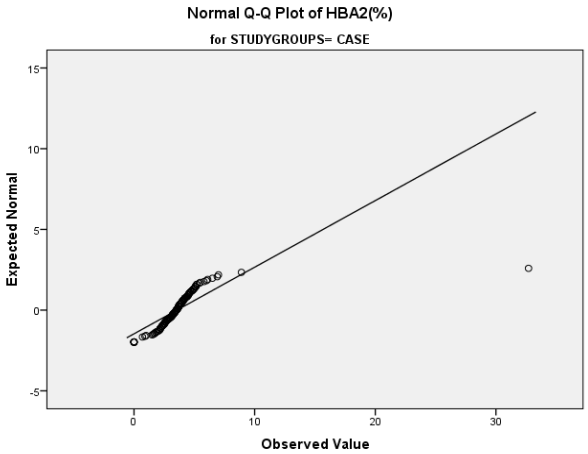


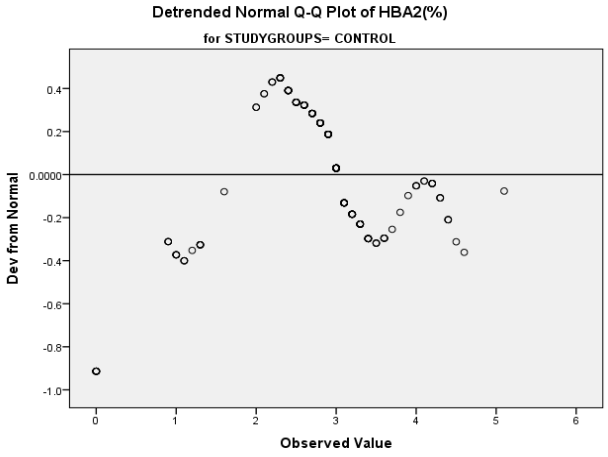
HBA1 Q-Q Plots and Box Plots for the Study Groups

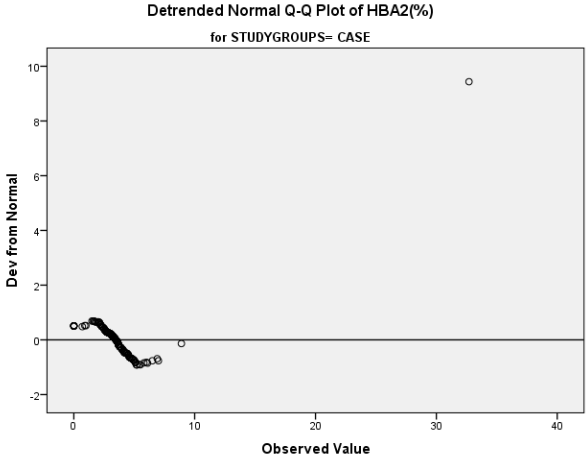


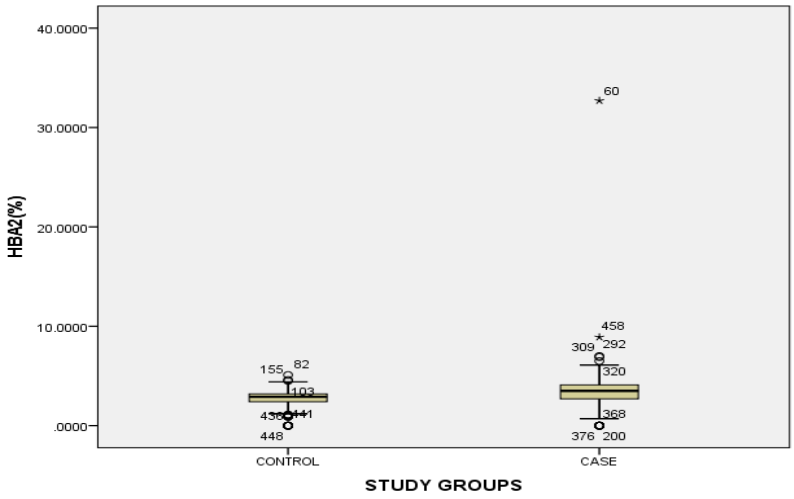
HBA2 Q-Q Plots and Box Plots for the Study Groups



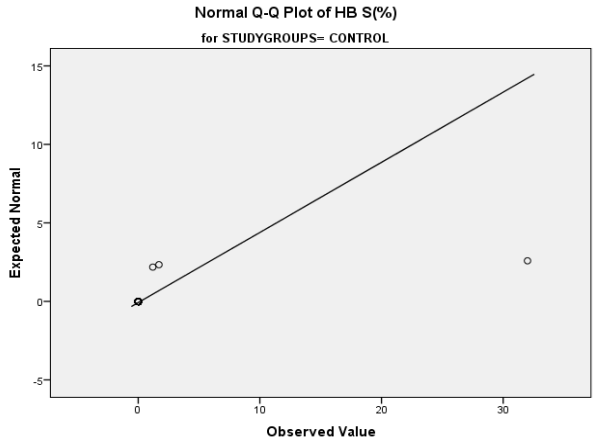


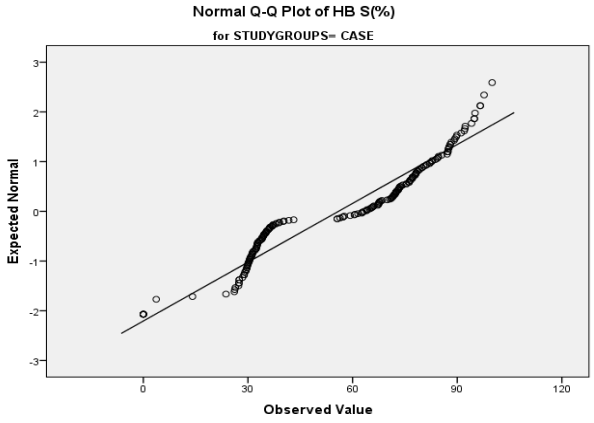


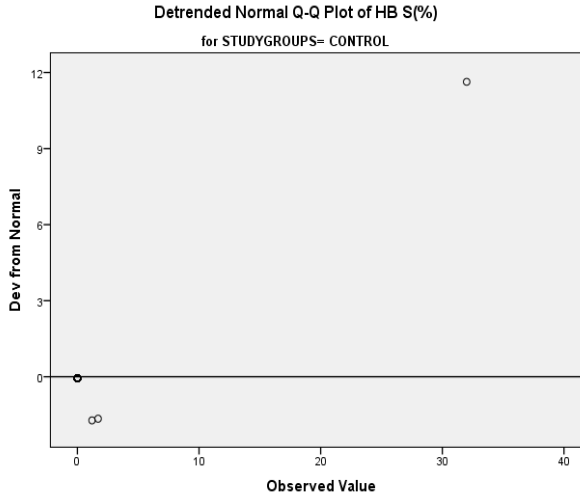


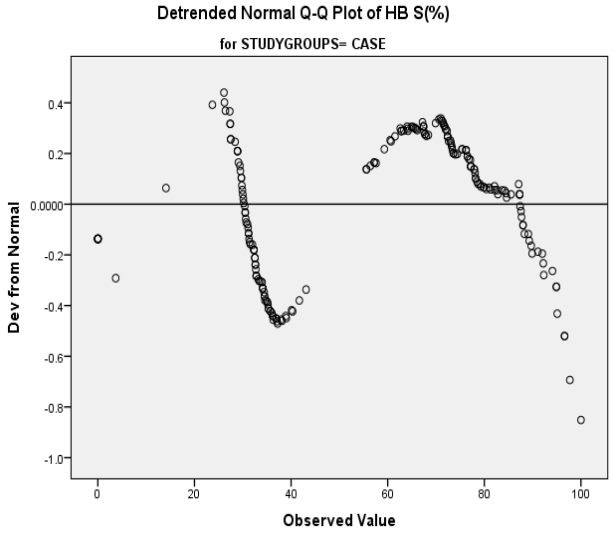


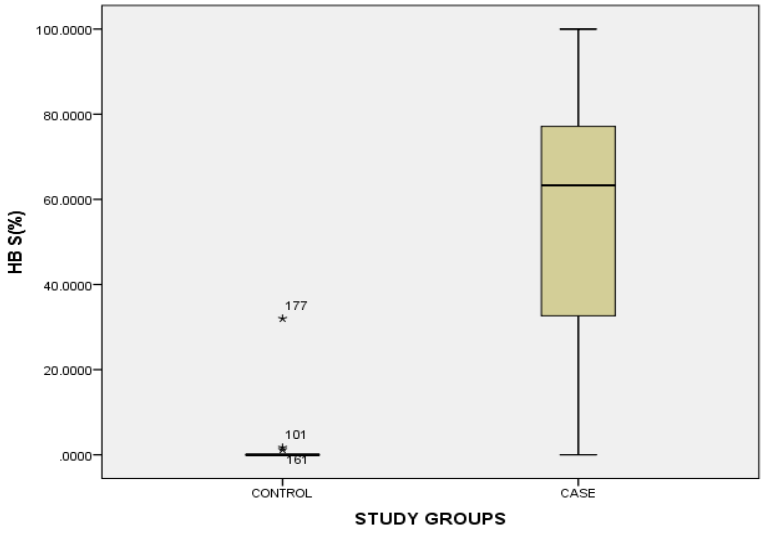
Hb S Q-Q Plots and Box Plots for the Study Groups











Hb F Q-Q Plots and Box Plots for the Study Groups