MODEL OF RISK PROBABILITY OF CHRONIC KIDNEY DISEASE IN NORTH CENTRAL REGION OF SRI LANKA

M.D.Nandana Gunaratne

(8207)

Degree of Masters of Science

Department of Mathematics

University of Moratuwa
Sri Lanka

February 2015

Declaration page of the candidate

I declare that this is my own work and this thesis does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature: Date:

Declaration of the Supervisor

The above candidate has carried out research for the Masters Dissertation under my supervision.

Name of the Supervisor : Mr. T. M. J. A. Cooray
Faculty of Engineering
Department of Mathematics
University of Moratuwa

Signature of the supervisor: Date

ABSTRACT

The chronic kidney disease of unknown etiology in farmers in North Central region has become an emerging health problem in Sri Lanka. Lack of scientifically proven experiments of epidemiological studies on the etiology is one of the major problems in solving this issue in Sri Lanka. Therefore, identifying the independent preventable risk factors mainly related to the occupation may helps in decreasing the number of patients suffering from CKD-U and slowing its progression. The objective of the study is to identify the epidemiological characteristics and potential risk factors related to agricultural activities in the development of CKD-U by using a case control study and modeling the risk probability of the disease occurrence.

The study was undertaken in Madawachchiya, Padaviya and Kebithigollawa area and two hundred and seventy four (274) patients with CKD-U (cases) and two hundred and seventy four (274) healthy individuals were selected as controls for the age and sex 1: 1 matched case control study. The relative risk of each factor was compared in terms of odds ratios (OR) and 95% confidence intervals (CI) by applying the conditional logistic regression model. The risk probability was calculated with the above model in cases and control separately and compared.

Involvement in agricultural activities, low protective measures against agrochemicals, and cultivating lands without labour exchange were identified as significant risk factors for the disease. Smoking and family history of CKD-U, drinking water source (shallow wells) and history of snake bite were identified as other life style related risk factors for the disease occurrence. The risk probability can be used as an index of the disease diagnosis and the receivers operating characteristic (ROC) curve plots the true positive fraction (TPF) against the false positive fraction (FPF) for different choice of cut off ranges. Therefore involvement in agricultural activities is significantly related for the disease. Low levels of protective measures in the application of agrochemicals indicate the importance to educate the farmers on protective measures of agrochemical applications.

ACKNOWLEDGEMENT

The contributions of the following are gratefully acknowledged:

At the time of putting demise to my thesis dissertation, on top of every one it is with heartfelt gratitude that I would like to extend my oozing out pleasures to my supervisor and teacher Mr. T.M.J.A. Cooray, Senior Lecturer, Department of Mathematics, University of Moratuwa

I am very gratitude to Mr. K.B. Jayasekara, Faculty of Medicine, University of Peradeniya who made his effort to collect data and technical assistance.

My sincere thanks goes to the Director and medical officers of the Department of Health Services North Central Provincial for providing us details of patients and medical advises.

Last but not least, I desire my deepest gratitude to my wife son and parents for their support and encouragement, to get through this agonizing period in the most positive way.

I also place on record, my sense of gratitude to one and all who directly or indirectly have lent their helping and hand in this venture.

CONTENT

Declaration of the candidate & supervisor i
Abstract ii
Acknowledgement iii
Content iv
List of figures v
List of tables vi
Abbreviations vii

1.0 INTRODUCTION 1
1.1 Chronic kidney disease 1
1.2 The chronic Kidney disease in Sri Lanka 1
1.2.1 Etiology of the chronic Kidney disease in Sri Lanka 2
1.3 Problem statement 5
1.4 Objective of the study 5
2.0 LITERATURE REVIEW 7
2.1 Chronic kidney disease 7
2.2 The Chronic kidney disease in Sri Lanka 7
2.3 The Case- Control Study 13
2.3.1 Matching 13
2.4 The logistic regression model 15
2.4.1 Conditional Logistic Regression 16
2.5 Estimation 17
2.6 ROC curves 17
3.0 METHODOLOGY 19
3.1 Study design 19

3.2 Sample size 19
3.3 Data and data collection 20
3.4 Data analysis 22
3.4.1 The Linear logistic model 23
3.4.1.1 Estimation..........24
3.4.2 Measures of association between diseased and exposure 25
3.4.3 Confidence Interval 26
3.4.4 Testing for the significance of coefficients..........27
3.4.4.1 The Likelihood Ratio Test 27
3.4.4.2 Wald Test 27
3.4.5 Computer software 28
4.0 RESULTS 29
4.1 Univariate analysis 29
4.2 Multivariate analysis 31
5.0 DISCUSSION AND CONCLUSION 37
REFERANCES .... Error! Bookmark not defined.

Annex 01..........45
Annex 02..........46
Annex 03..........49

LIST OF FIGURES

	Page
Figure 2.1: Prevalence of CKD-U in the North Central Area of Sri Lanka	8
Figure 4.1 : The "Model Fit Statistics" table of the SAS output	32
Figure 4.2: Box plot diagram of risk probabilities of the disease occurrence in case and control	35
Figure 4.3: The distribution of risk probability for diseased and non-diseased samples	35
Figure 4.4: The ROC curve for the diagnosis of the disease with risk probability	36

LIST OF TABLES

Page

Table 4.1: Odds ratios of agricultural and life style related risk factors 31

Table 4.2 : Analysis of Maximum Likelihood Estimates 33

Table 4.3: Odds ratios of agricultural and life style related risk factors in conditional logistic regression 34

Table 4.4: TPF and FPF at each cut off level 36

ABBREVIATIONS

AIC	Akaike Information Criterion
CI	Confidence Interval
CKD	Chronic Kidney Disease
CKD-U	Chronic Kidney Disease of Unknown Etiology
CRF	Chronic Renal Failure
CVD	Cardio Vascular Disease
FNF	False Negative Fraction
FNR	False Negative Rate
FPF	False Positive Fraction
FPR	False Positive Rate
GDP	Gross Domestic Production
ha	Hectare
ML	Maximum Likelihood
MRFIT	Multiple Risk Factor Intervention Trial
N	Sample size
NCR	North Central Region
OR	Odds Ratio
P	Probability
ROC	Receivers Operating Characteristic Curve
PREVEND	Prevention of Renal and Vascular End Stage Disease
TNF	True Negative Fraction
TNR	True Negative Rate
TPF	True Positive Fraction
TPR	True Positive Rate
SAS	Statistical Analysis System
SC	Schwarz Criterion

CHAPTER ONE

1.0 INTRODUCTION

1.1 Chronic kidney disease

Chronic kidney disease (CKD) has become a worldwide public health issue, with dangerous results of gradual loss of kidney function over time, cardiovascular disease (CVD), and premature death. Chronic kidney disease (CKD), is also known as chronic renal disease or chronic kidney failure. Kidneys filter wastes and excess fluids from the blood, which are then excreted in urine (World Health Organization & Ministry of Health and Nutrition, 2008). When chronic renal disease progresses it may lead to build up dangerous levels of fluid, electrolytes and wastes in the human body. In the early stages of chronic kidney disease, the patient may have few signs or symptoms. Because of the insidious nature of chronic kidney disease, it may not become apparent until the disease reached to an advanced stage and kidney function is significantly impaired. Early treatments for CKD can keep disease from becoming worse focusing on slowing it's progression of the renal damages by controlling the underlying cause. Chronic kidney disease can progress to end-stage kidney failure, which is fatal without expensive artificial filtering (dialysis) or a kidney transplant (National Kidney Foundation, 2013).

1.2 The chronic Kidney disease in Sri Lanka

Fifteen years ago health researchers and medical professionals in Sri Lanka have observed a high incidence of new form of chronic kidney disease of unknown etiology (CKD-U) or unknown origin in North Central Region (NCR) of Sri Lanka. The disease is prevalent in main rise cultivation areas under ancient irrigation systems and it has been on the increase dramatically over a period of last 15-20 years or last two decades. In recent years, a significant increase in some parts of the country especially in North Central, North Western, Uwa and Eastern Provinces and it is considered to be one of the most important emerging health issues. The total number of affected individuals is unknown, but a considerable percentage of the farming community is affected. It is estimated around 15,000 people are currently undergoing treatment for this disease. (Jayasekara et al.2013a). Males from poor socio-economic backgrounds who are involved in paddy cultivation are mainly affected segment of chronic kidney diseases and renal biopsies of these patients were reported as interstitial nephritis raising the concern of possible toxin exposure (Dissanayake, n.d).

1.2.1 Etiology of the chronic Kidney disease in Sri Lanka

Hypertension or chronic glomerulonephritis, diabetes mellitus, & infection (urinary tract) are three main causes for chronic kidney disease, however the disease in Sri Lanka was not related to any of the known causes and the etiology still remains a mystery. According to health professionals of Sri Lanka affected kidneys in the disease showed tubulo interstitial nephritis which is commonly related to toxic etiology (Jayasekara et al. 2013a).

Consumption of high fluoride groundwater, leaching of some heavy metals such as cadmium (Cd) and arsenic (As) from agrochemicals or chemical fertilizers in ground water, storing drinking water in aluminum (Al) utensils or containers considered as potential cause for chronic kidney disease and those studies clearly denote the CKD is affecting mainly the male farming community of the North Central Region of the island (Chandrajith et al. 2011b). Studies done so far in other countries identified that some different occupational related exposure factors and non occupational factors are interacting in a synergetic manner may cause for the disease. Some heavy metals has been identified as cause of CKD-U such as copper(Cu), lead (Pb), chromium(Cr), tin(Sn), mercury (Hg), and some other compounds such as silicon-containing compounds, welding fumes, grain dust, and oxygenated hydrocarbons etc may cause for the disease. However the etiology of the disease in North Central Region of Sri Lanka is still remaining mystery. (Wickramasinghe et al 2011)

Sri Lanka is an island in the North Indian Ocean and equatorial position gives its lowlands a tropical climate, with year round temperatures of 27-28^{0}C and a relatively constant day length. The island consists mostly of flat to rolling coastal plains, and rivers spread from the central hills outward to the coastal plains. The country has one hundred and three river basins.

Rainfall pattern is influenced by two monsoon winds from the Indian Ocean and Bay of Bengal which occur during two seasons of the year. From mid May to September the monsoon blows from southwest direction and brings in a greater amount of moisture than during December to February when the wind blows from northeast. The distinct inter-monsoonal periods receive conventional rains and at times cyclones. During the southwest monsoon, the position and dramatic relief of the southwestern side of the central high lands forces the moisture laden air upwards. The rapidly cooled air condenses, causing precipitation mostly on the windward slopes of the island's southeast. During the time the northwestern and southeastern parts of the island hardly receive any rain. On the other hand the northeast monsoon winds rise over the central highlands more gradually and the rain shadow effect is not nearly so distinct, allowing precipitation to fall on entire island. This has resulted in the division of the country into two major climatic regions; the wet zone which receives rain from both monsoons and the dry zone which receives rain from only northeast monsoon. The region in the wet zone of the country and some of the windward slope of the central highlands receives annual rainfall of 2200 mm or more than that covers only one third and the rest of the area a rainfall less than 1250 mm encompass two third of the island comprised of dry zone. The gradual change from the wet zone to dry zone allows an intermediate zone to exist. In addition two small areas at the extreme northwest and southeast of the country have a very dry climate and are known as arid zones (Fernando S, 2004).

Beside rainfall, temperature plays an important role in highland regions. For every 100m increase in elevation, the mean temperature fall 0.5^0C. On the plateau there is often grounds frost in lower lying areas between December and March. The evaporation rate in the dry zone is high about six mm per day and it is almost double compared to the wet zone of the island. Man made network of irrigation systems with reservoirs and canals which provide water for agricultural activities, animal husbandry, and human needs can be found in the dry zone from pre Christian era up to date. The chronic kidney disease is abundant in the dry zone in Sri Lanka; however it is prevalent in only few locations of the dry zone. North Central Region is the most CKD abundant area of the country. (Fernando S, 2004).

Agriculture sector is a very important in Sri Lankan economy and the primary form of agriculture in North Central Region is rice production. The area of the country is 65,610 square km and the land under agricultural practices comprises 2.3 million hectares. Around

80 percent of the agricultural area is located in the dry zone and intermediate zones of the country. The agricultural land area is approximately 2.3 million ha in extent in the region which is situated in the dry zone and intermediate zone of the country. Currently, agricultural lands covers around 2.26 million hectares in extend which is around 30 percent of the total land area of Sri Lanka. Around 40 percent of the agricultural lands are under small holders (Dharmaratne et al. 2006). According to the Central bank of Sri Lanka agriculture sector provides 11 percent contribution to the national GDP and provides employment around 30 percent of the national labour forces. Staple food in Sri Lanka is rice. Rice is cultivated by paddy farmers who are one of the poor population segments in the country. Rice is cultivated generally on a small scale in the rural areas in the North Central Region of the country and some of them are being operated under tenure agreements.

Around 0.8 million farm families are engaged in the paddy cultivation in Northern and Eastern plains. Theses farm families find their livelihood in the paddy cultivation mainly in small scale rice production and the average cultivating area is no more than 3 hectares per household. Traditionally men dominate the external economic scope in the household and women are generally engaged in internal household activities. Farmers depend on their traditional framing knowledge and methods that involves the participation of the whole community. Many farmers perform all farming activities such as land preparation, ploughing, tilling, leveling, planting, weeding, fertilizing, pest controlling and harvesting etc by manually. The paddy cultivated land area has increased approximately by 6 times from 1960 to 2000 and cultivated area is approximately 546,249 hectares in year 2000. Farmers will be promoted to use new agrochemicals by different private companies with their promotional activities. In addition to paddy cultivation, various other traditional agricultural crop productions are available for local consumption. They include coarse grains (Maize, Sorghum and millets), grain legumes or pulses (Green gram Cowpea and black gram), oil seed crops (sesame and groundnuts), tuber crops (Manioc/ Cassara, and yams) Condiments (onions and chilies), pulses, vegetables, and fruits etc. (Dharmaratne et al. 2006).

The epidemiological studies of CKD-U in Sri Lanka show that the disease is more prevalent in the male farmers over the age of 40 years age living in certain locations of the region. The presence of the disease in male farmers indicate the possibility that the exposure to the toxin is either more in male farmers of over the 40 years or there are other disease modifying factors operating in this group to enhance expression of the disease (Jayasekara et al. 2013a).

1.3 Problem statement

As mentioned earlier the CKD is being become a great threat for the country in economic aspects as well as physical quality of life the persons especially in the north central area of Sri Lanka. Because most of the victims of this threat are main income generators of the household as well contributors for the agricultural gross domestic production of the county. But the etiology of the disease is still unknown. Lack of scientifically proven experiments of epidemiological studies on the etiology is one of the major problems in solving this issue in Sri Lanka (Wickramasinghe et al 2011). Therefore, identifying the independent preventable risk factors mainly related to the occupation may helps in decreasing the number of patients suffering from CKD-U and slowing its progression.

1.4 Objective of the study

Most of studies so far carried out on the CKD-U indicate the possibility of an environmental toxin in the pathogenesis of the disease. Therefore, the aim of this study is to identify occupational and life style related risk factors of chronic kidney disease of unknown origin in the NCR of Sri Lanka that may help in the prevention of the disease.

1.5 Significance of the study

As discussed above the chronic kidney disease is an emerging health problem and it is often combined with poor health outcomes and high economical cost on patient, family, community and health system. In recent years a significant increase in CKD-U cases has been observed in some parts of the country especially in North Central Part of the country and the etiology of the disease remains a mystery. To resolve this public health issue research effort will generate conclusive evidence with existing information in implementing preventive actions of the disease.

1.6 Thesis Outline

This thesis consists of seven chapters including epidemiological background of the disease, and the objectives of the research effort in Chapter 1. Chapter 2 provides some related information and previous research done by other researchers in the same area. It includes a review of recent literature of chronic kidney disease. A valid and reasonable modeling method with relative risk of possible factors is presented in Chapter 3. Chapter 4 covers the mathematical modeling of relative risk using multiple linear logistic regressions developed in Chapter 3. Finally, Chapter 5 provides conclusions and summary of the research effort as well as the recommendations for future research interest.

CHAPTER TWO

2.0 LITERATURE REVIEW

2.1 Chronic kidney disease

Chronic kidney disease (CKD), also known as chronic renal disease, is a progressive loss in renal function over a period of months or years. Many patients may not have symptoms till the disease is worse and the symptoms of worsening kidney function are non-specific. It might include feeling generally more tired, trouble concentrating and having a reduced appetite etc. Usually , chronic kidney disease is detected as a result of screening of people known to be at risk of loss of renal function, such as those with having high blood pressure, diabetes, infections and those with a blood relative with chronic kidney disease (National Kidney Foundation, 2013). Chronic kidney disease may also be identified when it leads to one of its recognized complications, such as cardiovascular disease, anemia or pericarditis. Chronic diseases already report for seventy two percent of the total global burden of disease in persons over thirty (Cederborg 2011) and it will increase by seventeen percent over the next decade and much of this in developing countries.

2.2 The Chronic kidney disease in Sri Lanka

The CKD in Sri Lanka is still an unknown etiology (Wickramasinghe et al 2011). Sudden outbreaks of the CKD have been rampant in Sri Lanka. The high prevalence of the CKD had been chronically afflicted since 1990s in geographically discrete areas in the North Central Province of the country. But the situation was come to intense attention as late as year 2000. According to figure 2.1 three main high prevalent areas can be identified at Padaviya, Medawachchiya and Giranduru Kotte (Jayasekara et al. 2013b).

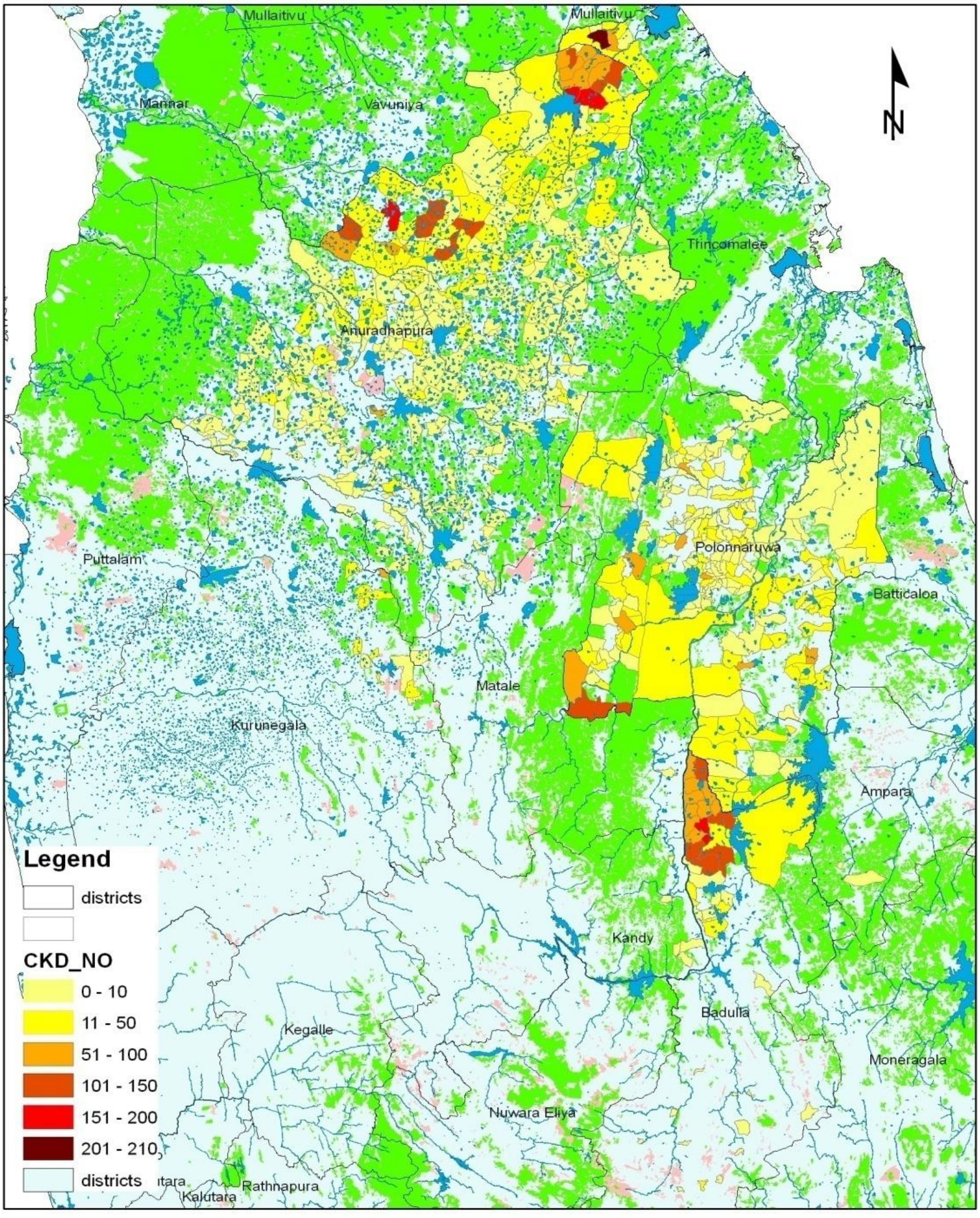

Figure 2.1- Prevalence of CKD in the North Central Area of Sri Lanka (Jayasekara et al. 2013b)

2.2.1 Etiology of CKD

The indistinct symptoms in early stages of the CKD-U, the diseases remains undiagnosed or undetected manner means many patients reach for medical treatments in its far gone stages when only more expensive kidney transplant or dialysis can maintain life. This situation remained rising due to the fact that it has a complex exposure scenario, and complications which confounds every possible causal factor, thus preventing taking case-specific risk alleviation or deduction measures. However in Sri Lanka the CKD-U is still an unknown etiology (Wickramasinghe et al. 2011).

2.2.2 Risk factors of CKD-U

During the last decade some medical researchers were undertaken to find out the prevalence and etiology of the CKD-U in the North Central Province of the country. Yet, some divergent and conflicting ideologies have been stated as causative and progmatic for this epic situation. Diabetes mellitus (Goonaratne et al. 2008), high blood pressure (Seneviratne et al. 2011), organophosphate pesticides (Pieris-John et al. 2006), exposure to mycotoxins like aflatoxins achratoxins, cyanobacterial toxins (Dissanayake n.d), chewing of betel nuts and snake venom (Herath et al. 2011) are some of the ideologies. Some general risk factors have also been stated views that North Central Province farming communities are more vulnerable than non-farmers, as they are exposed to possible disease modifiers, such as exposure to chronic dehydration and environmental contaminants such as fluoride in consuming water, that exceeding the upper level (0.5 mg/ L) declared by the World Health Organization (Warnakulasuriya et al 1992). It has been reported that excessive dependence on agro-wells or shallow dug wells for drinking water and application of considerable amount of agrochemicals such as weedicides, pesticides and fertilizers will lead to pollution of the quality of water, hence farmers' health will be affected (Environment Action Plan 2002).

Gender is a factor which is strongly associated with the disease and there is a male preponderance in chronic kidney disease afflicted patients. The age group more than 40 years, was strongly associated with the diseases (Athuraliya et al 2011). There is scientific consensus that this fatal disease in the country is not related to known causes commonly identified with kidney disease, such as, diabetes, hypertension, obesity, or other factors such as snake bite.

Due to differential causality factors, and difficulty in pinpointing any known risk factors the exact cause of this disease is being still determined. Therefore the World Health Organization and the Sri Lankan government continue to designate it as "Chronic Kidney Disease with unknown etiology (CKD-U) in scientific arena (Chandrajith et al. 2011a; 2011b). In this context, CKD-U is more likely to be multi-factorial disease rather than having a single cause.

The etiology of the disease has been subjected to many researches and ideologies over the last few years in which most researchers has focused on cases geo-environmental factors, which have been promulgated on CKD-U. For example CKD-U prevalent areas in the country are overlapping with the high ground water fluoride zone of Sri Lanka as suggested by Dissanayake (1996), that indicating at least to some extent, the fluoride content of drinking water may contribute to the CKD-U in Sri Lanka.

However, due to non-prevalence of the disease in some other geographical areas like Ampara, Hambantota and Monaragala with extremely high drinking water fluoride concentration within the dry zone of the country, some alternative mechanisms have been presented for increased fluoride activity in ground water. For example, more soluble and / or bio-available metal / ion complexes are formed with aluminum (Al) (Chandrajith et al. 2009) and Calcium (Ca^{2+}) (Chandrajith et al. 2011a), whereby enhancing alternative fluoride species such as CaF_4 , CaF^-, AlF_3 and AlF_4 in environmental media. More specifically, differences of sodium and calcium (Na and Ca) activities in consuming water between disease prevalent and non-prevalent areas are characterized by higher sodium /calcium ratio, i.e, subtle variation in sodium /calcium ratio (Na/ Ca) in the ground water consumed as drinking water, rich in fluoride can exacerbate CKD-U prevalence in endemic areas (Chandrajith et al 2011b). Chandrajith et al (2010) expressed views that possible sources for heavy metals like Uranium (U) and few other elemental groups with radionuclide properties like Potassium ^{40}K, Thorium ^{232} Th, and Uranium ^{238} U, which can come as impurities with agro chemicals (i.e. phosphate fertilizers) , are risk factors for the disease occurrence.

A research suggesting potential impact of heavy metals on groundwater as a result of fertilizer use, suggests significant potential groundwater pollution from cadmium (Cd), Selenium (Se), Molybdenum (Mo) and Uranium (U) caused as a result of long run application of phosphate fertilizers like triple super phosphate (TSP) (Bandara et al. 2010; 2011). The possibility contaminating the groundwater system with for Cd through soil compartment is indicated. Poor de-stilting of reservoirs in North Central Province could lead to progressive leaching of high Cd levels in water which in turn feed the paddy farming (Bandara et al. 2011). (Bandara et al 2011) further suggested that application of some popular pesticides like bisyribac-sodiumin in the region for paddy cultivation is communicated with high amounts of Cd levels. However, the Cd concentration of fertilizers and pesticides used in this primary agricultural setting is important and the source of Cd in soil and reservoirs are still not known.

A study undertaken of five reservoirs by Bandara et al. (2008), the total cadmium that is carries to reservoirs in NCP from both the cadmium input from phosphate- fertilized crop fields (rice and vegetables) in the NCP and the diverted River Mahaweli, and that eventually settles in the total concentration of Cd in sediment samples from collected the bottom of the five reservoirs ranged from 1.78 – 2.45 mg/kg. These sediments then release $Cd2^{+}$ into reservoir water, leading to high level of Cd in the irrigation and drinking water. The dissolved Cd in reservoir waters ranged from 0.03 to 0.06 mg/ L, which is a 10-20 fold increase over the maximum contamination level set by the WHO is 0.003mg/ L for drinking water and 0.005 mg/ L recommended by the United State Environment Protection Agency and the concentrations in sediments did not exceed the European Union Maximum Permissible level of 3.0 mg Cd per kg for agricultural soils. The consumption of Cd contaminated lotus rhizomes which have extraordinarily high levels (253.82 µg / kg) and reservoir fish species such as Thilapia (0.5-90.7 µg/ kg), rice grains collected from the farms of chronic renal failure patients in Medawachchiya, (ranged from 0.001 to 0.093 µg/kg in dry weight) and snake head (1.2-114.4 µg/ kg) have also suggested as possible risk factors (Bandara et al 2008). However, Chandrajith et al. (2010) eliminated arguments of Bandara et al. (2008) for possible etiology of Cd intake on NCP in the island could be due to progressive tubular damage caused by renal tubular toxins other than Cd.

Meanwhile, Prof. Nalin de Silva stated that the arsenate compounds responsible for the KDU in NCP are formed only when arsenic is mixed with hard water found in the

Rajarata areas (De Silva 2011). Jayasumana et al. (2011) theorized that arsenic (As) and mercury (Hg) in pesticides formulations sold in Sri Lanka is directly responsible for CKD-U incident in affected regions and identified high amounts of arsenic in rice collected from CKDu endemic regions, such as Giradurukotte and Nikawewa. In addition, rice has been said to be contaminated with arsenic, thus increasing the body burden to intolerable levels of arsenic element and arsenate compounds in the NCP communities. The above researchers hypothesized that arsenic in environmental media (mainly in water) forms strong bonds with calcium ions (Ca) presence in the hard water to form calcium arsenate which is difficult to elicit in 'hard water' and this is one of the major course of kidney dysfunction (i.e tubul interestitial nephritis which is toxic etiology).

The "calcium arsenate theory" was partly refuted by Illeperuma (2011). Because formulation of insoluble complexes by calcium arsenate that decreases human exposure and other by not in agreement with CKD-U prevalence in other parts of Sri Lanka with high application of agrochemicals and drinkable hard water. According to Chandrajith et al. (2005) the arsenic element in the soil where paddy is being cultivated in Sri Lanka is relatively smaller and the mean concentration of arsenic is around 1.1 µg per kg. However studies conducted on sediments of tanks could reflect that chemical input and anthropogenic environment pollutants (Chandrajith et al 2008) will not be related to create an alarming risk due to soil and sediment characteristics which has the ability of ameliorating deterioration of heavy metals such as arsenic (Chandrajith et al 2008). The results obtained by Chandrajith et at (2008) are vital for future pollutant management of tank ecosystem in Sri Lanka since detailed information on pollutants and element distribution within the sediment of tank is lacking. On the other hand some habits related to CKD such as smoking habits have been studied in so many foreign researchers.

In conclusion the etiology of CKDu in NCP of Sri Lanka is related to potential impact of multi-factorial risk factors with occupation related risk factors, some habitual risk factors and a possible genetic predisposition in vulnerable populations (Jayasekara et al.2013a). It is very important to understand about the environment which is made up of a complex interrelated web of disciplines (Jayakodiarachchige 2012)

In effective management of CKD-U, first of all we should understand the environment which is made up of a complex inter-related web of disciplines. Constraints and limitations in human resources, science, technology, socio-economic matters and financial aspects have to be intensively managed to achieve a success against the high prevalence of CKD-U in Sri Lanka. Epidemiological and other scientific researches have not ascertained a complete scientific answer on CKD-U in Sri Lanka which were conducted to date. Therefore it is important to encourage precisely designed, adequately powered experiments on this national health issue and studies which are linking the past and the present would enhance speculations on geo-environmental risk factors to provide a complete answer.

2.3 The Case- Control Study

Janet Lane-Claypon's study of breast cancer in 1926 was the first modern case control study. But case control designs were applied intermittently many in epidemiological and social sciences until 1950 by many researchers (Paneth et al.2002). In subsequent few decades early studies synthesized the essential elements in the comparison of cases and controls which developed conceptual designs shifting within epidemiological studies. This lead to great development in case control design.

The case control design is one of the more effective solutions to analyze epidemiological studies. In many such researches the aim is to detect possible risk factors of a disease. Individuals with a particular disease are compared to a group in whom the disease is absent with respect to exposures thought to be relevant to the development of the disease. Here individual with a particular disease is known as the case and individual who do not have the disease is known as the control. The expected effect of exposure on risk of disease often will be expressed in terms of the relative risk (disease rate ratio) RR parameter in epidemiological experiments. (Collett, 1991)

2.3.1 Matching

According to Collett (2001), a confounding variable is an extraneous factor, which is associated with both the exposure and the disease under study. Several approaches to control for confounding can be utilized in both the design and analysis of analytic studies. One strategy to control the influence of a confounding variable is to use a method of subject selection known as matching. In a matched case control study the referent group is chosen in accord with certain constraints, in order to match characteristics of individual cases thought to be confounders of the association under study

The matching variable can be treated in the analysis as either as categorical or continuous. The matching variables are inherently categorical in categorical matching. Otherwise data should be converted into categorical data. In practice age and gender are usually very strong confounders and exact age and gender are a matching variable in almost all matched case control studies.

Predetermined proportion of cases will be randomly selected in frequency matching and controls will be selected as the same proportion control per category. So it is known as individual matching which is selected specifically for each control instead of matching set of matched controls. R-1 matching refers to the situation where each case is individually matched with R controls. R can be vary from case to case that can be frequently observed when the matching ratio is intended to be R-1. Selection of more than many controls per case instead of selecting one control will increase the power of the study. However the marginal increase of the statistical power is small if it exceeds 4 controls per case.

Increasing the efficiency is an advantage of matching cases and controls when the matching variable is a very strong disease determinant. But the efficiency will reduce when the matching variable is strongly related to the exposure. This is referred as to over matching situations. There are several disadvantages in matching strategy. Sometimes it can be difficult, expensive and time consuming to find appropriate matches in practicing matching strategy. More over the effect of the confounding factor itself on the risk of the outcome cannot be explored.

Each case and control performs one stratum when analyzing the data in a matching case control study. Incase if either case or control has missing data the entire stratum will be lost for analysis in an individually matched case control study. Matching that is accompanied by an unmatched analysis can reduce the validity of case control comparisons. If the matching was ignored in the analysis it will be influenced in an odds ratio biased toward conservatism.

A cross-sectional study has conducted Jayathilaka et al (2013) in their study on "Chronic kidney disease of uncertain etiology: prevalence and causative factors in a developing country" the study has been conducted, to determine the prevalence of and risk factors for CKD-U. Arsenic, cadmium, lead, selenium, pesticides and other elements were analyzed in biological samples from individuals with CKD-U and compared with age and sex matched controls in the endemic and non-endemic areas. Food, water, soil and agrochemicals from both areas were analyzed for heavy metals.

These results indicate chronic exposure of people in the endemic area to low levels of cadmium through the food chain and also to pesticides. Significantly higher urinary excretion of cadmium in individuals with CKD-U, and the dose - effect relationship between urine cadmium concentration and CKD-U stages suggest that cadmium exposure is a risk factor for the pathogenesis of CKD-U. Deficiency of selenium and genetic susceptibility seen in individuals with CKD-U suggest that they may be predisposing factors for the development of CKD-U.

2.4 The logistic regression model

In 1944, Joseph Berkson introduced the logit model and he coined the term. Chester Ittner Bliss developed an iterative approach to finding maximum likelihood estimates in the probit method in binary data analysis. In 1949, log-odds; the log-odds of an event is the logit of the probability of the event was invented by G. A. Barnard (Nerds, 2014).

In a case control study dependent variable in a logistic regression is binary (dichotomous) i.e. it is the case control status (Y=1 for cases and Y =0 for controls). But the dependent variable(s) can be polychotomous. The logistic regression model can be used in binarary data analysis to find the best model to describe the relationship between the dependent and possible independent variables (predictor or explanatory). Theses independent variables are often called covariates.

Wanigasuriya et al (2007), study was conducted to determine the etiology of chronic renal failure (CRF) in the North Central Province of Sri Lanka. Patients (183 cases) with CRF of unknown etiology were compared with controls (200 controls) who had no evidence of chronic renal dysfunction. Exposure to possible risk factors were determined by an interviewer-administered questionnaire and analyzed with logistic regression. Being a farmer ($P < 0.001$), using pesticides ($P < 0.001$), drinking well water ($P < 0.001$), a family history of renal dysfunction ($P = 0.001$), use of ayurvedic treatment ($P < 0.001$) and a history of snake bite ($P < 0.001$) were risk factors for CRF of unknown etiology. Using logistic regression analysis, a family history of chronic renal disease, taking ayurvedic treatment and history of snake bite were found to be significant predictors for CRF of unknown etiology.

2.4.1 Conditional Logistic Regression

Cyrus et al. (2012) shows that exact inference for the logistic regression model is based on generating the permutation distribution of the sufficient statistics for the regression parameters of interest conditional on the sufficient statistics for the remaining (nuisance) parameters. Despite the availability of fast numerical algorithms for the exact computations, there are numerous instances where a data set is too large to be analyzed by the exact methods, yet too sparse or unbalanced for the maximum likelihood approach to be reliable. Cyrus et al. (2012) suggest Monte Carlo alternative to the exact conditional approach which can bridge the gap between the exact and asymptotic methods of inference. The problem is technically hard because conventional Monte Carlo methods lead to massive rejection of samples that do not satisfy the linear integer constraints of the conditional distribution. They propose a network sampling approach to the Monte Carlo problem that eliminates rejection entirely. Its advantages over alternative saddle point and Markov Chain Monte Carlo approaches are also discussed in “Efficient Monte Carlo Methods for Conditional Logistic Regression”.

The conditional method of maximum likelihood estimation for fitting logistic regression is the most appropriate method when analyzing individually matched data. Cases with disease are matched by cases without disease on variables such as age, gender, and so forth. There may be 1 to 1 or 1to many matched control for each case. The outcome has two categories (case and control) when there is only one matched control subject, whereas the outcome lead to multiple categories (disease, control1, control2. etc.) when multiple matched control subjects are in the case control design. The model is, as usual, based on the predictors that are included, but there is no constant (or intercept) (Collett ,2001).

2.5 Estimation

There are two alternative maximum likelihood (ML) approaches that are used to estimate the parameters (β) in the logistic model: the unconditional and the conditional maximum likelihood (Collett ,2001)

The likelihoods must be then solved iteratively. Ever though the logistic model was initially designed to analyze prospective studies it can also be applied to analyze case control studies also. Despite the definition of the constant term of the linear predictor, the likelihood function for the retrospective case control study is the same as for the prospective study. The likelihood function depends on the ratio of the probabilities for a case and control which is to be included in the case control study.

Incase if the number of parameters to be estimated in the logistic model is small compared to the number of subjects, the unconditional maximum likelihood method is more appropriate. Whereas when the number of parameters is large relative to the number of subjects, the conditional maximum Likelihood method is more precise in statistical analysis. However, the conditional likelihood approach will be the situation when the research design involves matching. When using a logistic model to analyze k matched pairs, the model will also contain covariates and interaction terms (Collett ,2001).

2.6 ROC curves

The ROC curve was developpped during World War II for the analysis of radar signals before it was employed in signal detection theory. Following the attack on Pearl Harbor in 1941, the United States army began new research to increase the prediction of correctly detected Japanese aircraft from their radar signals (Nerds, 2014)

In the 1950s, ROC curves were employed in psychophysics to assess human (and occasionally non-human animal) detection of weak signals. In medicine, ROC analysis has been extensively used in the evaluation of diagnostic tests. ROC curves are also used extensively in epidemiology and medical research and are frequently mentioned in conjunction with evidence-based medicine (Nerds, 2014). In radiology, ROC analysis is a common technique to evaluate new radiology techniques. In the social sciences, ROC analysis is often called the ROC accuracy ratio, a common technique for judging the accuracy of default probability models.

Jayathilaka et al. (2013) has been used receiver-operating characteristic (ROC) curve was to calculate the area under the ROC curve, to determine the cut-off values for cadmium and selenium with the best sensitivity and specificity ROC curves generated with urine arsenic, cadmium and lead and serum selenium concentrations.

The first application of ROC in machine learning was by Spackman. ROC curves proved useful for the evaluation of machine learning techniques. Spackman used to demonstrate the value of ROC curves by comparing and evaluating different classification algorithms. In signal detection theory, a receiver operating characteristic (ROC), is a graphical plot and it is important in graphical illustrations of the performance of a binary classifier system as its discrimination cut off point is varied. The ROC can be created by plotting true positive fraction out of the total actual positives (TPR = true positive rate) against false positives fraction out of the total actual negatives (FPR = false positive rate), at different cut off levels. TPR and FPR are important measures which deal with the sensitivity and specificity. Because TPR is known as sensitivity and FPR is “1- specificity” or known as true negative rate. In general, if both of the probability distributions for detection and false alarm are known, the ROC curve can be generated by plotting the cumulative distribution function (area under the probability distribution from –infinity to +infinity) of the detection probability in the y-axis versus the cumulative distribution function of the false alarm probability in x-axis (Nerds, 2014). The ROC is also known as a relative operating characteristic curve, because it is can be used to compare two operating characteristics (TPR and FPR) at different criterions.

CHAPTER THREE

3.0 METHODOLOGY

3.1 Study design

This study is a case control design. Comparisons were made between individuals who have the CKD-U, known as cases and individuals who do not have the disease known as controls. 1: 1 age and gender matched case control study was accomplished on the basis of potential confounding variables.

3.2 Sample size

Two hundred and seventy four (274) patients with CKD-U (cases) and Two hundred and seventy four (274) healthy individuals were selected as controls for the age and sex 1: 1 matched case control study. Both cases and controls were selected from Madawachchiya (272 persons), Padaviya (148 persons) and Kebithigollawa (128 persons) which are CKD-U high prevalence areas in NCR.

Minimum sample size *(N)* for a 1: M matched case control study could be determined as follows (Sinha & Mukherjee, 2005).

$$N_S^* = \frac{\left\{\dfrac{Z_{\frac{\alpha}{2}}(1+\varphi) + 2 \times Z_\beta \varphi^{1/2}}{\varphi - 1}\right\}^2}{\dfrac{(\varphi+1)P_{01}(1-P_{01})}{1+(\varphi-1)P_{01}}}$$

$$N = \frac{(M+1)N_S^*}{2M}$$

Where, φ is the odds ratio P_{01} is the exposure prevalence in the control population (as a parentage).

This formula for sample size (N) calculation in pair-matched studies is based on the normal approximation to testing a single binomial proportion. $Z_{\frac{\alpha}{2}}$ and Z_β for two tailed type I error and power respectively. Here we will be formed a 1: 1 matched case control design and therefore *M*=1.

Where, P_{01} denoted the percentage of involvement of agricultural activities in more than 2 hectares of land extent in the control = 50% (table 4.3). Appropriate odds ratio is φ = 3.42 in this risk factor (table 4.3). Minimum sample size was calculated assuming a two tailed 5% type I error rate and 20% type II error (80% power).

By substituting in the above equation,

$$N_S^* = \frac{\left\{\frac{1.96(1+3.42)+2\times 0.84\times 3.42^{1/2}}{3.42-1}\right\}^2}{\frac{(3.42+1)0.5(1-0.5)}{1+(3.42-1)0.5}}$$

$$N_S^* = 47.31$$

$$N = \frac{(1+1)47.31}{2\times 1}$$

Therefore required minimum sample size would be 47 individuals with the disease.

3.3 Data and data collection

Data were collected through questionnaires and oral informed consent which was obtained from all patients and controls by face to face interviews. All cases were disease identified patients from 2009 to 2012, who are registered in renal clinics in Madawachchiya, Kebithigollawa and Padaviya areas in NCR. Healthy individuals for the study were selected as control from the neighboring household unit of each case.

Data on involvement in agricultural activities, cultivated area, labour hiring for field agricultural activities, application of agrochemicals and protective measures against agrochemicals, were collected as occupation related risk factors for the disease. In addition to the possible occupational risk factors, data on family history of CKD-U, habitual alcohol consumption, smoking, water source for daily use for the last five (5) years, malaria, history of snake bites, regular use of indigenous medicine and betel chewing related data were also collected.

The occupational risk factors such as intense involvement in agricultural practices with cultivated land area in extent or operated land area, application of protective wears against exposure to agrochemicals, labour exchange in agricultural activities, and other life style related information were collected from both healthy individuals and patients The occupational risk factors and other life style related risk factors related data was collected for the previous five years from first date of diagnosing the disease. Each risk factor was classified as dichotomous variables by classifying them into two categories in modeling as exposed and unexposed groups.

Individuals who have a regular involvement in agricultural activities with manual activities for at least eight hours per day and four days per week for last five years was classified as exposed for agricultural activities.

The cultivating land for farming was considered high when the farming area is either two or greater than two hectares and low when it is less than two hectares

According to farming practices farmers were divided in to two groups considering physical exertion. The farmers who involve in agricultural activities in the field by themselves in their cultivated fields without hiring labour for more than eight hours per day for at least four days per week was considered as low level of labour hiring and the other segment who use hired labour for field agricultural practices without an intense physical exertion by themselves was considered as high level of labour hiring.

It is very difficult to classify the status of protective measures against agrochemical in its application in Sri Lanka as information is scarce. However according to considering practical aspects in agricultural practice farmers who are wearing long sleeved shirts, long trousers, covered shoes, gloves (water proof) and masks were considered as high level of protection against agrochemicals in its application and farmers or workers who do not use above safety measures were considered as low protective status against application of agro- chemicals.

Family history has also been considered as a risk factor for developing chronic kidney disease and if the chronic kidney disease was reported in a first, second or third degree relative the disease is said to be presence in the family history. The history of the disease occurrence in the generation was limited to three degree of relatives as that information might be precise instead of considering more generations.

A person who was infected malaria at least one time was considered as malaria experienced individual.

Life style related smoking, alcohol consumption and betel chewing were also considered as habitual risk factors. According to available information of history of regular smoking at least last five years individuals were stratified as smokers. A dichotomous exposure factor was corresponds to two levels of alcohol consumption. Regular alcohol users over five days per week for last five years were classified as high level of alcohol consumption. Betel chewing also in regularly at least last five years was classified as high level in its dichotomous exposure factor.

Most of the farmers use shallow dug wells in the area as the drinking water source and therefore water source was correspond to the shallow dug wells and other. Other water sources are tube wells, natural springs and reservoirs etc.

Regular use of aurvedic or indigenous medicines for not less than 3 months was considered as the high level of the exposed factor.

3.4 Data analysis

Logistic regression was used to analyze the binary outcome of the disease occurrence with the effects of possible risk factors as independent (dichotomized) variables. The ROC curve was applied to determine the threshold level of risk probability for a diagnosis of cases and control.

3.4.1 The Linear logistic model

The linear logistic regression model was used in both univariate and multivariate analyses. The univariate logistic regression model has one dichotomous outcome variable *Y and* X independent variable and, the logit of *Y* from *X*. So it has the following equation.

$$logit(P_j) = \beta_0 + \beta_1 X_j$$

Where, P_j is the probability that the disease occurs in an individual in the j^{th} exposure group j = 1, 2….*n*. *X* is an indicator variable that takes *X* = 0 for individuals in the unexposed group and *X*=1 for those in the exposed group. Here, β_0 and β_1 are regression parameters.

Fitting a series of univariate models rarely provide adequate analysis of the data in a study since the independent variables are usually associated with one another and many have different distributions within levels of the outcome variable. Thus multivariate analysis was done applying multivariate linear logistic regression model.

As the study design was matched, the probability that i^{th} person in the j^{th} matched set will be denoted as $P_j(X_{ij})$, where X_{ij} is the vector of *k* exploratory variables for the case is denoted X_{0j}, while the vectors X_{ij} for *i*=1,2,…*M*, will denote *M* controls in the j^{th} matched set. The disease probability$P_j(X_{ij})$ can be modeled using a linear logistic model with a different constant term $X_{j.}$ as follows.

$$Logit\{p_j(X_{ij})\} = \alpha_j + \beta_1 X_{1ij} + \beta_2 X_{2ij} + \ \dots \dots . + \beta_k X_{kij}$$

Where, X_{ij} is the value of the h^{th} independent variable or factor $X_{h,}$ $h = 1,2, \dots ., k$, for the i^{th} individual, *i*= 0,1,…..,*M* in j^{th} matched set. The parameter α_j in this model is a constant which summarises the effect of the matching variables on the probability of disease occurrence (Collett, 2001)

3.4.1.1 Estimation

Estimating the unknown parameters in the model will be involved by fitting a model. The method of maximum likelihood and the method of least square are the most important techniques in fitting models. The method of least squares is usually adopted in fitting linear regression models. But the most widely used general method of estimation is the method of maximum likelihood method which is generate in SAS output.

When the binomial data of the form y_i success out of n_i trials, i = 1,2,…….,n, the corresponding success probability is to be P_i.

The k+1 unknown parameters $\beta_0, \beta_1, \beta_2, \ldots \ldots \beta_k$ should be first estimated in order to fit a linear logistic regression model to a given data set. The likelihood function is as follows.

$$L(\beta) = \prod_{i=1}^{n} \binom{n_i}{y_i} P_i^{y} (1 - P_i)^{n_i - y_i}$$

The logarithm of the likelihood function is as given below.

$$\log L(\beta) = \sum_i \left\{ log \binom{n_i}{y_i} + y_i log P_i + (n_i - y_i) \log(1 - P_i) \right\}$$

$$= \sum_i \left\{ log \binom{n_i}{y_i} + y_i log_i \left(\frac{P_i}{1 - P_i}\right) + n_i \log(1 - P_i) \right\}$$

$$= \sum_i \left\{ log \binom{n_i}{y_i} + y_i \eta_i + n_i \log(1 + e^{n_i}) \right\}$$

Where, $\eta_i = \sum_{j=0}^{k} \beta_j x_{ji}$

Following evaluation shows the derivatives of the likelihood function with respect to the k+1 unknown parameters

$$\frac{\partial \log L(\beta)}{\partial \beta_j} = \sum y_i x_{ji} - \sum \eta_i x_{ji} e^{\eta_i} (1 + e^{\eta_i})^{-1} \qquad j = 0,1, \ldots \ldots k$$

Evaluating theses derivatives at $\hat{\beta}$ and equating to zero gives a set of k +1 non-linear equations in the unknown parameters $\hat{\beta}_j$ that can only be solved numerically. If the epidemiological experimentation is a matched case control study a conditional likelihood has to be applied to give up large number of nuisance parameters. The likelihood function (conditional on disease status) of a general form is as follows , L_{con} is:

$$L_{con} = \prod_{j=1}^{n} \left[1 + \sum_{i=1}^{M} exp\left\{ \sum_{h=1}^{k} \beta_h (X_{hij} - X_{h0j}) \right\} \right]^{-1}$$

Where, X_{hij} is the value for the i^{th} individual, i= 0,1,…..,M in j^{th} matched set in the value of the h^{th} explanatory variable, $h = 1,2, \ldots, k$,

As the design was 1: 1 matched case control study the conditional likelihood function gives as follows.

$$L_{con} = \prod_{j=1}^{n} \left[1 + exp\left\{ \sum_{h=1}^{k} \beta_h (X_{hij} - X_{h0j}) \right\} \right]^{-1}$$

As mentioned above since the study design was matched, the model parameters were estimated using conditional maximum likelihood in multivariate analysis. Estimates of univariate and multivariate analysis were used to calculate odds ratios and 95% confidence intervals. The analysis was done by using SAS statistical software. The probability of the disease occurrence was found at defined levels of significant individual variables and it was used to determine as an index of the disease occurrence.

3.4.2 Measures of association between diseased and exposure

Odds ratio is a ratio of "odds which was calculated as a measure of relative risk of each factor. The main purpose of estimating risk was to be able to make comparisons between the risks of the disease at different levels of exposure factors and thereby asses the associations between the disease and exposure factors. Here each possible risk factor has two distinct values known as "exposed" and "unexposed". Denote the risk of the diseased occurring during a given time period by P_e for the exposed person and P_u for the unexposed persons. The risk of disease occurring in an exposed person relative to that for an unexposed person is then $= \frac{P_e}{P_u}$.

This is a measure of the extent to which an individual who is exposed to a possible risk factor is more or less likely to develop the disease than someone who is unexposed.

Odds ratio was used as a measure of association between the occurrence of the disease and exposure factor. Odds of a person having the disease in the exposed group are then $P_e/(1-P_e)$ and similarly odds of disease for someone in the unexposed group are $P_u/(1-P_u)$. The odds ratio is the ratio of the odds of disease for an exposed relative to an unexposed person given by

$$\varphi = \frac{P_e/(1-P_e)}{P_u/(1-P_u)}$$

3.4.3 Confidence Interval

Confidence interval estimator's for the slope and intercept were determined based on their respective Wald test. The endpoint of a 100(1-α)% confidence interval for the slope coefficient (β_i) is

$$\hat{\beta}_i \pm z_{1-\frac{\alpha}{2}} SE(\hat{\beta}_i).$$

And for the intercept (β_0)

$$\hat{\beta}_0 \pm z_{1-\frac{\alpha}{2}} SE(\hat{\beta}_0)$$

Where $z_{1-\frac{\alpha}{2}}$ is the upper 100(1-α)% point from the standard normal distribution and $\widehat{SE}(.)$ denotes a model based estimator of the standard error of the respective parameter estimator. In the SAS output of the fitted model will provides the endpoints of the interval estimates. (α= significant level)

3.4.4 Testing for the significance of coefficients

After estimating coefficients assessment of significance of variables in the model was done.

3.4.4.1 The Likelihood Ratio Test

The likelihood ratio test is a statistical test that can be used to make a decision about the significance of an explanatory variable in the logistic model the maximized likelihood with and without the explanatory variable in the linear logistic regression model and the statistics can be written as

$$G = -2ln\left[\frac{(likelihood\ without\ the\ variable)}{(Likelihood\ with\ the\ variable)}\right]$$

The probability distribution of the likelihood test statistics is approximately a χ^2 distribution with one degree of freedom under the $H_0 : \beta_i = 0$. The likelihood ratio test evaluates the overall significance of the coefficients in a multiple logistic model in the same manner in both univariate and multivariate models. Under the null hypothesis that the *r* parameters all are equal to zero, the distribution of G will be χ^2 with *r* degrees of freedom. The likelihood ratio test can be used to compare a reduced model with a full model.

3.4.4.2 Wald Test

The Wald's statistics is a parametric statistical test in the logistic regression model which is given below.

$$W_j = \frac{\hat{\beta}_j^2}{Var(\hat{\beta}_j)}$$

The statistics can be used can be applied to test the true value of the parameter based on the sample estimate i.e, null Hypothesis (H_0): $\beta_i = 0$. where, *j=1,2,....r.* Under the null hypothesis these statistics is compared against $\chi 2$ distribution with one degree of freedom.

3.4.5 Computer software

SAS (Statistical Analysis system) (version 8) was used as the computer software system for linear logistic modeling. The procedure **proc logistic** was used to fit linear logistic models to binary data. That procedure allows a range of model checking diagnosis to be obtained directly and incorporates a number of sequential procedures for model selection. Minitab was used for graphic facilities in plotting ROC curves and box plot diagrams.

CHAPTER FOUR

4.0 RESULTS

The output of both univariate and multivariate test results will be considered. Output of the univariate analysis shows preliminary awareness about factors influenced for the disease but multiple logistic regression analysis was the more precise analysis since there is a single dichotomous outcome and more than one independent variable. One of the reasons for using linear logistic model in this analysis of data is that the coefficients of exploratory variables in the model can be interpreted as logarithms of odds ratios. This means that estimates of the relative risk of the disease and corresponding standard errors can be easily obtained from fitted models. The Chi-Square statistic in the proc logistic procedure of SAS output which test whether the parameter estimates equals zero. From the output, the *P* value of the test could be used reject or accept the null hypothesis (H_0) and conclude that each factor is significantly related to disease occurrence.

4.1 Univariate analysis

Univariate logistic regression analysis using SAS was applied to investigate the relationship between predictors and each variable. Estimates of odds ratios were obtained from the fitted linear logistic regression and epidemiological information are based on the magnitude of estimated odds ratios. Table 4.1 shows number of observations in cases and controls by levels of each explanatory variable with percentages, estimated odds ratios, confidence intervals and *P* values. *P* values of Likelihood Ratio Chi-Square statistics of each variable could be used to determine the significance of fitted linear logistic models.

Univariate linear logistic model analysis in Table 4.1 shows that 88% of cases and 78% of control group are involving in agricultural activities in the study area and status of protective measures against agrochemicals, cultivated land area and hiring labour for agricultural practices could be considered as occupation related risk factors. According to the table 4.1 the involvement in agricultural practices (OR = 2.12, 95% CI; 1.33–3.38), low preventive measures against agro-chemicals (OR = 8.4, 95% CI; 4.49–15.59), agricultural activities without labour exchange (OR=2.77, 95% CI; 1.85–4.14), high cultivated land area (more than 2 ha) (OR = 2.32, 95% CI; 1.52–3.53), were identified as significant occupation related risk factors for the disease occurrence (P < 0.01).

Concerning above mentioned odds ratio in agricultural activities related risk factors the magnitude of relative risk is highest in low protective level against agrochemicals. The table 4.1 shows that 92% of the cases and 58% of the controls have low level of protective measures. Pesticides, weedicides and some fertilizers (Urea) were widely used agrochemicals in the field and some of them are Paraquat, Bispyribac sodium, Glyphosate, Thioallophanate, quinalphos and methylcarbamoyl, TSP (Triple Supper Phosphate) and vinyl phosphate.

In addition to above mentioned occupation related risk factors history of smoking (OR=1.92 ,95% CI 1.37–2.71) (P<0.01), history of alcohol consumption (OR=1.63 ,95% CI 1.16–2.31) (P<0.01), and history of snake bites (OR=1.44 ,95% CI 0.96–2.159) (P<0.10) were identified as other possible risk factors when compared with healthy individuals. Ninety two (92%) of the cases showed the drinking water source in the water from shallow wells (OR=1.62, 95% CI 0.92–2.85) (P<0.10) and there is evidence to suggest that water quality plays a role in the development of the disease which could be identified as a contributory risk factor. Other drinking water sources are Tube wells, natural springs and reservoirs etc.

Genetic predisposition also play a major role in CKDu and having a family history of renal dysfunction due to CKD-U (OR=2.66 ,95% CI 1.79–3.96) (P<0.01) also a significant risk factor. Pedigrees were drawn for three generations to identify the likely inheritance of the disease considering the precision of the available data and thirty six percent of the CKD-U patients and family history of renal dysfunction suggests a genetic etiology of the disease but could also be due to the same geo-environmental exposure such as drinking same water or consuming same food items etc.

History of malaria (OR=1.26,95% CI 0.89–1.774), use of Aurvedic or indigenous medicine (OR=1.194,95% CI 0.79–1.81) and betel chewing (OR=0.789,95% CI 0.56–1.12) were not significant risk factors (P <0.10) when compared to cases and controls in the study.

Table 4.1. Odds ratios of risk factors

Risk factor (variable)	Level	Control		Case		Odds ratio	Probability	Confidence Interval (95%)
		count	percent	count	percent			
Involvement in agricultural practices	No	60	22%	32	12%	1	0.001	1.33–3.381
	Yes	214	78%	242	88%	2.12		
Status of protective measures against agro -chemicals	High	72	42%	14	8%	1	0.0001	4.49–5.59
	Low	98	58%	160	92%	8.4		
Agricultural activities with low labour hiring	Yes	52	24%	114	47%	2.77	0.0001	1.85–4.14
	No	162	76%	128	53%	1		
Cultivating land area	≤2ha (low)	78	36%	48	20%	1	0.001	1.52–3.53
	>2 ha(high)	136	64%	194	80%	2.32		
History of Smoking	Yes	102	37%	146	53%	1.923	0.0002	1.37–2.71
	No	172	63%	128	47%	1		
History of Alcohol consumption	Yes	93	34%	125	46%	1.63	0.0052	1.16–2.305
	No	181	66%	149	54%	1		
History of malaria	Yes	159	58%	174	64%	1.26	0.1893	0.89–1.774
	No	115	42%	100	36%	1		
History of snake bites	Yes	52	19%	69	25%	1.44	0.0796	0.96–2.159
	No	222	81%	205	75%	1		
Drinking water source	Shallow dug wells	240	88%	252	92%	1.62	0.0895	0.92–2.85
	Others	34	12%	22	8%	1		
History of Ayurvedic treatments	Yes	53	19%	61	22%	1.194	0.3997	0.79–1.81
	No	221	81%	213	78%	1		
History of Betel chewing	Yes	182	66%	167	61%	0.789	0.1826	0.56–1.12
	No	92	34%	107	39%	1		
Family history of renal dysfunction	Yes	48	18%	99	36%	2.663	0.0001	1.79–3.962
	No	226	82%	175	64%	1		

4.2 Multivariate analysis

In univariate analysis data from the case control study can be analyzed using linear logistic regression modeling. As in the case of linear regression, the strength of the modeling technique lies in its ability to model many variables simultaneously. This will be referred to as “multivariate case”. Central to the concentration of multiple logistic models will be estimation of the coefficients in the model and testing for the significance. Since the design was a matched case control design the multivariate analysis referred to conditional logistic regression techniques and it was used to investigate the relationship between the outcome of being a case or a control and a set of prognostic factors.

The dependent variable of interest, presence of the CKD was constructed as above independent variables in a conditional logistic regression model.

Figure 4.1 shows the "Model Fit Statistics" table of the SAS standard output which contains the Akaike Information Criterion (AIC), the Schwarz Criterion (SC), and the negative of twice the log likelihood (-2 Log L). The output contains theses statistics for the intercept-only without covariate model and the fitted linear logistic model. AIC and SC can be applied to compare different models, and a model which has variables having good significance level should result in smaller values. The "Testing Global Null Hypothesis: BETA=0" table (Figure 4.1) shows the test results of the likelihood method and the efficient score test for testing joint significance of independent variables or predictors. According to the SAS output the small *p*-values (P<0.01) reject the hypothesis that all slope parameters are equal to zero. The likelihood ratio of the fitted model was significant.

Table 4.2 shows the "Analysis of Maximum Likelihood Estimates" which contains the results of the Wald test for individual parameters and lists the parameter estimates with their standard errors. Conditional logistic regression model will not report an intercept in the output.

Model Fit Statistics

Criterion	Without Covariates	With Covariates
AIC	379.845	269.167
SC	379.845	299.311
-2 Log L	379.845	255.167

The LOGISTIC Procedure

Conditional Analysis

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	124.6774	7	<.0001
Score	101.5500	7	<.0001
Wald	66.2074	7	<.0001

Figure 4.1 : The "Model Fit Statistics" table of the SAS output

Table 4.2 : Analysis of Maximum Likelihood Estimates

Parameter	df	Estimate	Standard error	Wald Chi-Square	P value
Family history of CKD-U	1	1.0329	0.2576	16.0831	<.0001
Involvement of agricultural practices in > 2 ha	1	1.2281	0.2790	19.3691	<.0001
Agricultural activities with low labour hiring	1	1.4777	0.2757	28.7276	<.0001
Status of protective measures against agrochemicals	1	0.6758	0.2817	5.7544	0.0164
Smoking	1	1.0823	0.3395	10.1631	0.0014
History of Snake bites	1	0.4712	0.2757	2.9212	0.0874
Drinking water source	1	0.7176	0.3679	3.8048	0.0511

According to the test results of the conditional logistic regression model in Table 4.3, the involvement in agricultural practices in greater than 2 hectares of land area in extent, (OR=3.42, 95% CI 1.98 to 5.90, P<0.001), agricultural activities with low labour hiring (OR=4.38, 95% CI 2.55 to 7.52, P<0.001) and low protective measures against agrochemicals (OR=1.97, 95% CI 1.13 to 3.41, P<0.05), were identified as significant risk factors for CKD-U.

Meanwhile the occupational exposures giving rise to CKD in the region, some other significant risk factors could be found in multivariate analysis also. Having a family history of renal dysfunction due to CKD-U (OR=2.809, 95% CI 1.696-4.654. P<0.01), history of smoking (OR=2.951, 95% CI 1.517-5.741 P<0.01), drinking shallow dug well water (OR=2.050, 95% CI 0.997-4.215 P<0.05) and history of snake bite (OR=1.602, 95% CI 0.933-2.750,P<0.1) were identified as contributory risk factors. Whereas history of alcohol consumption, history of malaria, betel chewing and regular use of Ayurvedic or indigenous treatments didn’t show a significant association (P>0.1) to CKD-U when compared with the control group. Therefore according to above risk factors most of the variable in this region mainly affects from poor socio-economic backgrounds.

Table 4.3 : Odds ratios of agricultural and life style related risk factors in conditional logistic regression

Variable	Level	Control		Case		Odds Ratio	Probability	Confidence Interval (95%)
		count	percent	count	percent			
Involvement of agricultural practices in more than 2ha of land extent	Yes	136	50%	194	71%	3.42	0.0001	1.98–5.90
	No	138	50%	80	29%			
Agricultural activities with labour hiring	Low	52	19%	114	42%	4.38	0.0001	2.55–7.52
	High	222	81%	160	58%			
Status of protective measures against agro -chemicals	Low	102	37%	162	59%	1.97	0.0164	1.13–3.41
	High	172	63%	112	41%			
Family history of renal dysfunction	Yes	48	18%	99	36%	2.81	0.0001	1.70–4.65
	No	226	82%	175	64%			
Drinking water source	Shallow dug wells	240	88%	252	92%	2.05	0.0501	1.00–4.22
	Other	34	12%	22	8%			
Regular Smoking	Yes	102	37%	146	53%	2.95	0.0014	1.52–5.74
	No	172	63%	128	47%			
Snake bites	Yes	52	19%	69	25%	1.6	0.0874	0.93–2.75
	No	222	81%	205	75%			

Estimates of probability of the disease occurrence under each factor were calculated in the case and control separately. The distribution of disease occurrence probabilities in the case and control sample was as Figure 4.3.

Box plot diagram (Figure 4.2) shows the dispersion and central tendency of risk probability of the disease occurrence in the diseased and non diseased sample of the study. The case population shows the high risk and the dispersion is relatively low compared to the control.

Figure 4.3 showing the number of patients with and without the disease arranged according to the calculated risk probability and the distribution overlap the risk probability does not distinguish normal from disease with 100% accuracy. The area overlap indicates where the test cannot distinguish normal from disease.

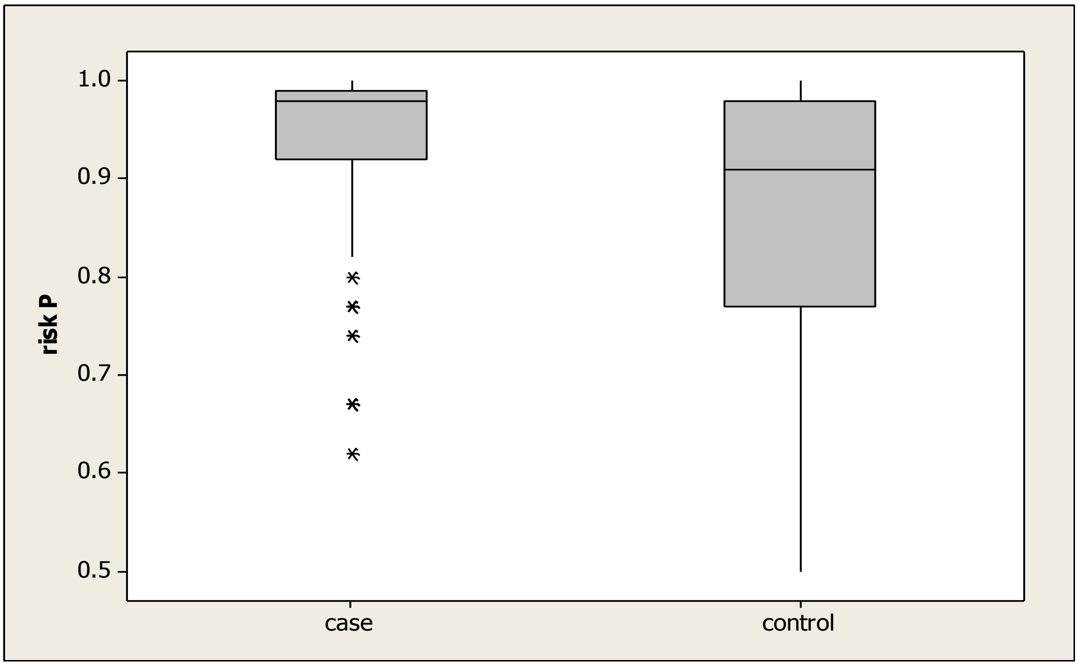

Figure 4.2: Box plot diagram of risk probabilities of the disease occurrence in case and control

The risk probability could be used as an index of the disease diagnosis. But it can be used with a determined threshold level. The true and false positive rates in turn, determine the magnitude of the revision of the disease's probability. The receiver operating characteristic (ROC) curve in figure 4.4 was constructed by varying the cut-point that determines which estimated event probabilities are considered to predict the event.

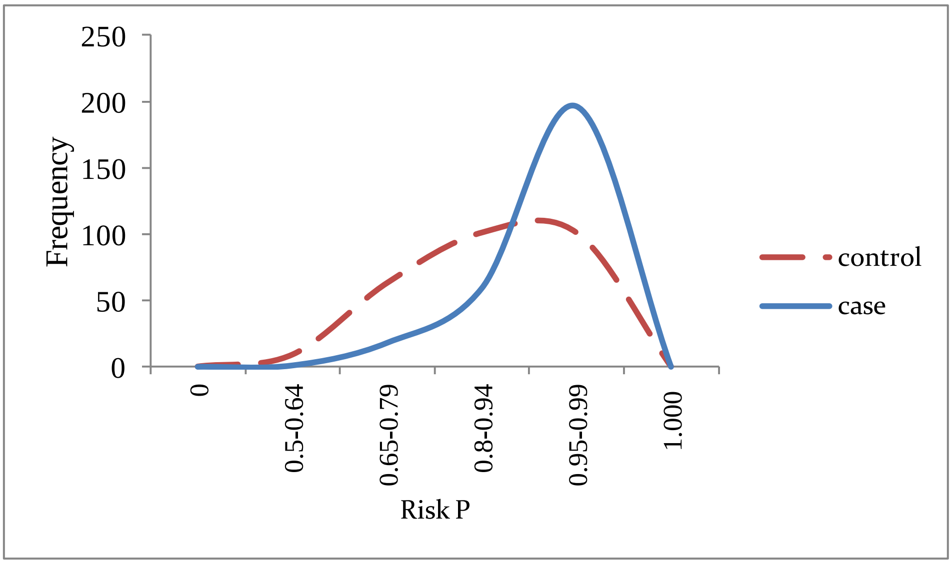

Figure 4.3: The distribution of risk probability for diseased and non-diseased samples

Table 4.4: TPF and FPF at each cut off level

Cut point	FPF	TPF
0.65	0.967	0.996
0.70	0.781	0.960
0.75	0.770	0.953
0.80	0.737	0.931
0.85	0.726	0.920
0.90	0.515	0.788
0.95	**0.369**	**0.715**
0.99	0.066	0.263

The receivers operating characteristic (ROC) curve plots the true positive fraction (TPF) against the false positive fraction (FPF) for different choice of cut off ranges. The slope of the ROC curve equals the likelihood ratio at that point. For a low cut off value almost all patients receive positive test results. This produces a high sensitivity, but low specificity for a high cut-off value, almost all patients receive negative test results (table 4.4). This produces a low sensitivity, but a high specificity. As the cut off increases from minus infinity to positive infinity, values along the ROC curve move from upper right to lower left corner of figure 4.4. Coordinates of the above ROC curve at each cut off point is as in the table 4.3 and the optimum cut off point will be 0.95.

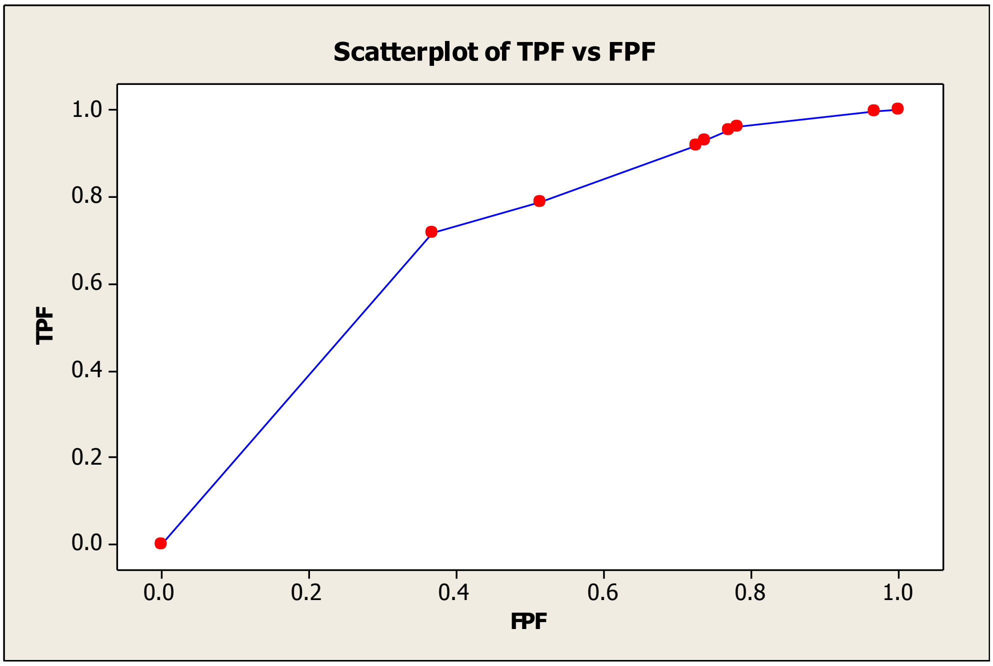

Figure 4.4 : The ROC curve for the diagnosis of the disease with risk probability

CHAPTER FIVE

5.0 DISCUSSION AND CONCLUSION

According to the results of the study we can find that there is a risk of developing the CKD-U in the north central region of Sri Lanka with multi-factorial risk factors involving certain agricultural activities related practices and certain life style related risk factors. Since the etiology of the CKD-U is likely to take a longer period of time, it is important to take preventive measures to minimize exposure to possible risk factors until the etiology is established.

Agrochemicals are routinely used in the rice production systems and as mentioned in the results Paraquat, Bispyribac sodium, Glyphosate, Thioallophanate, TSP (triple supper phonphate), methylcarbamoyl vinyl phosphate urea and quinalphos are widely used agrochemicals by farmers. Low protective status had significant contribution to the disease than the high protected farmers. Frequency of application of agrochemicals and dosages or quantity of chemical application was not considered in this study.

In univariate analysis we can find both smoking and alcohol consumption are associated with the disease and addition of metabolites and some toxins might be taken into the body from theses sources. In multivariate analysis also smoking can be identified as a significant risk factor. In most of the lifestyles, smoking and alcohol consumption are two mutually interrelated behavioral variables.

According to table 4.3 subjects with a family member with CKD-U were 2.81 times more likely to have the disease compared with those without such a family history. However, a clear Mendelian inheritance could not be indicated due to exposure of the siblings to the etiological agent rather than direct inheritance of the disease (Jayasekara et al.2013a) but it could also be due to the same environmental exposure such as drinking water source and having same food items etc. considering the reliability of data related to the family history of the disease.

Anuradhapura and Polonnaruwa are two districts in the North central region which is in the dry zone of Sri Lanka and it is very dry and hot compared to other parts of the country. Most of the farmers work in the open air environment who exposed to direct sunlight with dry air.

The odds ratio of agricultural workers is higher in both univariate and multivariate analysis. Thus, inferring that water and its attributes may play a role in the development of the disease. People who work under the hardest conditions sweat more and, as such, consume more water. Agriculture workers in North Central Region has an high exposure to sun and heat, and dehydration as critical factors associated with the occurrence of CKD. On the other hand pphysical exertion is higher in agricultural workers who work in more than 2 hectares and cultivating without labour exchange. According to a study conducted in banana and sugarcane agricultural workers in Niaragua by Torres *et al* possibility of chronic dehydration and electrolyte loss for the pathogenesis of renal disease is higher in working people with a heavy workloads in hot climatic conditions often lead to chronic dehydration. Therefore further studies will be important to find out the effect of dehydration on chronic kidney disease in north central region of Sri Lanka.

The risk probability could be used as an index of the disease diagnosis. For the diagnosis the distributions of risk probability result values in actually positive and in actually negative patients overlap, and no single threshold or decision criterion can be found that separates the populations cleanly; otherwise the test would be perfect. Usually a threshold value must be chosen arbitrarily, and different choices will yield different frequencies for the various kinds of correct and incorrect decisions. A threshold value must be selected that is believed to yield an appropriate compromise among these gains and losses. The concept of confidence threshold may be hard to quantify, but in most situations a confidence threshold can be varied, and the various decision fractions will vary with it. Recognizing the arbitrary nature of decision threshold selection might seem to complicate the problem even more.

Accuracy should be meaningful might be tempted to suppose that though this is true, accuracy should be meaningful at least as an index for comparison of diagnostic techniques applied to the population in which disease prevalence is known and fixed. However, here the index is limited. Two diagnostic modalities can yield equal accuracies but perform differently with respect to the types of correct and incorrect decisions they provide; the incorrect diagnoses from one might be almost all false negative decisions (misses), while those from the other might be nearly all false positive decisions (false alarms), and clearly, the usefulness of these two tests for patient management could be quite different in various situations. Though accuracy provides a single simple number for diagnostic performance, it is often too simple and must be interpreted with considerable caution. The limitations of this index force us to introduce some complexity into our evaluation scheme: We must sort out the effect of disease prevalence, and we must score separately the various kinds of right and wrong diagnostic decisions. Sensitivity and specificity represent two kinds of accuracy: the first for actually positive cases and the second for actually negative cases.

ROC is a plot of the TPF (true positive fraction) against the FPF (false positive fraction) for the different possible cutoff levels of a diagnostic experiment. If we explicitly change the decision threshold level different times as described above table 4.4, several different pairs of TPF and FPF can be obtained. These pairs were plotted as "y" and "x" coordinate values of points on a graph such as that shown in Figure. 4.4. The axes of this graph both range from zero to one because these are the limits of possible TPF and FPF values. Since we can imagine repeatedly changing the decision threshold and obtaining more and more points on this graph, and since TPF and FPF must always change together in a way determined by the test result distributions, the points representing all possible combinations of TPF and FPF must lie on a curve. This curve is called the receiver operating characteristic (ROC) curve for the diagnostic test, since it describes the inherent detection characteristics of the test (or, for subjective studies, the observer- test combination) and since the receiver of the test information can operate at any point on the curve by using an appropriate decision threshold. Fig. 4.4 shows three possible operating points that might correspond to use of a strict threshold (case called positive only if judged almost definitely positive); of a moderate threshold; or of a lax threshold (case called positive if any suspicion of disease). Conventional ROC curves of the kind described here (in which two actual states are possible and in which two decision alternatives are available) inevitably must pass through the lower left corner (FPF = 0, TPF = 0) of the graph because all tests can be called negative, and through the upper right corner (FPF = I, TPF = 1) of the graph because all tests can be called positive. Also, if the test provides information to the decision maker, the intermediate points on a conventional ROC curve must be above the major (lower left to upper right) diagonal of the ROC space, because in that situation a positive decision should be more probable when a case is actually positive than when a case is actually negative. i.e. the closer the curve follows the left-hand border and then the top border of the ROC space, the more accurate the test whereas when it comes to the 45-degree diagonal of the ROC space, the less accurate the test.

Therefore the optimal threshold level of risk probability was 0.95 and we can show on theoretical grounds that if the decision maker uses available information in a proper way, the slope of the ROC curve must steadily decrease as one moves up and to the right on the curve and the area under the ROC is a measure of text accuracy. However the threshold value has to be further evaluated and validated with more samples.

The study shows that there is a high risk of developing CKD-U with certain agricultural practices. As identification of the etiological agent is likely to take a longer period, minimizing the exposure to likely risk factors may help controlling the disease. Farmer education to minimize risk factors like application of chemicals, alcohol consumption, and smoking will help preventing the disease. Further studies are needed to identify the effect of dehydration and physical exertion on the renal functions to identify methods to minimize these effects. Risk probability of the multiple linear logistic regressions can be used as an index of the disease occurrence.

REFERANCES

Arachchige, S.J., 2012 . Overview of causal factors for chronic kidney disease (CKD) in Sri Lanka. *Sri Lanka plant protection Industry* 23: 75-77.

Athuraliya, N.T.C.,Tilak, D.J., Abeysekera, P.H.A., Kumarasiri, R., Bandara P., Karunaratne, U., Milton , A.H. and Jones , A.L., 2011. Uncertain etiology of proteinuric-chronic kidney diseasein rural Sri Lanka. *Kidney International* 80: 1212-1221.

Bandara, J.M,R,S., Seneviratthna D. M.. A. N., Dasanayake, D.M.R.S.B. Herath, V., Bandara, J.M.R.P., Abeysekara, T. and Rajapaksha, K.H. 2008. Chronic renal failure among farm families in cascade irrigation systems in Sri Lanka associated with elevated dietary cadmium levels in rice and freshwater fish (Tilapiya). *Environmental Geochemistry and Health* 30: 465-478.

Bandara, J.M,R,S., Wijewardena, H.V.P., Liyanage, J., Upul, M.A and Bandara, J.M.U.A 2010. Chronic renal failure in Sri Lankacausedby evaluated dietary cadmium: Trojan house of the green revolution. *Toxicology Letters* 198: 33-39.

Bandara, J.M,R,S., Wijewardena, H.V.P.,Bandara, Y.M.A.Y., Jayasooriya, R.G.P.T and Rajapaksha, H. 2011. Pollution of River Mahaweli and farmlands under irrigation by cadmium from agricultural inputs leading to a chronic renal failure epidemic among farmers in NCP, Sri Lanka. *Environmental Geochemistry and Health* 33(5): 439-453.

Cederborg,M-A 2011. Kidney Disease- Silent Killers. World Kidney Day Technical Brochure, 10 March, 2011.

Chandrajith R., Mahatantila, K., Jayasena, H.A.H. and Tobschall, H.J. 2008. Geochemical characteristics of sediments from a reservoir (tank) ecosystem in Sri Lanka. *Paddy water Environment* 6: 363-371.

Chandrajith, R., Seneviratna, S., Wickramaarachchi, K., Attanayake, T., Aturaliya, T.N.C and Dissanayake, C.B. 2010. Natural radio nuclides and trace elements in rice field soils in relation to fertilizer application Study of a Chronic Kidney Disease (CKD) area in Sri Lanka. *Environmental Earth Sciences* 60 (1) : 193-201.

Chandrajith, R., Nanayakkara S., Itai K., Aturaliya, T.N.C, Dissanayake, C.B., Abeysekara, T., Harada, K., Watnabe, T and Koisumi, A. 2011a. Chronic kidney diseases of Sri Lanka: Geographic distribution and environmental implications. *Environmental Geochemistry and Health.* 33:267-277.

Chandrajith , R., Dissanayake, C.B., Ariyaratne, T., Herath , H.M.J.M.K. and Padmasiri, J.P. 2011b. Dose-dependent Na and Ca in fluoride rich drinking water- A major cause of chronic renal failure in tropical arid regions. *Science of the Total Environment* 40(9): 671-675.

Collett, D., 1991. *Modeling Binary Data*. London: Chapman & Hall.

Cyrus, R.M., Nitin R. P., Pralay,S.,2012. Efficient Monte Carlo Methods for Conditional Logistic Regression. *Journal of the American Statistical Association*, Volume 95,[online] Available at http://amstat.tandfonline.com/doi/abs/10.1080/01621459.2000.10473906> [Acceessed 2 October 2013].

Desalegn, B., Nanayakkara S., Harads, K.H., Hitomi, T., Chandrajith R., Karunaratne U., Abeysekera, T and Koizumi, A. 2011. Mycotoxin detection in urine samples from patients with chronic kidney disease of uncertain etiology in Sri Lanka. Bulletin of Enviornmental Contamination and Toxicology 87: 6-10.

De Silva N,(2011), Arsenic, *The island* [Online] 7th June, 2011. Available from: http://www.island.lk/index.php?page_cat=article-details&page=article-details&code_title =27221 [Accessed: 22nd May 2012].

Dissanayake, C. B. (1996) Water quality and dental health in the Dry Zone of Sri Lanka. In JD Appleton, R Fuge, GJH McCall (Eds.), Environmental geochemistry and health. Geological Society UK, Special Publication, No. 113: 131–114.

Dissanayake, D.M. n.d. *The Cyanobacterial toxins: a hidden hazrd*. Available from < http://www.asiatox.org/9th%20apamt/OP/THE%20CYANOBACTERIAL%20TOXINS%20%20A%20HIDDEN%20HEALTH%20HAZARD.pdf/>.[Acceessed 2 October 2013]

Dharmaratne D, Gunawardane A, Marawila T D, Ranaweera Banda R M,Steele P, Weerakkody C. Poverty and Social Impact Analysis (PSIA) of Sri Lanka's Land Reform. Institute of Policy Studies of Sri Lanka 2006.

Environment Action Plan.2002. Final report submitted on analysis of water for major pesticides used in the dry zone of Sri Lanka to the Ministry of Environmental Authority under the Environment Action Plan 1, 1-26.

Fernando S, 2004. *Land and water in Sri Lanka: To whom they belong? .Law and Society Trust*. 14: 201.

Goonarante, I.K., Ranaweera, A.K.P., Liyanarachchi, N.P., Gunawardane, N. and Laneralle, R.D. 2008. Epidemiology of chronic kidney disease in a Sri Lankan population. *International Journal of Diabetes in Developing Countries* 28(2): 60-64.

Herath, H.M.N., Wazil, J.A.W.M., Abeysekara, D.T.D.J., Jeewani, N.D.C., Weerakoon, K.G.A.D., Ratnatunga, N.V.I., Bandara, E.H.C.K and Kalratne, S.A.M. 2011. Chronic kidney injury inSri Lanka : A descriptive study. Postgraduate Medical Journal doi: 10.1136/postgradmedj-2011-130225.

Illeperuma, O.A., Dharmagunawardhane, H.A. and Herath, K.P.R.P. 2009. Dissolution of aluminimum from substandard utensils under high florid stress: A possible risk factors for chronic renal failures in the North Central Province. Journal of the National Science Foundation of Sri Lanka 37: 219-222.

Ileperuma, O.A. 2011. Chronic renal failure in NCP and Arsenic: Science versus myth. *The Sunday Times*, 31 July 2011.

Jayakodiarachchige, S., 2012, Overview of causal factors of chronic kidney disease (CKD) in Sri Lanka.Crop life: Sri Lanka Plant Protection Industry Journal 5 (12) 75-77

Jayasumana, M.A.C.S., Paranagama, P. and Amarasinghe, M., 2011. Chronic kidney disease of unknown etiology and arsenic in ground waterin Sri Lanka. Paper presented in workshop on challenges in groundwater management in Sri Lanka, March 15, Colombo , Sri Lanka.

Jayatilake N, Mendis S, Maheepala P, Mehta F R.,2013, Chronic kidney disease of uncertain aetiology: prevalence and causative factors in a developing country,*Journal of medical case report*,< http://www.biomedcentral.com/1471-2369/14/180>.[Accessed 2 October 2013].

Jayasekara J M K B, Dissanayake D M, Gunaratne M D N, Thilakarathna S, and Sivakanesan R. 2013a. Agricultural and life style related risk factors of chronic kidney disease of unknown etiology: case control study. *International Journal of Medical and Pharmaceutical Sciences*, 3 (12), 21-29.

Jayasekara J M K B, Dissanayake D M, Adhikari S B and Bandara P. 2013b. Geographical distribution of chronic kidney disease of unknown origin in North Central Region of Sri Lanka. *Ceylon Medical Journal* 58: 6-10.

Jones, MD n.d., *Commentary on indigenous housing initiatives*. Available from: <http://www.architecture.com.au>. [Accessed 6 June 2009].

National Kidney Foundation 2013.*Kidney Disease*. [online] Available at <http://www.kidney.org/kidneydisease/aboutckd .cfm > [Accessed 2 October 2013]

Nerds, W., 2014. Machine Learning The Complete Guide According to Wikipedia: Everything you need to learn about Machinery [e-book] Singapore:Brighthouse. Available at : Google Books < http://books.google.lk /books?id=4ZayAwAAQBAJ&printsec =frontcover#v =onepage&q&f=false>[Acceessed 2 March 2014]

Orth S.R, and Hallan S.I., 2008, Smoking: A Risk Factor for Progression of Chronic Kidney Disease and for Cardiovascular Morbidity and Mortality in Renal Patients—Absence of Evidence or Evidence of Absence?, *Clinical journal of American Society of Nephrology*, <http://cjasn.asnjournals.org/content/3/1/226.full >.[Accessed 2 October 2013].

Paneth N, Susser E, Susser M , 2002. Origins and early development of the case-control study: part 2 ; Soz.- Präventivmed. Available at :
<http://www.epidemiology.ch/history/papers/SPM%2047(6)%20359-65%20Paneth%20et% 20al.%20%20Part%202.pdf> [Accessed 2 October 2013].

Wanigasuriya, J.K., Peris –John, R.J., Wickramasinghe, A.R. Dissanayake, W.P. and Hittarage, A. 2007. Exposure to acetylcholineestase-inhibiting pesticides and chronif renal failure. Celon Medical Journal 51: 42-43.

Seneviratne L.,Abeyselera, T., Nanayakkara, S., Chandrajith , R., Ratnatunga, N., Harada, K.H., Hiltomi. T., Korniya, T., Muso, E and Koizumi, A. 2011. Risk factors associated with disease progression and mortality in chronic kidney disease of unknown etiology: a cohort study in Medawachchiya, Sri Lanka. *Environmental Health and Preventive Medicine* 2011 Sep 1:1-8.

Sinha, S., & Mukherjee, B., 2005: A Score Test for Determining Sample Size in Matched Case-Control Studies with Categorical Exposure; *Biometrical Journal* 48 ,1, 35–53.

Torres C, Gonzalez M, Vanegas R, Aragon A. Prevalence of chronic renal disease in the communities of ‘‘La Isla’’ and ‘‘Candelaria’’, Chichigalpa, June 2008. Leon, Nicaragua: Universidad Nacional Autonoma de Nicaragua; 2008 (in Spanish).

Warnakulasuriya, K.A.A.S., Balasuriya, S., Perera, P.A.J., and Pieris, L.C.L., 1992. Determining optimal levels of fluoride in drinking water hot, dry climates- a case study in Sri Lanka. *Community Dental Oral Epidemiology* 2:364-367.

Wichramasinghe, R., Peris-John, R.J. and wanigasuriya , K.P.2011. Chronic kidney didease of unknown etiology in Sri Lanka: Trying to unravel the mystery. *Ceylon Medical Journal* 56(4): 143-146.

World Health Organization & Ministry of Health and Nutrition 2008. Chronic kidney disease of unknown etiology (CKDue) : a new threat to health. p4.

Mehtaa, C. R.1, Patelb, N. R. and Senchaudhuria, P.(2000) Efficient Monte Carlo Methods for Conditional Logistic Regression, *Journal of the American Statistical Association* <http://amstat.tandfonline.com/doi/abs/10.1080/01621459.2000.10473906>[Accessed 2 February 2013].