MODEL- BASED ASSESSMENT OF TRENDS FOR SOME COMMUNICABLE DISEASES IN KENYA

By

Rutto, Kiprotich James

A dissertation submitted in partial fulfillment of the requirements for the degree of master of science in statistics of Egerton university.

Oct,1999

APPROVAL PAGE

This is my original work and has NOT been presented in any other institution for award of any degree.

Signature: ....

James Kiprotich Rutto.

SUPERVISORS

Prof. Dankit Nassiuma

Mathematics Department

Egerton University.

Signature: ....

Date: ..../8/2000

Dr. Ali Salim Islam

Mathematics Department

Egerton University.

Signature: ....

Date: 9/8/2000

ACKNOWLEDGEMENTS

First and foremost, I thank my God who gave me good health, salvation, strength and patience during the entire period of my study.

I wish to express my deep appreciation to my supervisor Professor D. Nassiuma for his guidance and encouragement in preparation of this thesis. He has been both my mentor and a blessing to me. I would like to thank Dr. Ali Salim Islam for his useful comments, recommendations and suggestions which greatly improved my thesis. In addition, I am grateful to my lecturers: Prof Nassiuma, Dr. Ali, Prof. Kip'ngeno, Dr. Onyango and Dr. Njuho for teaching me what it takes to be a statistician.

Finally, I would like to thank my Parents, Wife, Son, Cherutoi, Amos and all my family members for their support, understanding and prayers during my studies.

ABSTRACT

Communicable diseases are the major causes of morbidity and mortality in Kenya. They retard both social and economic development since an unhealthy nation is riddled with poverty and underdevelopment. The inability to control and manage these diseases has been attributed among others to lack of resources, change of social patterns, lack of public health education and abject poverty. Subjective assessments of trends have thus been the basis for decision making regarding management of these diseases. In fact, actions are mostly taken when there are major outbreaks.

Statistical models are important tools that are used in analyzing causal relationships. In this study, statistical models are applied in assessing trends and forecasting of future trends of some communicable diseases in Kenya namely: Tuberculosis, Measles, Hepatitis, Tetanus and Poliomyelitis. These models were fitted to count-data based on reported cases of people infected by each of the five communicable diseases. The assessment of trends was done using generalized linear models while time series models were used to forecast future trends. The rate of infection of a particular type of communicable disease was evidently provided during the analysis and this issue is expected to contribute to the control and management of these diseases.

TABLE OF CONTENTS

TABLE OF FIGURES

LIST OF TABLES

CHAPTER ONE

INTRODUCTION AND LITERATURE REVIEW

1.0 INTRODUCTION

Infectious diseases are a continual threat to an individual's complete physical, mental and social well-being from birth until death. In addition to the physical suffering, pain, grief and death caused, infectious diseases lead to social and economic hardships, isolation and discrimination. From the economic point of view, medical interventions divert scarce national resources from other important uses such as building hospitals and implementation of other development projects.

The most common diseases in Kenya are the infectious diseases especially those which are spread from one person to another person commonly referred to as communicable diseases. Incidences of such diseases have drawn a lot of concern from government health authorities, world health organization (WHO) and other international bodies. For instance the spread of the acquired immunodeficiency syndrome (AIDS) continues unabated globally because there is no cure and it is 100% fatal. An infectious disease which poses an ominous health threat to the entire Kenyan population is Tuberculosis (TB) caused by *Mycobacterium tuberculosis*. Incidences of TB are expected to rise faster because of the AIDS epidemic. Tuberculosis kills more youth and adults than any other infectious diseases in the world today. It is a bigger killer than malaria and AIDS combined, and kills more women than all combined causes of maternal mortality. It is estimated that between now and the year 2020, nearly one billion more people will be newly infected, 200 million people will get sick and 70 million will die from TB if control is not strengthened (WHO bulletin, 1998). The major reason why TB is advancing so fast is because people refuse to seek medical attention during the early stages of infection.

However, in Kenya some epidemics have been eradicated for example small pox (Benenson, 1990). There is also a decline in the incidence of diphtheria and paralytic poliomyelitis brought about by world wide extended immunization program (EPI).

1.1 COMMUNICABLE DISEASES

Communicable diseases are those infectious diseases caused by specific infectious agents or their toxic products carried over through transmission from one infected person to another either directly or indirectly. They are the major causes of mortality and morbidity in Kenya. Unfortunately there is still a long way to go in terms of medical advances to eradicate communicable diseases completely. This is not only true for Kenya but also applicable in the developed countries as well. For instance, a disaster of major proportions struck southern Italy in 1976 when cholera caused great suffering, took a heavy toll of human life and snuffed out a thriving tourist business (Carrol, etal, 1976). In U.S.A, measles afflicted college students in Dakota while in Germany, polio crippled school boys in Connecticut in 1970 (Witte, 1977).

There are numerous communicable diseases in Kenya but we shall consider five of them namely: Tetanus, Infectious Hepatitis, Measles, Poliomyelitis, and Tuberculosis.

1.1.1 Tetanus

Tetanus is an acute disease which is caused by tetanus bacillus (*Clostridium tetani*) which grows anaerobically at the site of an injury or a wound. It is characterized by painful muscular contractions primarily of the masseter and neck muscles, secondarily of trunk muscles. The first sign is abdominal rigidity, though rigidity is confined to the region of the injury. The fatality rate ranges from 30% to 90% being highest in infants and the elderly.

The infectious agent (Tetanus bacillus) stays and grows in intestines of man, soil or formites (dishes, laundry, bedding, clothes and personal effects) contaminated with human feces. Therefore if a person with a punctured wound/injury steps on the contaminated feces, the tetanus spores will find their way into the body through the wound.

The incubation period is usually three to twenty one days, although it may range from one day to several months depending on the character, extent and the location of the wound. Most cases occur within fourteen days. In general, shorter incubation periods are associated with more heavily contaminated wounds, more severe diseases and worse prognosis.

Preventive measures usually include educating the public on the necessity of complete immunization with tetanus toxoid, the kinds of injury particularly liable to be infected by tetanus bacillus and the potential of active and passive prophylaxis. Prophylaxis in wound management is based on careful assessment of whether the wound is clean or contaminated, the immune status of the patient, proper use of tetanus toxoid, wound cleansing and proper use of antibiotics. Whereas complete immunization means that tetanus toxoid is administered at booster doses every ten years or if this standard schedule is impractical, then the third successive doses can be done at the intervals of one month or more. The toxoid is usually administered together with diphtheria toxoid and pertussis (DTP) vaccine as a triple antigen or double diphtheria toxoid (DT) antigen for children under seven years or tetanus toxoid (TT) for older people. In Kenya DTP, DT and TT are available combined with inactivated polio vaccine.

Persons who have not completed a full primary series of tetanus toxoid require a dose of toxoid as soon as possible after the development of the wound and may require passive immunization with tetanus antitoxin if it is a major wound or if it is heavily contaminated. DTP, DT or TT as determined by the age of the patient and previous immunization history is used when there is a wound and ultimately to complete the primary series.

Active protection is maintained and this aspect is important for workers in contact with soil, sewage and domestic animals, members of the military forces, policemen and others with greater than usual risk of traumatic injuries.

Other control measures include the search for contaminated drugs such as heroine and cocaine, increased need for tetanus antitoxin and toxoid for injured patients as a result of social upheavals (wars, riots) and natural disaster (floods, hurricanes, earthquakes). Lastly the Kenyan Ministry of Health has recommended that every individual should seek for an upto-date immunization against tetanus and pregnant women ought to go for vaccine-induced maternity (KMAR, 1994) which is important in preventing neonatal tetanus.

There is currently considerable evidence to show that in Kenya, the disease has been contained due to the widespread immunization campaign that has been going on in the country though some cases have been reported as a result of injuries considered too trivial for medical consultations. In January 1994, a team from World Health

Organization (WHO), Ministry of Health (MOH), Centres for disease control (CDC-USA) assessing Kenya Expanded program on immunization (KEPI) quality of surveillance and level of EPI disease control concluded that KEPI had developed one of the best immunization programs in Africa and brought the country close to the goals of neonatal Tetanus elimination and Polio eradication (WHO annual report, 1994).

1.1.2 Infectious hepatitis

Infectious hepatitis is caused by hepatitis A virus (HAV) which is a 27-nm picornavirus (RNA virus). This virus is classified as enterovirus type 72, a member of the family picornaviridae. This disease starts abruptly with fever, malaise, anorexia, nausea and abdominal discomfort followed within a few days with jaundice. The disease varies in clinical severity from a mild illness lasting one or two weeks to a severety disabling disease lasting several months. Many infectious diseases are asymptomatic, mild and without jaundice especially in children and recognizable only by liver function tests. The fatality rate is low (about 0.6%), the rare death usually occurs in older patients in whom the disease has a fulminant course (sudden occurrence).

Diagnosis is done by using 1gm antibodies against Hepatitis A virus in the serum of acutely or recently ill patients; 1gm may remain detectable for four to six months after onset. Also the virus and the antibody can be detected by radioimmunoassay (RIA) or enzyme-linked immunosorbent assay (ELISA). If laboratory tests are not available, epidemiological evidence can provide support for the diagnosis (Benenson, 1990).

Since the infectious agent is found in feces and reaches peak levels the week or two before onset of symptoms in an infected person, the disease is transmitted directly from one person to the other through sexual intercourse particularly among the homosexuals. Another common source has been related to contaminated water, food contaminated by infected foodhandlers including sandwiches and salads which are not cooked and are handled after cooking. Therefore if a person takes this contaminated food he becomes infected. Rare cases have been reported of transmission by transfusion of blood from a donor during the incubation period. The incubation period of this disease is from fifteen to fifty days and on average it is twenty eight to thirty days (Benenson, 1990).

Studies of infectious hepatitis transmission in humans and epidemiological evidence, indicate that there is maximum infectivity during the later half of the incubation period continuing for a few days after onset of jaundice. Most cases are probably noninfectious after the first week of jaundice. The people susceptible to these diseases are the intravenous drug abusers, school-age children, young adults and those who travel to areas where the disease is endemic. Currently, cases of infectious hepatitis have been on the increase in Kenya especially in the highly populated areas because the public is not aware of the source of infection and patients are not yet informed on how to curb this menace even after treatment. The second factor is that it has been hard for people to change their moral behaviors especially the homosexuals and those who practice prostitution.

Methods of control of this disease include good sanitation and personal hygiene with special emphasis on careful handwashing and sanitary disposal of feces. Proper treatment and sewage disposal facilities is provided at any level. If two or more cases of HAV occurs in a family, [illegible] school or any other institution then everybody in these places is to be given [illegible] globulin IG (serum) including those who travel to endemic areas. For expected exposures upto three months, a single dose of 0.02ml/kg IG or 2ml for adults is recommended; for more prolonged exposures 0.06ml/kg or 5ml should be given and repeated every four to six months if exposure continues. Disposal units, properly sterilized syringes and other equipment used for parenteral injection are usually used. Food which include fish, oysters, meat, clams are heated to a temperature of 85-90 degrees centigrade for four minutes before eating. In case of an outbreak the health authorities are notified and patients isolated especially in the first two weeks of illness, but not more than a week after the onset of jaundice.

1.1.3 Measles

Measles is an acute, highly communicable disease caused by measles virus, a member of the genus Morbillivirus of the family paramyxoviriade. Its symptoms are; promodial fever, conjuctivitis, coryza, cough and koplik spots on the buccal mucosa and it is characterized by red blotchy rash which appears in the third to the seven day beginning on the face then spread all over the body lasting four to seven days and sometimes ending in branny desquamation (peeling off). Leukopenia is also common (Benenson, 1990).

Complications may result from viral replications, bacterial superinfection and this creates other opportunistic diseases like otitis media, pneumonia, diarrhoea and encephalitis. The disease is more severe in infants and adults than in children but death can occur in children due to opportunistic disease such as pneumonia and encephalitis.

Measles can be very severe in young adults and in malnourished children, in whom it may be associated with rash, protein-losing enteropathy, mouth sores, dehydration, diarrohea, blindness and severe skin infections. The case fatality rate may be 5-10% or more (KMAR, 1994).

The infection virus can be diagnosed on clinical and epidemiological grounds. It can be confirmed by the presence of measles virus specific antibodies or a significant rise in the antibody concentrations between acute and convalescent sera. Diagnostic techniques include virus isolation from blood in cell culture, conjunctiva, nasopharynx or urine specimens taken before the third day of the rash.

Measles is one of the highly communicable diseases which is airborne and spread by droplets through direct contact with nasal or throat secretions of infected persons and less commonly by articles freshly soiled with nose and throat secretions. Its incubation period is about ten days varying from seven to eighteen days from exposure to the onset of fever. The disease is highly communicable from slightly before the beginning of the prodromal period to four days after the appearance of the rash and communicability is minimal after the second day of the rash (Benenson, 1990).

All persons who never had the disease or who have not been previously immunized are susceptible. An interesting aspect about this disease is that the immunity acquired after this disease is permanent and infants born of mothers who had the disease are immune for approximately the first six to nine months or more depending on the amount of residual maternal antibody at the time of pregnancy and the rate of the antibody degradation. Vaccination at the age of fifteen months produces 95-98% immunity of recipients and revaccination may produce immunity levels to as high as 99%.

Prior to widespread immunization, measles was common in childhood, so that over 90% of people had been infected by the age of twenty. Few people went through life without the attack. Measles was epidemic in large metropolitan communities and institutions like schools, colleges and slum areas where outbreaks tend to be more widely spread and more severe.

Currently with effective childhood programs introduced by EPI, the number of cases have reduced although there are some set-backs in this program. For instance, in 1978, the immunization program could have been 100% successful were it not for an improper immunization schedule, where the right age for a child to be immunized was not specified (Njogu, and Tukei, 1982). Consequently the children vaccinated below nine months remained susceptible. However, the current success of this program (almost 100%) is attributed to fact that health education has been intensified, the disease is under surveillance by WHO and up-to-date immunization practices based on research outcomes and developments (WHO bulletin, 1998).

All individuals susceptible to measles are vaccinated using a single injection of the live measles vaccine which can be administered concurrently with other vaccines (for mumps, rubella, smallpox). This induces active immunity in more than 95% of the susceptible individuals possibly for life by producing a mild or unapparent noncommunicable infection. The current recommendation is a routine two-dose measles vaccine schedule with the initial dose administered at fifteen months of age or slightly higher. The second dose is given at school entry-age (four to six years of age) and both doses are given as a combined measles, mumps and rubella vaccine (MMR).

The use of edmonston-zagreb strain has been recommended by WHO to be used as a vaccine. For instance, at age of six months in areas at a high risk of significant morbidity and mortality from measles, infants are to be given the same drug at age less than nine months. In addition to the above, vaccine shipment and storage facilities are handled and stored properly since dried measles vaccine is relatively stable and can be stored at refrigerator temperatures (2-8 degrees centigrade) with safety of a year or more. It is recommended that those who enter institutions such as hospitals, schools, colleges should be revaccinated unless they have a documented history of measles or they have received two doses of measles-containing vaccines.

Measles vaccines are not given to pregnant women because of the risk of fetal wastage. Also patients who cannot eat eggs because of specific allergy, those suffering from leukemia, lymphoma or generalized malignancy are not supposed to be given live vaccines.

Children are kept out of school for at least four days after appearance of the rash. In hospitals, respiratory isolation from onset of catarrhal stage of the prodromal period through fourth day of the rash reduces the exposure of other patients at risk. Strict segregation of infants is adhered to if measles occurs in an institution.

1.1.4 Poliomyelitis

Poliomyelitis is an acute viral infection caused by poliovirus. Its severity ranges from inapparent infection to nonspecific febrille illness, aseptic meningitis, paralytic disease and death. The first symptoms are fever, nausea, malaise, headache and vomiting and if the disease progresses, severe muscle stiffness of the neck and back with or without flaccid paralysis may occur. The site of the paralysis depends upon the location of the nerve cell destruction in the spinal cord or brain stem but it is characteristically assymmetric.

This disease can be diagnosed in laboratories equipped with molecular biological laboratory facilities by isolation of the virus and inoculating cell culture systems with human fecal material or o[illegible] secretions.

The disease can b[illegible] from one person to another person through close contact association (airborne). In rare instances, milk food stuffs and other fecally contaminated materials have been incriminated as vehicles. There is no reliable evidence of the spread by insects or virus-contaminated sewage. Water is rarely involved while fecal oral is the major route of transmission where sanitation is poor.

The incubation period of the disease is commonly seven to fourteen days with a possible range of three to thirty five days. It is most infectious during the first few days before and after onset of first symptoms. Poliovirus is more easily detectable and for a long period in feces than in throat secretions. The ingested virus multiplies first in the alimentary tract, viremia may then follow leading to invasion of the central nerves system and selective involvement of the motor neuron cells resulting in flaccid paralysis (Benenson, 1990).

Poliomyelitis affects mostly children, infants, adolescents and improvement in living standards has been associated with emergence of paralytic cases in older people who did not acquire immunizing infections in childhood. In children, most paralytic diseases occur at the age of three and pregnant women are very susceptible to paralytic poliomyelitis. Poliomyelitis effect depend on races, for instance in 1947, poliomyelitis outbreaks were rampant in Kenya among all races with high incidences among the Europeans than other races (KMAR, 1952).

Methods of control include immunization in early childhood. The use of both injectable noninfectious inactivated poliovirus vaccine (IPV) and live attenuated oral poliovirus vaccine (OPV) which are commercially available, give excellent protection in most populations. They are more efficient as compared to the Salk vaccine which was used extensively throughout the country before (KMAR, 1952-1957).

Patients suffering from immune deficiency states (B-lymphocyte deficiency, thymic dysplasia), immunosuppressive therapy, AIDS, lymphoma, leukemia, generalized malignancy are not to be given OPV and instead IPV is recommended. Travelers going to areas prone to polio, laboratory workers who may handle specimen containing poliovirus and health care workers who may be exposed to patients excreting poliovirus are given primary immunization.

Throat discharge, feces and articles soiled therewith are usually disinfected concurrently. Source of infection and contacts are investigated and search for sick persons, especially children is done thoroughly to achieve early detection and to facilitate control and permit appropriate treatment of unrecognized and unreported cases. Foot rot, scoliosis and other deformities resulting in functional impairment may be late manifestation of initially mild illness.

Strategically located centres are usually provided for specialized medical care of acutely ill patients and those patients with significant paralysis. Immunization against other diseases and elective surgery (especially nose and throat operations) is postponed until after immunization of poliomyelitis is done.

Recently there have been rare cases of people suffering from this disease because of perfect immunization which has been done appropriately in this country. This is one of the goals of EPI under the auspices of the WHO to reduce morbidity and mortality by providing immunization against (pertusis, tetanus, measles, poliomyelitis and tuberculosis) for every child in the world (WHO annual report, 1994). Therefore, in between July-Aug 1994 the results of an immunization coverage survey in eighteen randomly selected districts revealed that Kenya was close to WHO/UNICEF immunization coverage. One of the reasons is because they opened KEPI centres in rural areas. The other factor contributing to reduction of cases is because the disease is under surveillance by WHO. National health administration are expected to inform WHO promptly by telex, fax and to supplement these reports as soon as possible with details of the source, nature and the extend of the epidemic and the identity of the type of the epidemic virus involved. The world health assembly declared a goal to eradicate polio by the year 2000 and expects that immunization to be most efficiently provided along with other essential health services (WHO Annual report, 1994) although this has not been fully achieved.

1.1.5 Tuberculosis

Tuberculosis is a highly infectious disease caused by *Mycobacterium tuberculosis* and *Mycobacterium africanum*. In more than 80% of cases of TB the disease attacks the lungs causing pulmonary tuberculosis which is characterized by persistent cough for three weeks or more, mostly with sputum (sometimes blood stained), chest pain, fever, shortness of breath and loss of weight.

Extra-pulmonary tuberculosis affects various organs such as the lymph nodes, bones and joints, the urigenital tract, the nervous system (causing meningitis) and intestines. It is characterized by swelling of joints, headache, fever, stiffness of the neck and mental confusion (symptoms of meningitis). If untreated, about half of the patients suffering from this disease will die after a two-year period.

Diagnosis is usually done using laboratory techniques for demonstration of acid-fast bacilli stained smears from sputum or other body fluids: a positive smear justifies initiation of antituberculosis therapy. The diagnosis is confirmed by the isolation of the tubercle bacilli on culture. This also permits determination of drug susceptibility of the infectious organism. Identification and examination of suspects by direct sputum-smear examination is done in all health institutions which have a microscope with a high-power lens. Chest X-ray examination can help in the diagnosis of pulmonary TB and pleural effusion.

The incubation period of TB is from four weeks and the degree of communicability depends on the number of bacilli to the sun (ultraviolet light) and opportunities for their aerosolization by coughing, sneezing, talking and singing. A point worth noting is that patients suffering from extra-pulmonary TB and children (TB cases) hardly transmit this disease because they do not have sputum-smear being positive (Kibuga, 1995).

TB can be transmitted from person to person through exposure to bacillus in airborne droplet-nuclei produced by persons with pulmonary or laryngeal TB during respiratory efforts such as coughing, singing and sneezing. Public health nursing and outreach services for home supervision of patients (to supervise therapy directly and to arrange for examination and preventive treatment of contacts) are provided by the health authorities. The transmission of tuberculosis in a community can be reduced by providing effective chemotherapy with the highest possible cure rate to infectious patients at the earliest possible stage of the disease.

Occasionally, the community is informed on the importance of respiratory symptoms, especially persistent and productive cough, blood-stained sputum and chest pain particularly if they persist for more than three weeks and are of recent origin, i.e. the past few months or so. Also close surveillance of patients with predisposing factors e.g. HIV infection (AIDS), diabetes and malnutrition is encouraged.

If patients are prescribed with the right drugs in the right dosage and taken regularly, as prescribed for the correct period of time, then all patients will be cured (Kibuga, 1995). Preventive treatment with isoniazid is recommended because it has been shown to be effective in preventing the progression of latent infection to clinical disease in a high proportion of individuals. It is routinely preferred for infected persons under thirty five years of age and not those above because of the increased risk of isoniazid-associated hepatitis among older persons. Isoniazid therapy is not to be used where there is a history of previous severe adverse reaction to drug or when there is an acute liver disease. For pregnant women, it may be wise to postpone preventive treatment until after delivery and isoniazid should be given with added caution to persons who use alcohol regularly . Patients infected by organisms resistant to isoniazid or rifampin are vaccinated with Bacilli calmet guarine (BCG) and this include children with negative tuberculin tests who cannot be placed on isoniazid preventive therapy and persons with immunodeficiency diseases such as symptomatic HIV infection.

Selective tuberculin testing is done on groups at a high risk of tuberculosis or HIV infection as a case finding measure e.g. immigrants from areas where tuberculosis is prevalent , groups at high risk for HIV infection such as intravenous drug users and homosexuals. In population groups where disease still occurs, systematic tuberculin testing surveys may be used to monitor trends in the incidence of infection. X-ray examination is especially preferred whenever persistent chest symptoms are noted and where bacterial tests are negative.

Persons infected with HIV are normally skin tested at the time of identification of the HIV infection and if found positive, prophylatic treatment is given. Testing for HIV infection is also considered necessary in all persons with evidence of TB infection. If a TB patient is suspected of being HIV positive then he/she is not supposed to be tested without pre-and post-test counseling. This is as per the guidelines of the national AIDS control program. This task should be delegated to health staff who have received training on counseling.

Patients with sputum-positive pulmonary TB need to be placed in private room with special ventilation. Patients are taught how to cover the mouth and nose when coughing or sneezing. Persons entering the room should wear masks if the patient is coughing and does not cover his mouth.

Patients with primary TB and those whose sputum are bacteriologically negative, who do not cough and who are known to be on adequate chemotherapeutic need not be isolated. The need to adhere to the prescribed chemotherapeutic regimen must be re-emphasized repeatedly to all patients. Handwashing and good housekeeping practices should be maintained according to routine policy and there are no special precautions when handling formites. Decontamination of air may be achieved by ventilation and sunlight (ultraviolet light).

Patients with TB are given prompt treatment with an appropriate combination of antimicrobial drugs which include isoniazid (INH) combined with rifampin (RIF) with or without pyrazinamide (PZA) and ethambutol (EMB). If sputum fails to become negative after three to four months of regular therapy or reverts to negative after a series of positives or if clinical response is poor, then examination for drug taking compliance and development of drug resistance is investigated because treatment failure is commonly as a result of irregularity in taking drugs . If drug resistance is observed, at least two drugs to which the organism is susceptible is usually included in the regimen (combination). If both INH and RIF cannot be included in the regimen, the minimum duration of the therapy is eighteen months. Children are treated with the same regimens as adults although many experts advise prolonging therapy for three to six months for those with life-threatening extra-pulmonary TB such as meningitis. EMB is not used until colour vision has been checked and PZA has not been approved by health authorities for use in children.

There has been an increase of people suffering from TB in Kenya in recent years and this is suspected to be due to an increase of HIV infection in adult population which is already infected with mycobacterium tuberculosis. It is also very likely that HIV-seropositive persons who have never been infected with mycobacteria tuberculosis are at a risk of developing TB when they become infected for the first time (Kibuga, 1995). For instance, during 1994, TB case finding increased by to 22,601 cases as compared to 20,131 in 1993 which is a 10% increase. New smear-positive cases increased by 11% from 10,023 in 1993 to 11,159 in 1994. This was a much lower increase than the previous year when the increase was 37% for all cases and 30% for smear positive cases. The huge increase during 1993 might not have been wholly due to HIV but possibly also as a result of increased program performance and the economic changes that year which led to rising cost of medication in private hospitals with more patients seeking treatment in government hospitals (Kibuga, 1995). In order to ascertain the added TB disease burden due to HIV infection, a sentinel surveillance of HIV in TB patients had commenced in 1993 and continued during 1994. Nineteen districts were included in the study, supported by both the WHO and the Royal Netherlands TB association. The results revealed that in most districts HIV co-infection in TB patients was as high as 20-35% (WHO annual report, 1994).

Another factor which has led to an increase in TB cases is the fact that the treatment of the disease is poor and this results in treatment failure, patients' death and acquired drug resistance. Late diagnosis is an aspect which has contributed to ineffective treatment because of patient's delay or doctor's delay (WHO annual report, 1994).

1.2 STATISTICAL MODELS FOR BIO DATA ANALYSIS

Statistical models are used in the analysis of relationships for data obtained from a wide variety of sources and are presented in numerous ways. They were developed with the goal of finding solutions to certain mathematical and statistical problems encountered in daily life. For example, Hamer (1906) tried to use simple but precise mathematical statements to find solutions of two specific quantitative problems; the regular occurrence of measles epidemics and the relationship between the number of mosquitoes and the incidence of malaria (Moshkovkii, 1950) and one of the models developed by Hamer (1906) is the discrete time series model which is based on the "mass action principle" which states that the net rate of the spread of an infection is assumed to be proportional to the product of the density of susceptible people and the density of the infectious individuals.

The importance of s[illegible]dels is in its application to data-oriented problems and it is summed up by the science-fiction writer Poul Anderson: "I have yet to see any problem, however complicated which when you look at it in the right way, did not become still more complicated" (Anderson and Robert , 1992). There are three main models used in the analysis namely: regression models, time series models and categorical data models. In this study, time series models, extended loglinear models as well as extended logistic regression models are applied in assessing trends of communicable diseases. The last two models are blend of the regression models and categorical data models for dependent data.

1.2.1 Regression models

A regression model is one in which the expected value of one variable called the response variable is related to the actual values of other variables. Therefore, regression analysis is a statistical tool that utilizes the relationship between two or more quantitative variables so that one variable (response variable) can be predicted from the other or others (independent or explanatory variables). For example if one knows the relationship between advertising expenditures and sales, one can predict sales by regression analysis once the level of advertising expenditures has been set.

Regression models include Simple regression models, Multiple regression models, Polynomial regression models and Non-linear regression models.

1.2.2 Time series models

Time series models deal with data which consists of time-ordered sequences of measurements on some phenomenon of interest and are applicable in various fields which include the field of medicine where an epidemiologist might be interested in the number of infectious hepatitis cases over time, diastolic and systolic blood pressure measurements traced over time for various individuals could be useful for evaluating drugs used in treating hypertension and effect of a particular infection traced over time.

In economics, time series models can be used when dealing with stock market quotations, monthly daily unemployment figures and in social sciences where demographers are interested in following population series such as birth-rates or school enrollments. In geophysical sciences, time series models are used in earthquake prediction or for indicating whether a given seamic record was produced by an earthquake or a nuclear explosion (Shumway, 1988). Time series models also work well in engineering and sciences where events tend to be generated by periodic phenomenon.

There are many types of time series models and these include the Autoregressive moving average (ARMA) and Autoregressive intergrated moving average (ARIMA) models, transfer function models, multivariate time series models, intervention models and nonlinear time series models.

1.2.3 Categorical data models

Categorical analysis deals with the analysis of variables for which the measurement scale consist of a set of categories. For example, political philosophy may be measured as liberal, moderate or conservation. Categorical models are applicable in the field of behavioral sciences, public health, ecology, education, marketing, engineering, sciences and industrial quality control. The main models that are used in categorical analysis are the logistic (logit) models and loglinear models. These models closely resemble regression models for continuos response variables but they assume binomial, multinomial or poisson response distributions rather than the normal distribution.

1.3 LITERATURE REVIEW

There have been important developments in the study of infectious diseases due to their social and economic impact on individuals, community and the fact that they do not respect international borders. As a result of the studies of these diseases, important information on medical advances is readily available. This enables health personnel to recognize specific diseases, and thus facilitate how to manage patients so that they do become sources of new cases and they also provide guidance for their treatment to preserve lives. In addition, regulations and legal requirements have been set up for the control and management of infectious diseases and for developing programs for health education of the public.

The study of infectious diseases has progressed a long way and the growth of knowledge in the field of immunology and cellular and molecular Biology has provided many tools which help in the control and management of these diseases. New effective drugs are being developed based on the latest research findings.

The frequency and magnitude of epidemics of some communicable diseases in Kenya has been on the increase as a result of changing social patterns, rapid urbanization, poor working conditions, lack of public health education and the limited resources. This has adversely affected the ability to control these diseases. The other major problem has been attributed to the fact that these diseases have become resistant to antibiotics (KMAR, 1994). It is thus essential to formulate strategies to minimize the emergence and spread of drug resistance.

However, there has been a decline of cases of people suffering from smallpox, polio and tetanus due to the fact that public health education has been intensified, there are better standards of living and effective medical interventions have been carried out. For instance, smallpox, was eradicated as a consequence of the spread of inoculation, vaccination and isolation procedures (Benenson, 1990).

Immunization programs in Kenya have been effectively performed because of the fact that health education has been intensified in all communities. There is a continuous surveillance of these diseases by World health organization (WHO) to detect changes by which the organism become resistant to previous effective vaccine (WHO bulletin, 1998) coupled with the fact that Kenya expanded programs on immunization (KEPI) centres have been set up in rural areas and upto-date immunization practices based on the latest research outcomes and developments are always implemented. However, in some situations immunization programs have not succeeded due to cost, imperfect distribution of vaccines and less acceptance in some communities.

The use of statistical models dates back to the work of Farr (1840) who effectively fitted a normal curve to smoothed quarterly data deaths from smallpox in Wales and England. This approach was developed further by Browlee (1906) who published a paper entitled statistical studies in immunity in which he fitted personian frequency distribution curves to a large series of epidemics. This concept was further developed into a discrete time series model (Hamer, 1906). Later, Ross translated this model to a continuous time series model based on dynamics of malaria (Ross 1911, 1915, 1916, 1917).

The most widely used time series models developed by Wold (1954) who combined the Autoregressive (AR) model due to Yule (1927) and the moving average process (MA) due to Slutzky (1937) is of the form

$$X_t + \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} = \theta_1 e_{t-1} + \ldots + \theta_q e_{t-q} + e_t.$$

These models have been applied in various fields especially for forecasting purposes and these include the field of medicine, economics, geographical sciences and engineering. These models provide extremely accurate forecasts due to serial dependency existing between the data values and they also provide a good approach to model building. For example, to fit this model to a time series data involves three major steps namely: Identification of the order of the ARMA (p,q) model, estimation of the model parameter, diagnostic checking of the fitted model and finally the model is used for forecasting.

Before fitting an ARMA (p,q) model, the stationarity of a time series data need to be checked. If it is non-stationary then it is differenced d times to make it stationary because reliable information can be drawn from a stable process. Identification of the ARMA (p,q) is based on both subjective and objective techniques. The popular subjective technique is referred to as the Box and Jenkins (1976) approach which is based on plotting of the sample autocorrelation function (SACF) and sample partial autocorrelation function (SPACF) then these plots are compared with the corresponding theoretical behaviour. The objective techniques are usually based on information theory and these include Akaike (1974), Schwartz (1978) developed the Bayesian Information Criteria (BIC) and the final prediction error Criteria (F.P.E).

In estimation of the model parameters, we have the maximum likelihood estimates for which the various algorithms can be used to compute the parameter estimates (Ansley, 1979). The other commonly used estimation criteria is the Yule walker technique. Others are the conditional least squares developed by Klimko and Nelson, (1978) and the optimal estimating functions by Godambe (1985).

Diagnostic tests have also been suggested to examine the adequacy of the model identified. These tests are based on comparing the response of the fitted model with that of actual process for a number of tests after which the model is used for forecasting. During forecasting the criteria often used is to minimize the mean square error (MSE) or the use of the conditional expectation of X_{n+h} given the observation X_1 , X_2 ,...,X_n.

Zeger (1988) developed a parameter driven extension of log-linear model which is a blend of a regression model and a categorical data model. This model is primarily meant for assessing trends of a dependent data and he applied it to U.S. Polio incidence to show whether this data provided evidence of a long term increase or decrease in the rate of polio infection. In this model, the observation Y_t is assumed to be affected by the unobservable noise process e_t and thus the autocorrelation and overdispersion is introduced through a latent process where the canonical parameter $\theta_t = \log \mu_t$, that is

$$\phi_t = \phi(e_t, y_{t-1}, \ldots, y_1)$$

Hence this model is of the form

$$E(y_t / e_t) = \exp(X_t\beta)\, e_t$$

where y_t is an observation at time t which assumes a poisson distribution, X_t is a (1xp) vector of covariates. The poisson distribution plays a vital role in analyzing and describing count data and being a member of the exponential family implies that the associated inferential procedures are optimal.

Estimation of the parameters is done using the quasi-likelihood method which allows for a variety of variance-mean relations since in practice var $(y_t) > \mu_t$. This leads to quasi-likelihood estimators which are consistent, asymptotically gaussian and are optimal in the extended Gauss-Markov sense. Also, this approach is robust in the sense that consistent inferences about β can be made given only that $E(y_t) = \mu_t$.

Zeger (1988) also proposed a log-linear model of the form $E(y_t) = \mu_t = \exp(X_t\beta)$ to be used for assessing trends of independent data. Any likelihood method apart from the quasi-likelihood can be used to estimate the parameters.

Cox and Snell (1970) proposed a markov chain that is an extension of the logistic regression model given as

$$\text{Logit}\ (\mu_t) = X_t\beta + \phi_1 y_{t-1} + \ldots + \phi_q y_{t-q}$$

where μ_t is the conditional mean, β and $\phi = (\phi_1, \ldots, \phi_q)$ are the parameters to be estimated. He used this model to assess the trend of a dichotomous (binary data) time series data. In this model, $\exp(\phi_i)$ is the odds of a positive response at time t given $Y_{t-i} = 1$ relative to the odds that $Y_{t-i} = 0$ with all other components being held constant.

When $\phi_i = 0$ ($i = 1,\ldots,q$), it reduces to the logistic regression model.

1.4 OBJECTIVES

A major goal of this study is to assess trends of these five communicable diseases using regression models and the time-dependent regression models. The other goal was to obtain forecasts of future values by fitting time series models. The results obtained are expected to contribute to the management and control of these communicable diseases.

CHAPTER TWO

TIME-PLOTS OF MORBIDITY COUNTS

2.1 METHODS

Data for the analysis were collected from the Ministry of health headquarters and they are in the form of time series counts. These data are based on reported incidences of patients infected, in-patients who survived and those who died from the five communicable diseases in Kenyan government hospitals. These include 1989-1997 morbidity data, 1989-1997 in-patient data based on alive and dead cases and 1952-1997 morbidity data. Appropriate statistical models were fitted to the data.

Simple regression models were fitted to both 1989-1997 morbidity and mortality data while Log-linear models and Logistic regression models were fitted to 1989-1997 morbidity data and 1989-1997 mortality data respectively. These models were fitted basically for the assessment of trends while time series models were fitted to 1989-1997 morbidity data for forecasting purposes.

2.2 TIME PLOTS

Time-plots on 1989-1997 monthly morbidity data and 1952-1997 annual morbidity data for each kind of disease were drawn and every feature is clearly illustrated in the subsequent figures and discussed thereafter.

2.2.1 Tuberculosis

TB incidences over the period 1952-1997 are illustrated below in the figure1. It is clearly shown that TB was not a serious problem in the period 1952-1980. This could be due to the establishment of [illegible] sensitivity testing for TB and the fact that BCG was administered to all cases of TB infection (KMAR, 1957, 1982). Alternatively, the number of TB was low due to underreporting of cases, lack of proper medical records, low population over this period and most people relied on traditional herbs for treatment of TB (Njogu and Tukei, 1982).

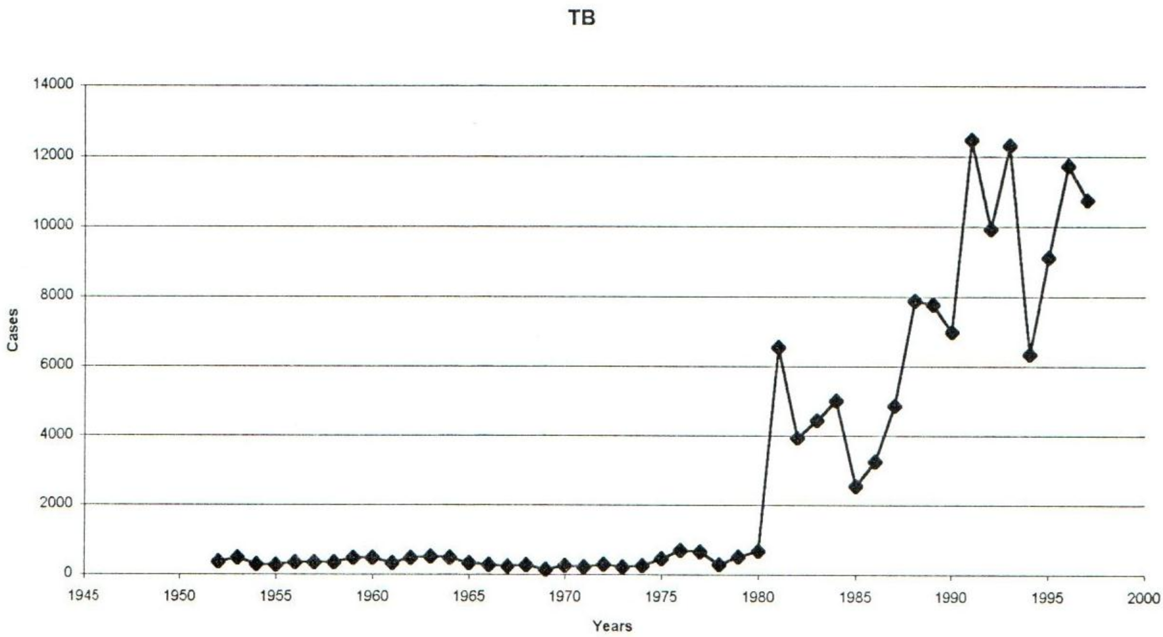

Figure 1: 1952-1997 annual TB morbidity time-plot.

It can also be seen in fig 2, that from 1980, the number of TB cases rose sharply and it was highest in 1993, then it slightly decreased in 1994 but was high again in 1995 and 1996. In 1997, it declined slightly. The increase of TB incidences and deaths in Kenya, from 1980 is attributed increase in population, poor living conditions especially in slum areas and unavailability of drugs in government hospitals (Njogu and Tukei,1982). The other reason was lack of public health education on the disease and as a result, most people did not know the need of observing health practices such as covering the mouth when one is coughing and sneezing. The disease can be transmitted from person to person through exposure to bacillus in airborne-nuclei produced by the person with pulmonary or laryngeal TB during respiratory efforts such as coughing, sneezing and singing.

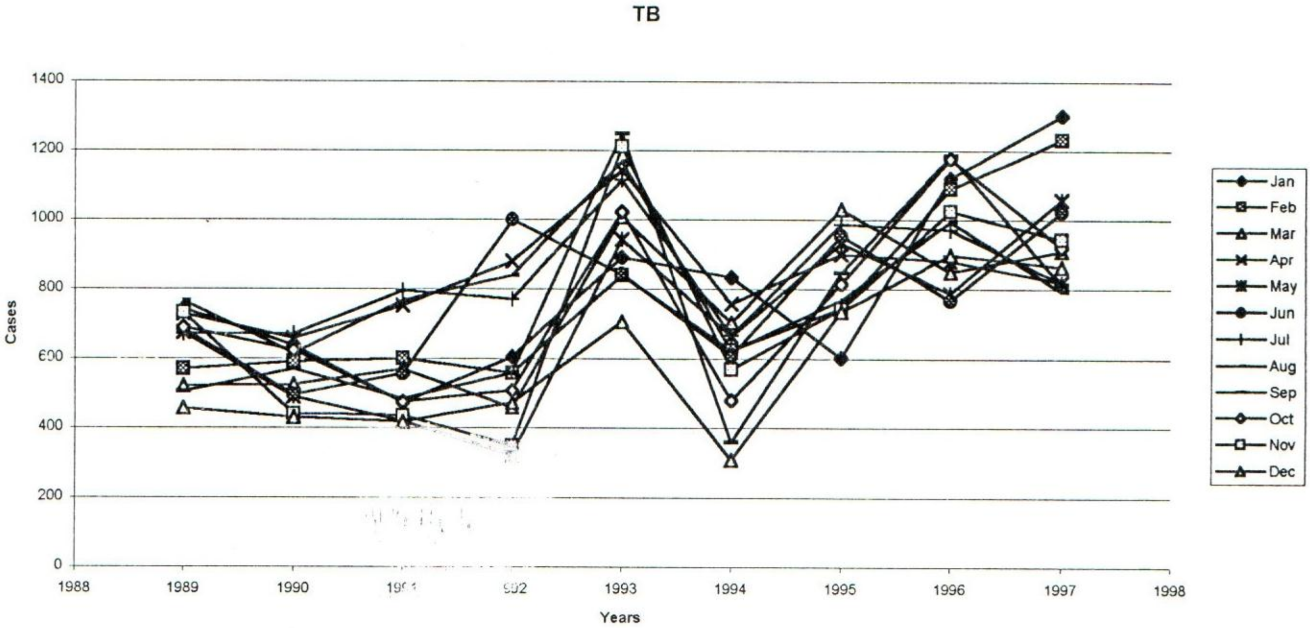

Figure 2: 1989-1997 monthly TB morbidity time- plot.

It can also be seen clearly from a time-plot of TB cases over the period 1989-1997 in figure 2 , that there were large number of TB incidences in 1993. The increase might not have been wholly due to the association of TB with HIV infection and AIDS, but also largely as a result of inflation caused by the devaluation of the shilling in 1993. Inflation made the cost of medication especially in private hospitals high and as a result, most patients went for treatment in government hospitals (Kibuga,1995).

Another factor that could have led to increase of TB incidences is the fact that the treatment compliance might have been poor and this could result to treatment failure, patients' death and the disease became resistant to drugs. In principle, it is expected that patients are prescribed the right drugs in the right dosage and take them regularly as prescribed, for the correct period of time, then almost all TB patients will be cured (WHO annual report, 1994).

Also, the introduction of the treatment of infectious cases of TB in all districts with short-course chemoth[illegible] previously only reserved for the treatment of nomads. From 1995, TB incidenc[illegible] again largely due to the association with HIV infection and AIDS (Kibuga,1995). The slight decrease of TB cases in 1997 was due to NLTP policies that were implemented and aimed at reducing TB incidences and the transmission of HIV infection by all available means (Kibuga,1995).

2.2.2 Poliomyelitis

Polio cases over the period 1952-1997 are depicted below in fig 3. It is clearly shown that there were few incidences of poliomyelitis, about 250 cases reported annually over the period 1952-1970. From 1959 to 1973, the number of incidences decreased slightly possibly as a result of the introduction of live attenuated oral polio vaccine at a free-cost to all children in 1959 and 1960. Sabin vaccine was later introduced in 1962 (KMAR,1958, 1962). In general , number of polio incidences was at a low level due to low population, underreporting and lack of proper records.

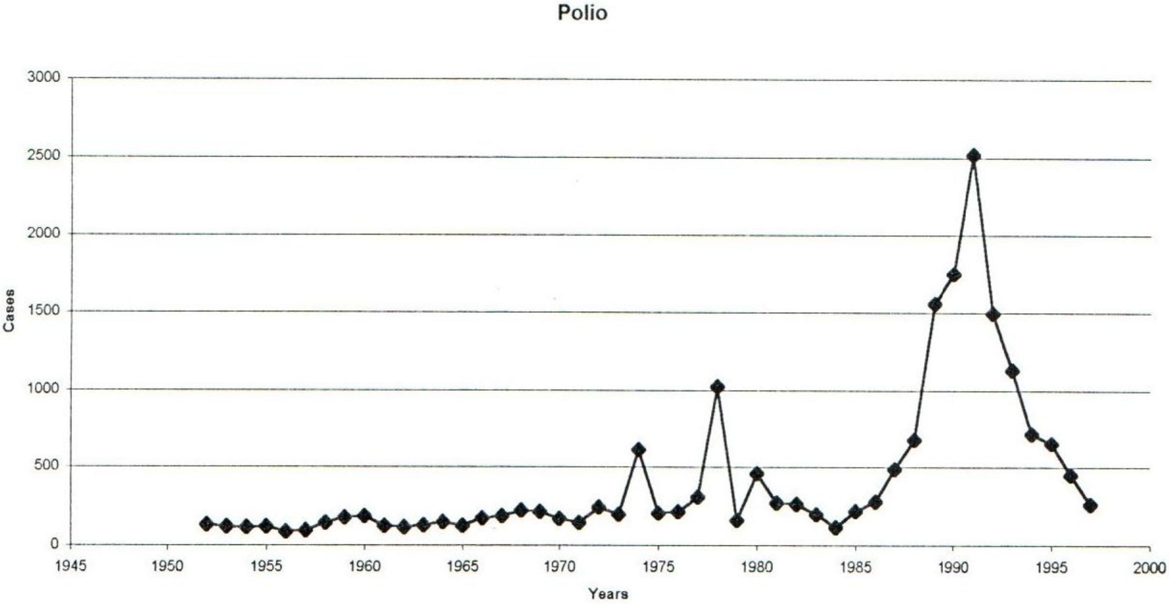

Figure 3: 1952-1997 annual polio morbidity time-plot.

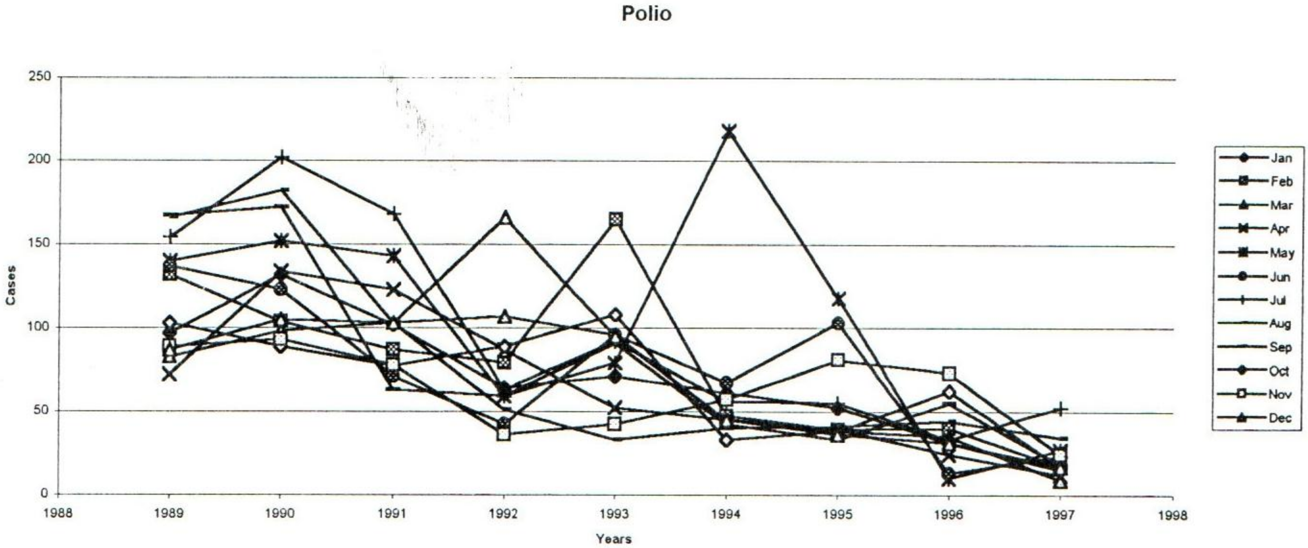

Figure 4: 1989-1997 monthly polio morbidity time-plot.

Cases of Poliomyelitis rose in 1974 and there was a sharp increase in 1978 and there was an outbreak of poliomyelitis especially among school-going children who had not been immunized before (Njogu and Tukei,1982). The outbreak was contained through the introduction of mass immunization. This approach never worked effectively since there was no public health education on immunization and as a result, there was less acceptance of immunization program in some communities (KMAR,1995). Thus the intervention gains were short lived.

This crucial aspect made the number of polio incidences to increase tremendously from 1985 to 1991. In figure 4 above, it is clearly illustrated that the number decreased from 1989 to 1997 due to [illegible] immunization that has been implemented effectively and efficiently in the country through KEPI.

2.2.3 Tetanus

The figures depicting trends in tetanus cases over the period 1952-1997 and the period 1989-1997 are given in figures 5 and 6 below. It is clear that the number of the people infected with tetanus was low, slightly below 400 tetanus cases in the period 1952-1965. Since 1966, tetanus incidences began to increase drastically with a sharp increase in 1989 and it was highest in 1990. There was a steady decrease in the number of tetanus incidences in 1990-1997.

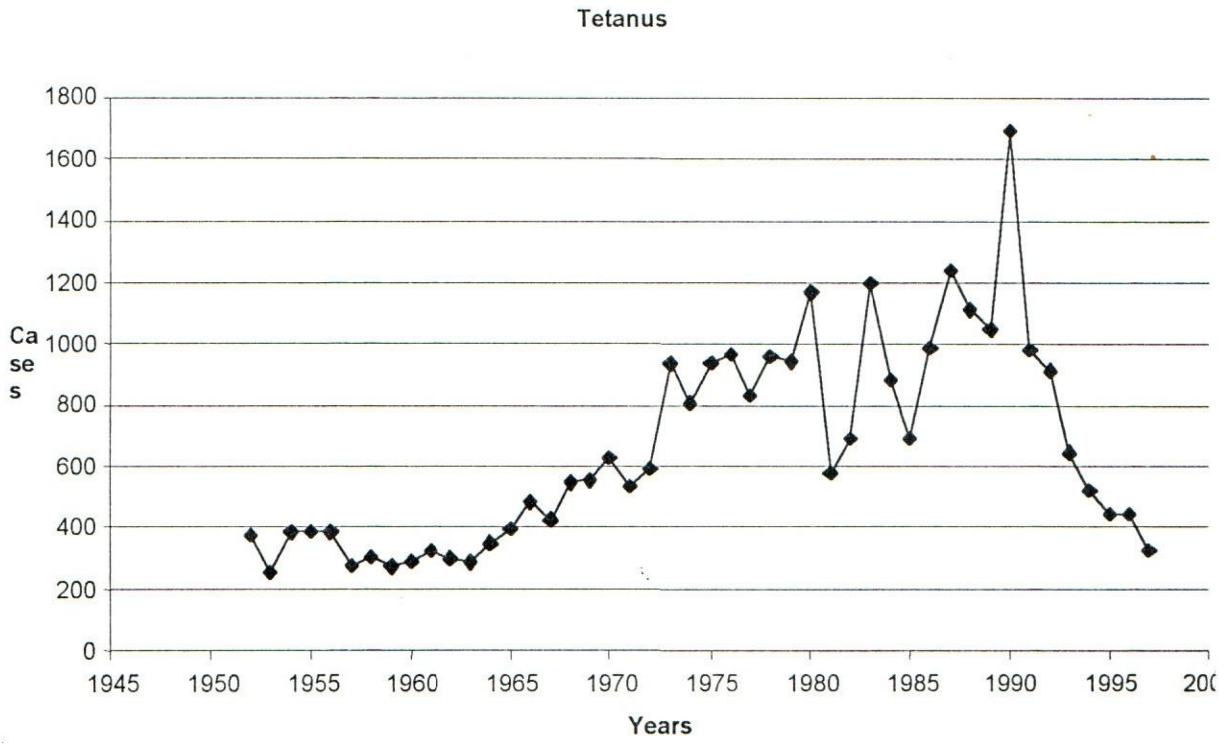

Figure 5: 1952-1997 annual tetanus morbidity time-plot.

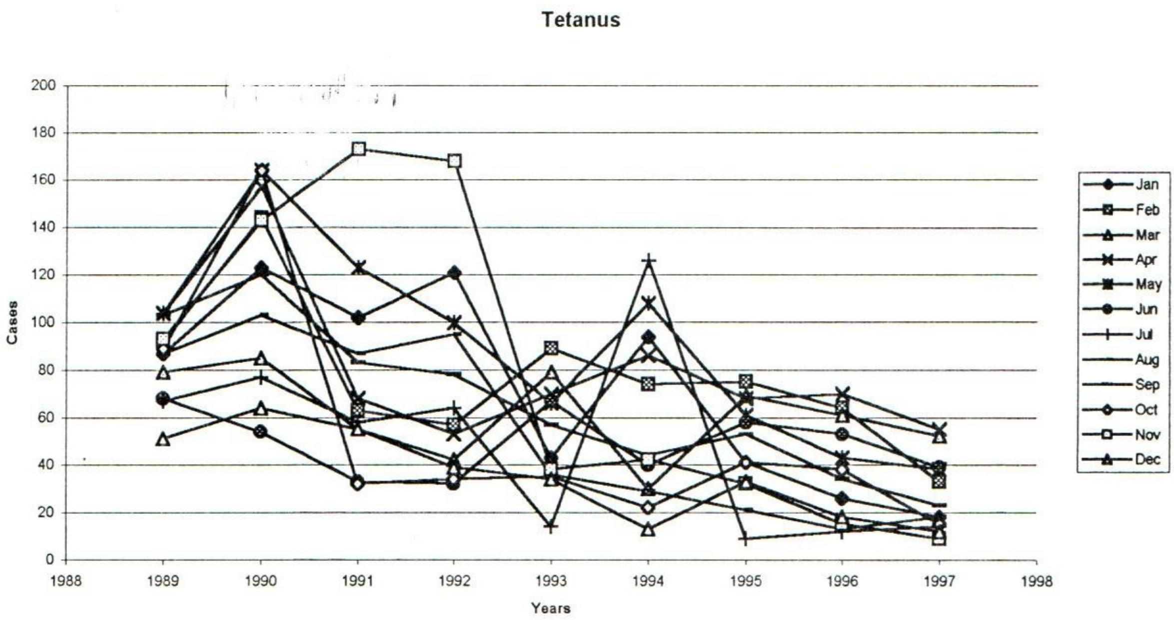

Figure 6: 1989-1997 monthly tetanus morbidity time-plot.

There was low population in the period 1952-1965 and BCG was widely used. Tetanus toxoid and DTP vaccines were introduced in 1960 and 1962 respectively (KMAR, 1958-1962). Increase in population, lack of public health education on kinds of injuries liable to be infected by tetanus bacillus, imperfect immunization practices and lack of government commitment in eradicating this disease led to this increase of tetanus incidences in the country (KMAR, 1994). In addition, the sharp increase in the period 1989-1990, was as a result of the fact that some communities accepted immunization practices as a way of eliminating the disease, imperfect distribution of the vaccines and lack of public awareness about the fatality of the disease (KMAR,1994).

From 1990, the steady decrease was due to the widespread immunization campaign that has been going on in the country. Further to this, Kenyan ministry of health has recommended that every individual should seek for an upto-date immunization against tetanus while pregnant women ought to go for vaccine-induced maternity which is important in preventing neonatal tetanus (KMAR,1994).

Immunization became a success as a result of intensification of public health education on the kinds of injuries liable to be infected by tetanus bacillus since some cases were reported as a result of injuries considered too trivial for medical consultations (KMAR,1995).

2.2.4 Measles

Measles can be considered as one of the important childhood diseases in Kenya. This was not a serious problem in Kenya until 1955 when an outbreak occurred. Since then , the incidences of measles have been high. Slightly over 10,000 cases especially among young adults and malnourished children were reported in government hospitals to be suffering from measles in the period 1955-1978. This was relatively a high figure as per the population of the people in this period . This scenario is quite evident in figures on measles cases over the period 1952-1997 in the figure 7 below.

However attenuatted live measles vaccines has been used since 1965, but the impact in rural areas in Kenya has been limited due to several reasons. These include the high cost of vaccines, the severe reactions experienced with earlier generation, the logistics of keeping the vaccine refrigerated all the way down the factory to the child receiving the vaccine and the scarcity of trained staff at health delivery points.

As a result of these limitations, low immunization coverage rates and a considerable failure rate in protection of children against measles were experienced (Njogu and Tukei,1982). For instance, since 1978, the number of measles cases rose steadily due to ineffective immunization as a result there was some outbreaks in schools, colleges and slum areas with the major outbreak occurring in 1989 (KMAR,1995).

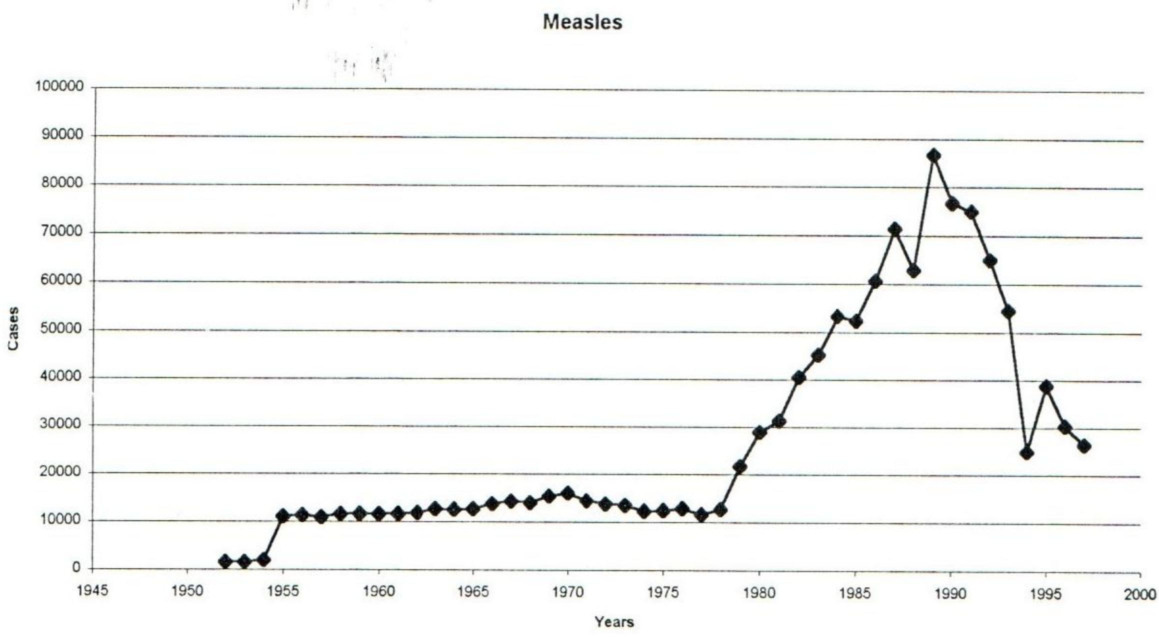

Figure 7:1952-1997 annual measles morbidity time-plot.

The 1989-1997 morbidity of measles is given in figure 8 below and is similar to that of poliomyelitis and tetanus. The number of measles incidences was reduced to around 25,000 cases in 1994. It rose slightly to 38,000 in 1995 as evidenced by outbreaks (KMAR,1995).

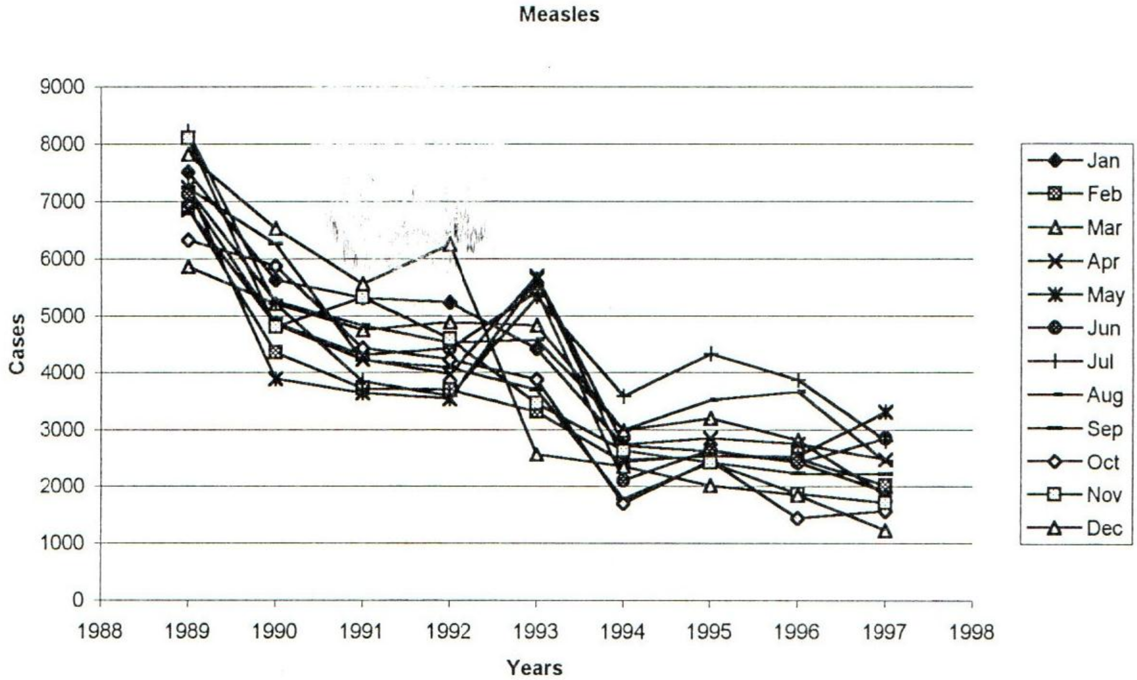

Figure 8: 1989-1997 monthly measles morbidity time-plot.

Incorporation of poliomyelitis and measles vaccines into regular childhood program in accordance with EPI under the auspices of the WHO led to the reduction of measles cases to low levels in 1995-1997. This improvement in management and the control of measles was also due to the intensification of public health education and the fact that the disease has been under surveillance by WHO as a result up-to-date immunization practices based on the latest research outcomes and developments have been implemented. For instance these vaccines of further attenuated live measles virus have appeared on market which have been shown to have more stable temperature. It was also claimed that attenuvax vaccine could be kept at ambient temperature for four weeks and for 24 hours when reconstituted (Njogu and Tukei,1982). Figure 8 shows a general decreasing trend over the years.

2.2.5 Infectious Hepatitis

Infectious hepatitis is predominantly a childhood disease. It occurs in children but many adults have also been infected. It constitutes about one third of acute hepatitis cases, with large fluctuations occurring from year to year. Hepatitis was not a problem until 1960, when some few hepatitis incidences were reported in some government hospitals in the country as shown in figure 9.

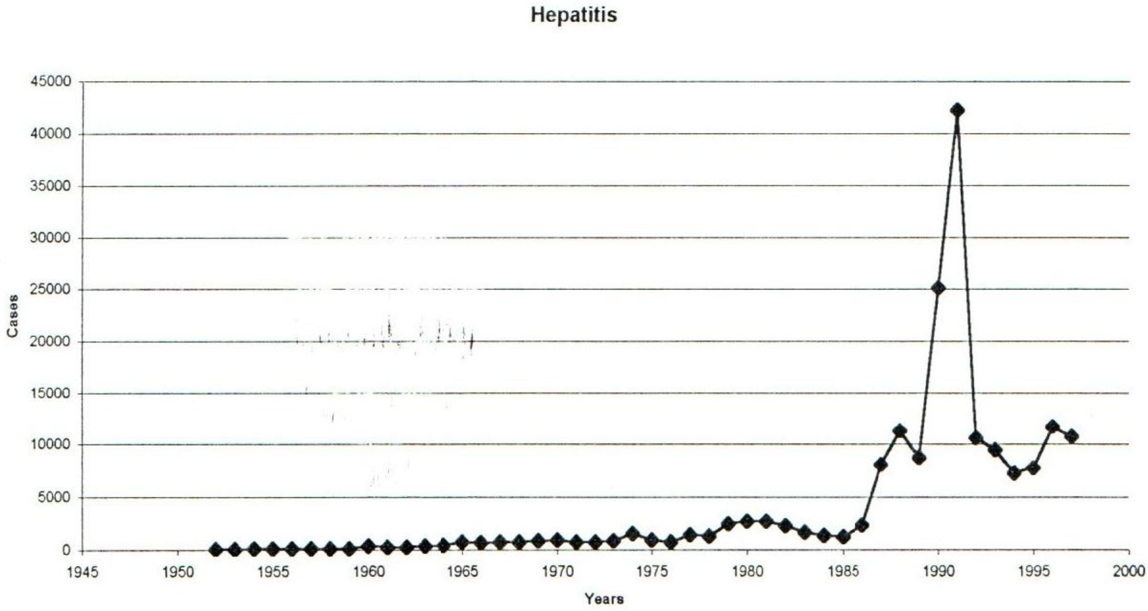

Figure 9: 1952-1997 annual hepatitis morbidity time-plot.

In 1964, the number of cases began to increase and from 1986, there was a sharp increase in hepatitis incidences and was highest in 1991 as also shown in figure 10.

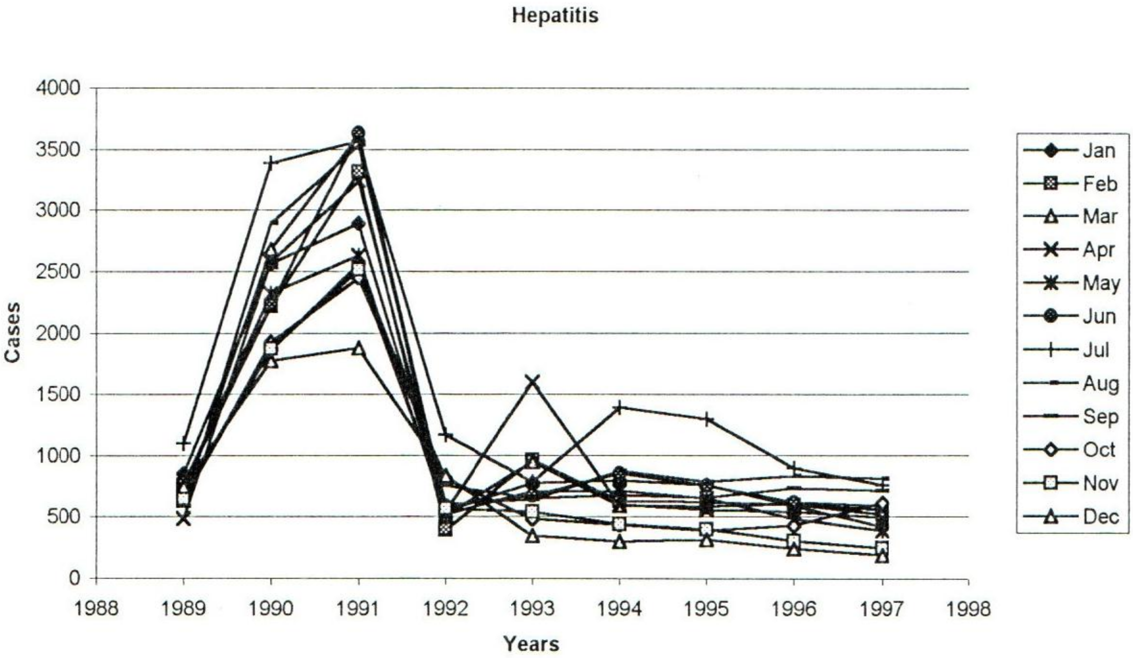

Figure 10: 1989-1997 monthly hepatitis morbidity time-plot.

The major reason for this increase was because there was an increase of intravenous drug users, lack of public health education, increased number of slum-dwellers, homosexuals and prostitutes in the period 1989-1991 (KMAR, 1995).

There was also a sharp increase in 1991 because public health education was not intensified on control methods which included good sanitation and personal hygiene with special emphasis on careful hand-washing and sanitary disposal of feces. Consequently, incidences of hepatitis have decreased due to the availability of drugs and trained health personnel in government hospitals. In addition, hepatitis A live vaccine is currently under use.

CHAPTER THREE

TRENDS BASED ON REGRESSION MODELS AND TIME-DEPENDENT REGRESSION MODELS

3.1 REGRESSION MODELS

Regression analysis is one of the widely used statistical techniques for investigating and describing the relationships between dependent variables and independent or explanatory variables and in most cases, the purpose of the analysis is for prediction purposes. In this analysis essentially the data was transformed using the logarithm transformation technique plus one. A simple regression model was fitted to both 1989-1997 morbidity and mortality data for each disease. The simple regression model used is of the form

$$Y_t = \beta_0 + \beta Y_{t-1} + e_t$$

where Y_t is the count at time t, β_0 is the intercept, β is the regression coefficient, Y_{t-1} is the previous count at time t-1 and e_t is the innovation at time t that is assumed to be normally distributed with zero mean and a constant variance σ^2 .

The estimated regression coefficient $\hat{\beta}$ corresponds to the trend of a particular communicable disease. In essence, the estimate of the parameter β provides evidence of a long term increase or decrease in the rate of infection. Y_t is the number of people infected at time t and x_t is the previous figure at time $t-_{1}$.

3.1.1 Results

Tuberculosis

The results for both TB morbidity and mortality cases are given in the following models respectively

$$Y_t = \text{[illegible]} + 0.00885\,X_t \;,\; P=0.0001$$

$$Y_t = 3.162 + 0.013X_t, \quad P=0.0001$$

These results provide strong evidence of significantly increasing TB morbidity data mortality incidences between 1989 to 1997.

Poliomyelitis

The simple regression models are as follows:

$$Y_t = 3.305 + 0.0106X_t, \quad P=0.0001$$

$$Y_t = 0.307 + 0.027X_t, \quad P=0.6704$$

These results also provide evidence of both increasing morbidity and mortality trends in the rate of polio infection. Though the mortality trend is insignificant.

Tetanus

The models for both tetanus morbidity and mortality cases are given as follows:

$$Y_t = 10.713 + 0.8264X_t, \qquad P=0.0001$$

$$Y_t = 0.401 + 0.146X_t, \qquad P=0.0001$$

These results provide evidence of increasing tetanus incidences from 1989 to 1997.

Measles

Significantly increasing morbidity and mortality trends are obtained when simple regression models were fitted to measles incidences. The results are given in the following models

$$Y_t = 7.286 + 0.0002X_t, \qquad P = 0.0001$$

$$Y_t = 1.193 + 0.06X_t, \qquad p=0.0001$$

Hepatitis

The regression models for both hepatitis morbidity and mortality data were obtained as

$$Y_t = 6.005 + 0.0006X_t, \qquad (P=0.0001)$$

$$Y_t = 0.187 + 0.0196X_t, \qquad (P=0.775)$$

These results show slight significant increases in morbidity and mortality trends from 1989 to 1997.

The results obtained when regression models were fitted do not provide adequate inferences since some of the results do not concur with the trends that are exhibited in the time plots of these diseases. Patterns exhibited by the time series data are generally far more complex to be accommodated by regression models. Also, the observations in a time series are unlikely to be independent of one another and the distribution of the innovation process of the time series of counts may not necessarily be gaussian. We thus fit time-dependent regression models for evaluating the trends in the various diseases.

3.2 TIME-DEPENDENT REGRESSION MODELS

Generalized linear models include linear regression models, analysis of variance models, polynomial regression models, response surface models, non-linear regression models and some commonly used models for survival data. Log-linear models and logistic regression models are some of the special cases of generalized linear models. All these models share a number of properties such as linearity, that can be exploited to good effect and there is a common method for computing parameter estimates.

These models have distributions that are useful for the analysis of data which do not necessarily have normal error distributions and they also permit the study of patterns of systematic variation in much the same way as ordinary linear models are used to study the joint effects of treatments and covariates.

A generalized linear model is different from a general linear model in two different ways: First, in a general linear model, it is assumed that the distribution of the random components are independent and normally distributed with constant variance (σ^2) and mean (μ). In the generalized linear model, the distribution of the random components is from an exponential family such as binomial, negative binomial, poisson, gamma and inverse gaussian. Second, the identity link in a general linear model is replaced by a monotone differentiable function η =g(μ).

In general, a generalized linear model has the following salient features: The random component y which has an independent distribution from the exponential family; the systematic component constituting covariates $X_1, X_2, ..., X_P$ that produce a linear predictor $\eta_i = \sum_{j=1}^{p} \beta_j X_j$ a[illegible] that relates the mean (μ) of the datum (y) and the linear predictor η. The link function (g) is thus given as $\eta = g(\mu)$ or $\mu = g^{-1}(\eta)$.

In addition, each class of generalized linear models corresponds to a member of the exponential family of distributions with a log-likelihood function of the form

$$L(\theta,\phi,y) = \sum_i \{(y_i\theta_i - b(\theta_i))/[a(\phi) + c(y_i,\phi)]\} \qquad (3.2.1)$$

where the summation is over all the observations. Also, E(y)=$b'(\theta) = \mu$, var(y)=$b''(\theta)a(\phi)$, θ_i is the canonical parameter, μ_i =E(y_i)= $b'(\theta_i)$ is the mean of the datum y, $\eta_i = \sum_{j=1}^{p} \beta_j X_{ij}$ is the linear predictor and $\eta_i = g(\mu_i)$ is the link function.

Since the distribution belongs to the exponential family, it means that unique maximum likelihood estimates exist and thus we can obtain the estimates of β_j, j=1,2,…,p by solving

$$\partial L / \partial \beta_j = 0, \text{ j=1,2,…,p.}$$

The likelihood L is function of θ where θ is the function of μ, μ is the function of η and η is a function of β. This implies that

$$\partial L / \partial \beta_j = (\partial L / \partial \theta_i) \times (\partial \theta_i / \partial \mu_i) \times (\partial \mu_i / \partial \eta_i) \times (\partial \eta_i / \partial \beta_j) \qquad (3.2.2)$$

On omitting the summation, we have

$$\partial L / \partial \theta_i = (y_i - b'(\theta_i))/a(\phi) = (y_i - \mu_i)/a(\phi),\ \mu_i = b'(\theta_i)$$

$$\partial \mu_i / \partial \theta_i = b''(\theta_i) \Rightarrow \partial \theta_i / \partial \mu_i = 1/b''(\theta_i),\ \partial \eta_i / \partial \beta_j = X_{ij}$$

So substituting equation (3.2.2), we have

$$\partial L / \partial \beta_j = \sum_i \{ [(y_i - \mu_i)/(a(\phi)b''(\theta_i))] \times (\partial \mu_i / \partial \eta_i) \times X_{ij} \}$$

Now since $\mathrm{var}(y_i) = b''(\theta)a(\phi) = V_i$, it follows that

$$\eta_{ij} = \partial L / \partial \beta_j = \sum_i \{ [(y_i - \mu_i)/v_i] \times (\partial \mu_i / \partial \eta_i) \times X_{ij} \}$$

This equation reduces to the estimating equation of the form

$$S(B)= \sum_{i=1}^{p} (\partial \mu_i / \partial \beta) v_i^{-1} (y_i - \mu_i(\beta)) \qquad (3.2.3)$$

Where v_i=var(y_i). Note that S(β) is the derivative of the logarithm of the likelihood function. The solution, $\hat{\beta}$, which is the maximum likelihood estimate, can be obtained by the iteratively reweighted least squares method (McCullagh and Nelder, 1983).

3.2.1 Log-linear models

Log-linear models (Poisson regression models) are applicable in problems where the response variable represents the number of events occurring in fixed period of time. Because of the discrete and non-negative nature of count data, a reasonable assumption is that the logarithm of the expected count is a linear function of the explanatory variables so that the log-linear model is of the form

$$\mathrm{Log}(E(y_i))= X_i'\beta$$

where β is a vector of parameters to be estimated. The random component (y) is assumed to have a poisson ditribution with parameter μ. The density function is given as

$$f(y; \mu) = e^{-\mu} \mu^y / y! = \exp(y \log \mu - \mu - \log y!)$$

Comparing with equation (3.2.1), we conclude that the canonical parameter $\theta = \log \mu$. Where the log is the link function relating the mean response to the vector of covariates X_i , $b(\theta) = \mu = e^{\theta}$, $\phi = 1$, $a(\phi) = 1$, $c(y, \phi) = \log(y!\ )$. Hence $E(y) = b'(\theta) = e^{\theta} = \mu$ and $var(y) = b''(\theta)a(\phi) = e^{\theta} = \mu$.

For independent data, E(y)=var(y)= μ, therefore the maximum likelihood method can be used to estimate β. In this study, the data are in the form of time series of counts. Therefore, the neighbouring observations are unlikely to be independent and as a result, the extended log-linear model was applied to account for dependence necessary to obtain valid inferences. The autocorrelation is introduced through the latent process since the canonical parameter θ is assumed to depend on an unobservable noise process, e_t. This model is also referred to as parameter-driven model (Zeger,1988) and is of the form

$$E(y_t / e_t) = \exp(X_t' \beta) e_t$$

where y_t is the response at time t, X_t is a $p \times 1$ vector of covariates and β is a $p \times 1$ vector of parameters to be estimated.

Since the latent process introduces both overdispersion and autocorrelation in y_t , it implies that $var(y_t) > \mu_t$. In this case, quasi-likelihood methods of estimating the parameters are appropriate because they allow a variety of variance-mean relations (Wedderburn,1976; McCullagh and Nelder,1983). The two most common assumptions are $var(y_t) = \mu_t \phi$ and $var(y_t) = \mu_t + \mu_t^2\ \sigma_t^{\ 2}$ where ϕ and σ^2 are known constants for members of the generalized linear model family, whereas in others they are additional parameters to be estimated.

Estimation of the regression coefficients

An estimating equation approach analogous to the quasi-likelihood is used to estimate the regression coefficients. In this approach, the first two moments of y_t are needed to be specified. From equation (3.2.3), the estimating equation is of the form

$$S(B) = \sum_{t=1} (\partial\mu_t / \partial\beta) v_t^{-1} (y_t - \mu_t) = 0$$

The quasi-likelihood estimator of β to be obtained is consistent, asymptotically normal and is optimal in the extended Gauss-Markov sense (McCullagh, 1983). This approach is also robust in that consistent inferences about β can be made given only that

$$E(y_t) = \mu_t .$$

In this study, the quasi-likelihood estimation approach was generalized to the time series case by letting $y = (y_1, \ldots, y_n)$, $X = (x_1, \ldots, x_n)$, $\mu = (\mu_1, \ldots, \mu_n)$ and V=var(y) so that

$$S(B) = (\partial\mu / \partial\beta) V^{-1} (y - \mu) = 0$$

Where V included off-diagonal terms which depend on nuisance parameters. The solution for β was obtained using SAS program under a generalized model (genmod) procedure.

3.2.2 Logistic regression models

The logistic regression models are members of a class of models known as generalized linear models introduced by Nelder and Wedderburn (1972). These models are fitted to some binary data to explore the relationships between binary response variables and one or more explanatory variables. Such data are said to be binary when each observation falls into one of the two categories such as alive or dead, positive or negative, defective or non-defective and success or failure. Binary data are often encountered in agriculture, biological and medical sciences. Underlying the analysis of binary data is the assumption that the observations are from a binomial distribution.

Suppose that we have n binomial observations of the form $x_i / n_i \ ,i = 1,2,...,n$ where $y_i = x_i / n_i$ and $\mathrm{E}(x_i) = n_i p_i$ and p_i is the success probability corresponding to the i^{th} observation. Let y=x/n and x~b(n,p) then

$$\mathrm{P(x)}=\binom{n}{x} p^x (1-p)^{n-x}$$

$$=\exp(x\log(p/1-p)) + n\log(1-p) + \log\binom{n}{x}$$

Therefore, the density function can be written as

$$f(x,\theta,\phi) = \exp\left[\ x\theta - n\log(1+e^\theta) + \log\binom{n}{x}\ \right]$$

$$= \exp\left[\ \{(x/n)\theta - \log(1+e^\theta)\}/(1/n) + \log\binom{n}{x}\ \right]$$

since y=x/n $\Rightarrow$ $f(y,\theta,\phi) = \exp\left[\ \{y\theta - \log(1+e^\theta)\}/(1/n) + \log\binom{n}{ny}\ \right]$

On comparing this equation with equation (4.11), it follows that $\theta = \log[p/(1-p)]$=logit(p) is the logit function which links the predictor to the mean of the datum y. Thus

$$\mathrm{Logit(p)=log[p/(1\text{-}p)]}=\sum_{j=1}^{n} \beta_j x_j \ .$$

The above logistic regression model is used to analyze independent binary data but for binary time series, Cox and Snell (1970) proposed a markov chain that is an extension of the logistic regression model of the form

$$\text{logit}\,(\mu_t) = X_t'\beta + \theta_1 y_{t-1} + \ldots + \theta_q y_{t-q}$$

This is the markov chain of order q . Here $\exp(\theta_i)$ is the odds of positive responses at time t given $y_{t-i} = 1$ relative to the odds when $y_{t-i} = 0$. When $\theta_i = 0$ $(i = 1,\ldots,q)$, it reduces to the logistic regression model.

Estimation of regression coefficients

Fitting a model to a set of data first entails estimating the unknown parameters in the model

$$\text{Logit}\,(p) = \log(p/1-p) = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p$$

The parameters in this model, $\beta_0, \beta_1, \ldots, \beta_p$ can no longer be estimated by the least squares method but by using the maximum likelihood method (Cox and Snell,1970 and Collet,1991). The least squares approach has several drawbacks: First, although the response variable in this case is constrained to be between zero and one , the least squares estimation may lead to parameter estimates which give fitted values of P outside the range (0-1). This would clearly be unsatisfactory and a further problem is the assumption of error terms with equal variance which is unlikely for this type of response (Collet,1991).

The likelihood function is given by

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

The problem now is to obtain $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$ which maximize L(β), or equivalently log L(β). The logarithm of the likelihood function is

$$\text{Log L}(\beta) = \sum_{i=1}^{n} \left\{ \log \binom{n_i}{y_i} + y_i \log p_i + (n_i - y_i) \log(1 - p_i) \right\}$$

$$= \sum_{i=1}^{n} \left\{ \log \binom{n_i}{y_i} + y_i \log[\ p_i / (1 - p_i)\ ] + n_i \log(1 - p_i) \right\}$$

$$= \sum_{i}^{n} \left\{ \log \binom{n_i}{y_i} + y_i \theta_i - n_i \log(1 + e^{\theta_i}) \right\}$$

where $\theta_i = \sum_{i=0}^{n} \beta_i x_i$ and $x_0 = 1$ for all values of i. So

$$L(\beta, x) = \sum_{i=1}^{n} \left\{ y_i \sum \beta_i x_i - n_i \log(1 + e^{\sum \beta_i X_i}) + \log \binom{n_i}{y_i} \right\}$$

The maximum likelihood estimator of β is obtained by setting the derivative of L(β) with respect to β to zero i.e

$$\partial \log L(\beta) / \partial \beta_j = \sum y_i x_{ji} - \sum n_i x_{ji} e^{\theta_i} (1 + e^{\theta_i})^{-1} = 0 \quad , \text{j=0,1,\ldots,n}$$

This leads to a set of n+1 non-linear equations with the unknown parameters $\hat{\beta}_j$ that can only be solved numerically. An algorithm known as Fisher's method of scoring can be used to obtain the maximum likelihood estimates $\hat{\beta}$. This procedure is equivalent to using an iteratively weighted least squares procedure in which values of the adjusted dependent variable $Z_i = [\theta_i + (y_i - n_i p)] / n_i p_i (1 - p_i)$ are regressed on the n explanatory variables $x_{1i}, x_{2i}, \ldots, x_{ni}$ using weights $w_i = n_i p_i (1 - p_i)$.

In this study, the logit link function and the binomial distribution were specified and the crucial aspect of dependence between neighboring observations was taken into consideration.

3.2.3 Results and Discussions

The extended log-linear models (Parameter-driven models) were applied to morbidity cases of the five communicable diseases while the logistic regression models proposed by Cox(1970) were fitted to the binary time counts which are based on both morbidity and mortality incidences. In the later case, the proportions of the incidences of the dead patients were taken as the response variables. Of interest is whether these data provided evidence of a decrease or an increase in the rate of infection of a particular disease in Kenya. This is depicted in the values of the estimated regression coefficients.

Tuberculosis

The results of the both estimated morbidity and mortality trends of TB incidences are as shown are given in the models

$$\log(E(y_t)) = 5.9357 + 0.0009X_t \quad , \text{P=0.0001}$$

and $$Logit(E(y_t)) = 6.08456 + 0.3833X_t \quad , \text{P =0.0001}$$

These results indicate that there were significant increases in the number of patients who suffered and those who died from TB in the period 1989-1997. The mortality trend is the same as the one exhibited in TB incidences although the rate of increase in deaths is much higher.

This scenario as already observed could be due to the association of TB with AIDS pandemic that was on a rapid increase and unavailability of drugs and as a result, patients could not be treated as required. In addition, most communities are not informed on the importance of respiratory symptoms such as persistent and productive cough, blood-stained sputum and chest pain particularly if they persist for more than three weeks and are of recent origin (Kibuga,1995).

Tetanus

The results for tetanus incidences were impressive as there were strong evidence of both decreasing morbidity and mortality trends. The results are given in the models

$$Log(E(y_t)) = -2.4778 - 0.0044X_t \quad ,(\text{ P }=0.0012)$$

and $$Logit(E(y_t)) = -0.9811 - 0.1874X_t \quad ,(\text{P }=0.0364)$$

The major reason for these decreasing trends was as a result of widespread immunization that has been going on in the country under the auspices of WHO (KMAR, 1994).

Poliomyelitis

The results for poliomyelitis give significant evidence of both decreasing morbidity and mortality trends as shown in the models

$$Log(E(y_t)) = -2.3384 - 0.0065X_t \quad , \text{P }=0.0001$$

and $$Logit(E(y_t)) = -2.1032 - 0.2559X_t \quad , \text{P }=0.0001$$

These happened as a result of immunization practices that have been performed effectively in Kenya through KEPI. Inspite of this, only a few children above one year are immunized against polio, tetanus and measles in especially difficult-to-reach districts (KMAR, 1995).

Measles

The parameter-driven model and logistic regression model for both measles morbidity and mortality incidences were obtained as

$$Log(E(y_t)) = -5.3389 - 0.0002X_t \quad , \text{P} = 0.0001$$

and
$$Logit(E(y_t)) = -1.9562 - 0.2816X_t \quad , \text{P} = 0.0001$$

These results indicate both decreasing morbidity and mortality trends in the period 1989-1997 as a result of up-to-date immunization practices based on the latest research outcomes and developments. In addition, public health education on the disease has been intensified (KMAR,1995).

Hepatitis

Hepatitis morbidity and mortality incidences show decreasing trends in the period 1989-1997 as given in the models

$$\log(E(y_t)) = -4.2954 - 0.0005X_t \quad , \text{P} = 0.0001$$

and
$$Logit(E(y_t)) = -2.3537 - 0.2246X_t \quad , \text{P} = 0.0001$$

This decrease has been as a result of improved health practices such as availability of drugs for this disease and increased public health education (KMAR,1995).

CHAPTER FOUR

APPLICATION OF TIME SERIES MODELS

4.1 INTRODUCTION

Time series models are used for forecasting future values of the series. Accurate forecasts of a time series are of great value in many areas of medicine and public health for monitoring the expected frequencies of diseases in the coming years so as to plan how to allocate the often limited resources.

The essential first step in the analysis of time series is to plot observations against time. Such plots are often valuable in highlighting trends, seasonality and outliers, although such patterns are frequently obscured by noise and this makes them extremely difficult to detect without more formal analysis. On the other hand, a simple plot of the data can often suggest patterns in the data which on further investigation, are found to be unreal.

An initial formal step in the analysis of a time series is to check for stationarity because most time series encountered in practice are not stationary. The importance of stationarity requirement is that it gives a sense of stability to the series such that a model developed using a single realization can be used for prediction. In essence, reliable information can be drawn from a stable series.

A time series X_t $\{t \in T\}$ is said to be strictly stationary if the joint distribution of $(X_1, X_2, \ldots, X_n)$ is the same as the joint distribution of $(X_{1+k}, X_{2+k}, \ldots, X_{n+k})$ for all k. This implies that the properties of the time series are unaffected by change in time origin. It also implies that if we consider the first two moments of the series then they will be constant, i.e $E(X_t)=\mu$, var $(X_t)=\sigma^2$ and cov $(X_t, X_{t+k})=f(k)$. Furthermore the autocovariance between values of the series separated by a particular number of time points depends only on the time difference or lag, not on the time instance. When we consider only the first two moments of the series, then a time series is said to be stationary in the weak sense or in the wide sense.

Transformation of a series to stationarity is usually a necessary prerequisite to fitting time series models. If a series has a trend component, then this component is removed by differencing , moving average, and least squares techniques. Time series are often non-stationary because of the presence of seasonality and cyclicity. This can be removed by the differencing technique. In some circumstances it may be appropriate to take some simple transformation of the original series before attempting to remove trend, cyclicity and seasonal effects. If for example, it is found that the variance is related to the mean, then a variance-stabilizing transformation should be applied.

A major goal underlying time series analysis is to develop models which adequately describe the mechanism generating the observations. A time series model accounts for patterns in the past movements of a variable and such information is used to control and predict its future movements.

Time series data are often modeled by autoregressive models or by a combination of autoregressive and moving average models. Autoregressive models express the observation X_t, as a linear function of the past values of the series. This type of process dates back to the work of Yule (1927) where he developed the first order autoregressive process (AR(1)) which is usually referred to as the first order markov process and is given by the relation

$$X_t = \phi X_{t-1} + e_t$$

where ϕ is the parameter to be estimated from the values of the series and the e_t are a series of independent, identically distributed random variables whose distribution is approximately normal with mean zero and variance σ^2.

The general autoregressive model of order p has the form

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + e_t$$

This model expresses the current observation as a linear combination of the past observation plus a residual disturbance term e_t, usually regarded as white noise. In many instances however, there is evidence that although the series has some features characteristic of an autoregressive model, the e_t are not white noise but are themselves correlated. A plausible model for the disturbance is that they are produced by a superposition of lagged white noise terms leading to the so called moving average processes.

The moving average (MA) process was developed by Slutzsky (1937). The functional form for the first order moving average (MA(1)) process is given by the equation

$$X_t = \theta e_{t-1} + e_t..$$

The moving average of order q is given by

$$X_t = \theta_1 e_{t-1} + \theta_2 e_{t-2} + \ldots + \theta_q e_{t-q} + e_t$$

where e_t is a white noise series and $\theta_1, \theta_2, \ldots, \theta_q$ are parameters to be estimated. Wold (1954) combined the AR and MA processes forming the autoregressive moving average (ARMA) process These kind of models are the most widely known and applied set of time series models given by the equation

$$\phi(B)X_t = \theta(B)e_t$$

where B is a backshift operator such that $BX_t = X_{t-1}$, $\{e_t\}$ is a sequence of uncorrelated random variables with zero mean and variance σ^2. The polynomials

$$\phi(B) = 1 + \phi_1 B + \phi_2 B^2 + \ldots + \phi_p B^p$$

and $$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \ldots + \theta_q B^q$$

represent the autoregressive and moving average operators of order p and q respectively, while the coefficients in $\phi(B)$ and $\theta(B)$ represent some of the model parameters. The most common model of this type is the first order autoregressive moving average ARMA(1,1) process. This is a linear process given by the relation

$$X_t = \phi X_{t-1} + \theta e_{t-1} + e_t$$

An autoregressive moving average model of order p and q usually abbreviated as ARMA(p,q) is given as

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \ldots + \theta_q e_{t-q} + e_t$$

Box and Jenkins (1976), suggested differencing as a means of transforming a non-stationary ARMA(p,q) process into a stationary ARMA process known as the autoregressive intergrated moving average abbreviated as ARIMA process. For instance a nonstationary ARMA(p,q) which requires differencing d times before it becomes stationary is said to follow an autoregressive intergrated moving average of order (p,d,q) abbreviated as ARIMA(p,d,q). The difference operator $\overline{V}$ when applied to the entry X_t yields the difference

$$\overline{V} X_t = X_t - X_{t-1}$$

Fitting a good model to a finite set of observed data involves an iterative approach to model building (Box and Jenkins,1976) which is based on identification of the order of the ARMA(p,q), model estimation of the parameters, diagnostic checking of the fitted model and finally the model is used for forecasting.

4.1.1 Model identification

A major task in time series analysis is the determination of the order of a process using a finite set of observations such that it leads to the most parsimonious model, i.e, it must have the least number of parameters which are not complicated. This in essence facilitates the estimation of the model parameters and the general usage of the model.

Identification of the ARMA(p,q) is based on both subjective and objective techniques that have been developed. The popular subjective technique is referred to as the Box and Jenkins (1976) approach which is based on plotting of the sample autocorrelation function (SACF), the sample partial autocorrelation function (SPACF) and sample inverse autocorrelation function (SIACF) of the series which are compared with the corresponding theoretical functions for various ARMA models (see Abraham and Ledolter,1983, pg.250). As an example, an autoregressive model of order one, has a correlogram in which the correlations gradually die away as the lag increases and a moving average process of order one, has a correlogram in which all the correlations for the lags greater than one are zero.

Another procedure is based on the autocorrelation function to identify the order (p,q) of an ARMA process. In this procedure the autocorrelation function is used to calculate arrays of numbers called the R- and S- arrays for which p and q may be determined. There is also the corner method of Bequin et al, (1981) which uses the autocorrelation to generate an array of numbers for identifying the order of a time series process.

Objective procedures are based on information theory. Akaike (1970) proposed a method, which chooses the order p of the autoregression so that the expected one-step-ahead forecast variance is minimized. It is called the final prediction error (FPE) and is given by

$$FPE(p)=\sigma^2 (n+p)/(n-p)$$

where σ^2 is the variance that is to be estimated and n is the sample size. The order is chosen such that the final prediction error (FPE(p), p=0,1,2,...,L) attains the minimum value with L being the maximum order cut off imposed for the process. Another objective method developed by Akaike (1973,1974) is referred to as the Akaike information criterion (AIC) and expressed as

$$AIC(p,q)=n\log \sigma^2 +2(p+q).$$

where n is the sample size and σ^2 is the variance to be estimated.

Other identification procedures are those that are based on Bayesian techniques. They make use of the priori knowledge about the model, which is in the form of a probability density function. The Bayesian information criterion (BIC), Bayesian estimation criterion (BEC) and Hannan and Quinn criterion (1979) are some of the identification criteria under Bayesian methods.

4.1.2 Parameter estimation

Once a particular model has been identified, the next crucial aspect is the estimation of parameters. Some estimation techniques employed include the Yule-walker estimation criteria , the maximum likelihood criteria, the conditional and the unconditional least squares method and the optimization criterion. The Yule-walker method is based on the autocovariance and autocorrelation values estimated from the sample data as outlined by the Box and Jenkins (1976). Yule-walker estimates for AR(p) are efficient and have sim[illegible]on properties as the maximum likelihood estimates. These estimates [illegible]icient estimates in the case of MA(q) processes because the estimates are expressed as a function of unobservable noise process.

Conditional maximum likelihood estimation procedures entails the maximization of the likelihood function g(ϕ, θ, λ) and in practice, the likelihood function is complicated. However this has been overcome by the development of various algorithms for computing maximum likelihood estimates (Ansley,1979). The unconditional maximum likelihood criteria is a popular estimating technique for non-independent variables although the evaluation of the analytical forms of the maximum likelihood estimates for the case of dependent variables is complicated. Another major drawback is that the nature of the likelihood function for models with the dependent variables is very difficult to set up.

Conditional least squares estimate (CLSE) are computationally simple estimates compared to the maximum likelihood estimates. This procedure entails minimizing of the conditional sum of squares.

$$S_c(\phi, \theta) = \sum_{t=p+1}^{T} e_t^2$$

where the starting values $e_p, e_{p-1}, \ldots, e_{p+1-q}$ are set equal to their expected value of zero. The conditional least squares approach is comparable to the maximum likelihood when values are away from the invertibility boundaries, say in the case of MA(1), when the values are away from $0 \leq \theta \leq 1$. It also tends to overestimate the variance of the process X_t while the use of the unconditional least squares procedure leads to underestimation.

For the case of the AR(p) processes, the conditional least squares estimates are functions of the observations and thus they can be shown to be very efficient.

Another method is the technique based on the estimating functions. This method of estimating functions has an advantage over the least squares, maximum likelihood and unbiased minimum variance methods because optimal estimates are obtained in the sense that unbiased, consistent estimates are obtained and it combines the strengths of the three methods while at the same time eliminating their weakness (Godambe,1991). One major advantage with the optimal estimation procedure is the fact that the probability density function for the stochastic process whose parameters are being estimated need not exist except the first two moments. This technique is thus superior to other techniques in situations where there is dependence of the conditional mean and variance on time and also in situations whereby the distribution may not be fully specified apart from the first two moments.

The next step in time series model formation and analysis is diagnostic checking. This is aimed at determining the adequacy of the fitted model. This procedure entails the comparison of the ACF for the series based on the fitted model with the sample ACF of the original series.

4.1.3 Forecasting

Once a time series model has been identified, and its parameters estimated, it can be used for forecasting. A major use of time series is in their application to forecasting (Box and Jenkins,1976). There are various forecasting techniques used in time series analysis and these include forecasting by extrapolation of polynomial trends, exponential smoothing and by the minimization of the mean square error criteria. The commonest criteria used when forecasting is to minimize the mean square error, i.e for the process X_t, we minimize

$$E(X_{n+h}-X_{n(h)})^2$$

where $X_{n(h)} = E(X_{n+h})/\underline{X}$ ... ecast function based on a sample of size n, h is referred to as the lead tim... s the future value. This leads to minimum mean square error (MMSE) forecasts. The forecast for X_{n+h} is a linear combination of past and present values of X_t and it is thus regarded as the conditional expectation of X_{n+h} given $X_n, X_{n-1},\ldots$

The h-step forecast error for the origin n is defined as

$$e_{n(h)}=X_{n+h}-X_{n(h)}$$

This is also a linear combination of the unobservable future stocks entering after time n, and in particular, the one-step ahead forecast error (h=1) is obtained as

$$e_{n(1)}=e_{n+1}$$

Since $E[e_{n(h)}/X_n,X_{n-1},\ldots] = 0$, $X_{n(h)}$ is said to be an unbiased forecast. The variance of the forecast error is then obtained as

$$Var(e_{n(h)})= Var[X_{n+h}-X_{n(h)}]$$

For the AR(1) process, $X_t=\phi X_{t-1}+ e_t$,

given a set of n observations, the h steps ahead future value is given by

$$X_{n+h}=\phi X_{n+h-1}+e_{n+h}$$

and thus the forecast value h steps ahead is

$$X_{n(h)}= E[[\phi X_{n+h-1}+e_{n+h}/\underline{X}]]$$

When h = 1, we have

$$X_{n+1}=\phi X_n+e_{n+1}$$

and thus the forecast value ... d is

$$X_{n(1)} = E(X_{n+1} / \underline{X}) = \phi X_n$$

The forecast error is obtained as

$$e_{n(1)} = \phi X_{n+1} + e_{n+1} - \phi X_n = e_{n+1}$$

and the corresponding variance of the forecast error is

$$\text{var}[e_{n(1)}]=(1-\phi)E(e^2_t)=(1-\phi)\omega$$

When h = 2, we have

$$X_{n+2} = \phi X_{n+1} + e_{n+2} = \phi^2 X_n + \phi e_{n+1} + e_{n+2}$$

and therefore the forecast value two steps ahead is

$$X_{n(2)} = \phi^2 X_n$$

The corresponding forecast error is obtained as

$$e_{n(2)} = \phi e_{n+1} + e_{n+2}$$

and the variance of the forecast error is

$$\text{var}[e_{n(2)}] = [(1-\phi) + \phi^2]E(e_t^2)$$

$$=[1-\phi + \phi^2]\omega$$

In general the h-step ahead forecast is obtained as

$$X_{n(h)} = \phi^h X_n$$

and the corresponding h-step ahead forecast error is obtained as

$$e_{n(h)} = X_{n+h} - X_{n(h)}$$

The corresponding variance of the forecast error is obtained as

$$\text{var}[(e_{n(h)})] = (1-\phi)[1 + \phi + \phi^2 + \ldots + \phi^{h-1}]\omega$$

$$= (1-\phi^{h})\omega$$

Tuberculosis

An ARMA(1,1) model was used to obtain both back and future forecasts and the parameters were estimated using the maximum likelihood procedure with no constant. This model was found to be the best as compared to the others because it had the minimum variance estimate, the forecasts were close to the actual data values and mean sum of forecast residuals was minimal (MSE = 73145.185).

Back forecasts was done from 1997 to Oct 1996 and it indicated that there was an increase of TB cases from 1996 to 1997 but become constant towards the end of 1997 which is the actual case from the time plot of TB cases versus years (1989-1997) . Future forecasts for 1998 and 1999 was done and it showed that TB cases were still high but under the current conditions, a slight reduction in the number of cases in1998 and 1999 is expected. The back and future forecasts are shown in Table 1 and Table 2 respectively.

Table 1: Back forecasts for variable TB cases from oct 1996 to dec 1997.

Observation	Forecast	Actual	Residual
1996 oct	750.2130	1177.0000	426.7870
1996 nov	750.5102	1024.0000	273.4898
1996 dec	750.7495	899.0000	148.2505
1997 jan	750.9423	1302.0000	551.0577
1997 feb	751.0976	1232.0000	480.9024
1997 mar	751.2226	910.0000	158.7774
1997 apr	751.3234	819.0000	67.6766
1997 may	751.4046	1057.0000	305.5954
1997 june	751.4700	1023.0000	271.5300
1997 july	51.5226	816.0000	64.4774
1997 aug	751.5650	792.0000	40.4350
1997 sep	751.5992	812.0000	60.4008
1997 oct	751.6267	920.0000	168.3733
1997 nov	751.6489	942.0000	190.3511
1997 dec	751.6668	863.0000	111.3332

Minimum mean square error (MMSE)= 73145.185_

Table 2: Forecast for Jan 1998 to Dec 1999 for TB variable Cases.

Observation	Forecast	Std. error
1998 jan	854.6795	168.1878
1998 feb	830.4363	187.1759
1998 march	816.4704	204.7858
1998 april	803.4568	213.8384
1998 may	793.5335	219.9030
1998 june	785.3627	223.6511
1998 july	778.8373	226.0748
1998 aug	773.5631	227.6273
1998 sep	769.3204	228.6306
1998 oct	765.9010	229.2789
1998 nov	763.1471	229.6986
1998 dec	760.9287	229.9705
1999 jan	759.1418	230.1468
1999 feb	757.7024	230.2611
1999 march	756.5430	230.3352
1999 april	755.6090	230.3833
1999 may	754.8567	230.4145
1999 june	754.2507	230.4347
1999 july	753.7625	230.4479
1999 aug	753.3693	230.4564
1999 sep	753.0526	230.4619
1999 oct	752.7975	230.4655
1999 nov	752.5919	230.4678
1999 dec	752.4264	230.4694

Mean standard error (MSE)=223.0653

Poliomyelitis

In the analysis, the ARMA(1,1) model was used and the results produced were optimal. Back forecasts were obtained for 1997 to Oct 1996 and they indicated that there was a great decline of polio cases from eleven cases in Oct 1996 to two cases in Dec 1997 which is almost true in comparison to actual values. Forecasts were obtained for 1998 to 1999 and the results showed in that by the end of 1999 and beyond, it is expected that there will be nobody to be infected by poliomyelitis under the present immunization program. This goes hand in hand with the World health assembly objective to eradicate this disease by the year 2,000 and expects that immunization to be most efficiently provided along with other health services (WHO annual report,1994). Back and future forecasts are given in Table 3 and Table 4.

Table 3: The back forecasts for polio variable cases from Oct 1996 to Dec.

Observation	Forecast	Actual	Residual
1996 oct	10.6767	62.0000	51.3233
1996 nov	9.4548	73.0000	63.5452
1996 dec	8.3727	31.0000	22.6273
1997 jan	7.4144	15.0000	7.5856
1997 feb	6.5658	17.0000	10.4342
1997 march	5.8144	9.0000	3.1856
1997 april	5.1489	12.0000	6.8511
1997 may	4.5596	27.0000	22.4404
1997 june	4.0378	23.0000	18.9622
1997 july	3.5757	52.0000	48.4243
1997 aug	3.1664	34.0000	30.8336
1997 sep	2.8040	18.0000	15.1960
1997 oct	2.4831	17.0000	14.5169
1997 nov	2.1989	24.0000	21.8011
1997 dec	1.9473	17.0000	15.0527

MMSE=1.5×10^9

Table 4: Forecasts for Jan 1998 to Dec1999 for polio variable cases.

Observation	Forecast	Std.error
1998 jan	15.0544	41.4366
1998 feb	13.3314	55.3485
1998 march	11.8056	64.1822
1998 april	10.4545	70.3377
1998 may	9.2580	74.8113
1998 june	8.1984	78.1406
1998 july	7.2601	80.6553
1998 aug	6.4292	82.5738
1998 sep	5.6934	84.0477
1998 oct	5.0418	85.1856
1998 nov	4.4647	86.0675
1998 dec	3.9538	86.7527
1999 jan	3.5013	87.2864
1999 feb	3.1005	87.7026
1999 march	2.7457	88.0276
1999 april	2.4314	88.2816
1999 may	2.1532	88.4803
1999 june	1.9067	88.6359
1999 july	1.6885	88.7576
1999 aug	1.4953	88.8530
1999 sep	1.3241	88.9277
1999 oct	1.1726	88.9863
1999 nov	1.0384	89.0321
1999 dec	0.9195	89.0681

MSE = 81.328

Tetanus

An AR (2) model was used in the analysis of tetanus data and it was the best in terms of minimum variance estimate and minimum mean square forecast error as compared to the other models. Back forecasts from 1997 to Oct 1996 indicated that there was a dramatic decline of Tetanus cases from seventeen cases in Oct 1996 to seven cases in Dec 1997. Forecasts for 1998 to 1999 indicated that there will be a decline of cases or no case at all to be reported in future. This aspect augurs well with KEPI's goal of

Tetanus elimination by the year 2,000 (WHO annual report, 1994). The results of the back and future forecasts are given in Table 5 and Table 6.

Table 5: Back forecasts from Oct 1996 to Dec 1997 for tetanus variable cases.

Observation	Forecast	Actual	Residual
1996 oct	17.1272	38.0000	20.8728
1996 nov	16.0888	15.0000	-1.0888
1996 dec	15.1154	18.0000	2.8846
1997 jan	14.1998	18.0000	3.8002
1997 feb	13.3403	33.0000	19.6597
1997 march	12.5325	52.0000	39.4675
1997 april	11.7737	55.0000	43.2263
1997 may	11.0608	38.0000	26.9392
1997 june	10.3912	39.0000	28.6088
1997 july	9.7620	14.0000	4.2380
1997 aug	9.1709	18.0000	8.8291
1997 sep	8.6157	23.0000	14.3843
1997 oct	8.0940	15.0000	6.9060
1997 nov	7.6040	9.0000	1.3960
1997 dec	7.1436	12.0000	4.8564

MMSE=412.83693

Table 6: Forecasts for Jan 1998 to Dec 1999 for tetanus variable cases.

Observation	Forecast	Std.error
1998 jan	9.5080	36.7657
1998 feb	9.8115	40.1884
1998 march	8.7796	46.9978
1998 april	8.4661	50.4528
1998 may	7.8449	54.0899
1998 june	7.4240	56.7566
1998 july	6.9476	59.1661
1998 aug	6.5403	61.1428
1998 sep	6.1377	62.8681
1998 oct	5.7694	64.3375
1998 nov	5.4184	65.6138
1999 dec	5.0912	66.7169
1999 jan	4.7825	67.6769
1999 feb	4.4931	68.5124
1999 march	4.2210	69.2417
1999 april	3.9655	69.8788
1999 may	3.7254	70.4365
1999 june	3.4998	70.9250
1999 july	3.2879	71.3533
1999 aug	3.0888	71.7292
1999 sep	2.9018	72.0594
1999 oct	2.7261	72.3495
1999 nov	2.5611	72.6046
1999 dec	2.4060	72.8290

MSE=63.112

Measles

During the analysis, an ARIMA (1,1) model was used due to the optimal results it produced as compared to other models. Back forecasts from 1997 to Oct 1996 indicated that the number of cases increased slightly. Forecasts indicated also that there was going to be a slight increase of cases in future, possibly due to increase of population and high-level of poverty. The results of back and future forecasts are given in Table 7 and Table 8.

Table 7: Back forecasts from Oct 1996 to Dec 1997 for measles variable cases.

Observation	Forecast	Actual	Residual
1996 oct	2746.7867	1435.0000	-1311.787
1996 nov	2790.5563	1862.0000	-928.5563
1996 dec	2832.7734	1850.0000	-982.7734
1997 jan	2873.4930	1898.0000	-975.4930
1997 feb	2912.7683	2019.0000	-893.7683
1997 mar	2950.6505	1850.0000	-1100.650
1997 apr	2987.1890	2480.0000	-507.1890
1997 may	3022.4315	3324.0000	301.5685
1997 june	3056.4239	2852.0000	-204.4239
1997 july	3089.2106	2812.0000	-277.2106
1997 aug	3120.8343	2432.0000	-688.8343
1997 sep	3151.3363	2213.0000	-938.3363
1997 oct	3180.7565	1560.0000	-1620.756
1997 nov	3209.1331	1712.0000	-1497.133
1997 dec	3236.5031	1226.0000	-2010.503

MMSE = 1149523.4

Table 8: Forecasts for Jan1998 to Dec 1999 for measles variable cases.

Observation	Forecast	Std.error
1998 jan	1545.2642	788.0332
1998 feb	1631.6521	927.7742
1998 mar	1714.9757	1041.0696
1998 apr	1795.3438	1136.3730
1998 may	1872.8613	1218.3599
1998 june	1947.6292	1289.9634
1998 july	2019.7451	2019.7451
1998 aug	2089.3030	1409.4471
1998 sep	2156.3936	1459.8478
1998 oct	2221.1046	1505.2218
1998 nov	2283.5202	1546.2390
1998 dec	2343.7220	1583.4444
1999 jan	2401.7884	1617.2888
1999 feb	2457.7951	1648.1509
1999 mar	2511.8153	1676.3525
1999 apr	2563.9194	1702.1693
1999 may	2614.1753	1725.8405
1999 june	2662.6486	1747.5743
1999 july	2709.4026	1767.5538
1999 aug	2754.4982	1785.9403
1999 sep	2797.9943	1802.8773
1999 oct	2839.9475	1818.4924
1999 nov	2880.4127	1832.9000
1999 dec	2919.4425	1846.2027

MSE=1457.096

Hepatitis

Back and future forecasts were obtained using an AR (2) model, which had minimum variance estimate, and minimum mean square forecast error. The optimal results produced indicated that for back forecasts from 1997 to Oct 1996, there was a slight decrease of hepatitis cases. Forecasts indicated also a slight reduction of people suffering from hepatitis in future. This slight decrease could be due to the fact it has been hard for people to change their moral behaviors especially the homosexuals and those who practice prostitution. The results of future and back forecasts are given in tables 9 and 10.

Table 9: Back forecasts from Oct 1996 to Dec1997 for hepatitis variable cases.

Observation	Forecast	Actual	Residual
1996 oct	205.7460	431.0000	225.2540
1996 nov	197.1167	302.0000	104.8833
1996 dec	188.8494	243.0000	54.1506
1997 jan	180.9288	418.0000	237.0712
1997 feb	173.3404	520.0000	346.6596
1997 mar	166.0702	556.0000	389.9298
1997 apr	159.1050	489.0000	329.8950
1997 may	152.4319	385.0000	232.5681
1997 june	146.0387	588.0000	441.9613
1997 july	139.9136	756.0000	616.0864
1997 aug	134.0454	812.0000	677.9546
1997 sep	128.4234	712.0000	583.5766
1997 oct	123.0371	614.0000	490.9629
1997 nov	117.8768	248.0000	130.1232
1997 dec	112.9328	187.0000	74.0672

MMSE = 1.013×10^{12}

Table 10: Forecasts for Jan 1998 to Dec 1999 for hepatitis variable cases.

Observation	Forecast	Std.error
1998 jan	182.0973	404.5868
1998 feb	174.2890	544.3016
1998 mar	166.9890	647.2292
1998 apr	159.9847	728.9870
1998 may	153.2747	796.6813
1998 june	146.8461	854.1068
1998 july	140.6872	903.6097
1998 aug	134.7866	946.7718
1998 sep	129.1334	984.7255
1998 oct	123.7174	1018.3180
1998 nov	118.5285	1048.2045
1998 dec	113.5573	1074.9056
1999 jan	108.7945	1098.8429
1999 feb	104.2315	1120.3643
1999 mar	99.8599	1139.7607
1999 apr	95.6716	1157.2780
1999 may	91.6590	1173.1266
1999 june	87.8147	1187.4874
1999 july	84.1316	1200.5177
1999 aug	80.6030	1212.3546
1999 sep	77.2224	1223.1187
1999 oct	73.9836	1232.9160
1999 nov	70.8806	1241.8407
1999 dec	67.9078	1249.9763

MSE=1007.909

CHAPTER FIVE

CONCLUSION AND RECOMMENDATIONS

In this study, the application of regression models, time-dependent regression models and time series models to count data of some communicable diseases in Kenya has been studied, with major emphasis on the assessment of trends and the evaluation of forecasts of these diseases.

The results of this analysis indicate that most of these communicable diseases have been contained as a result of immunization that has been effectively and efficiently implemented in the country. These diseases include Poliomyelitis, Tetanus and Measles. However, the trend of TB incidences is on the increase and hence there is need to find ways of curbing it.

Application of regression models constitute the traditional approach of assessing trends of various diseases. In this study, inadequate inferences were obtained when these models were applied because the patterns exhibited by time series data are generally far more complex to be accommodated by the regression models. Time-dependent regression models such as poisson regression models and logistic regression models provide valid inferences about the trends of the diseases as opposed to the ordinary regression models. This is due to the fact that the autocorrelation between the data values, the link function and the distribution of the innovations are accounted for in these models. Therefore application of these models is highly recommended as opposed to the subjective assessments of trends which have been the basis for decision making regarding the management and control of the diseases.

Accurate forecasts of future values are obtained when time series models are applied due to the fact that all the crucial aspects pertaining to time series data have been taken into consideration. As a result, the use of time series models is recommended in various sectors. For instance in medicine, the future situation such as outbreaks of a particular disease can be controlled before hand.

BIBLIOGRAPHY

Abraham, B., and Ledolter, J. (1983). Statistical methods for forecasting. John wiley and sons. New York, U.S.A.

Agresti, A. (1990). Categorical data analysis.John wiley and sons,Inc.U.S.A.

Akaike, H. (1970). Statistical predictor identification. *Ann. Statist. Math*, **22**, 203-217.

Akaike,H. (1973). Information theory and an extension of the maximum likelihood principle. *2nd int,symp.on inf. Theory.* eds.B.N. Petrov and F.Csaki. Pg. 267-281. Akademia kiado,Budapest.

Akaike, H. (1974). Markovian representation of stochastic processes and its application to the autoregressive moving average process. *Ann. Inst. Statist. Math* . **26**, 363-387.

Anderson, R.M and Robert, M.M. (1992). Infectious diseases of humans. Oxford science publications. New york, U.S.A.

Ansley, C. F. (1979). An algorithm for the exact likelihood of a mixed autoregressive moving average process. *Biometrika vol.* **66,** 547-555.

Bequin, J.M. Gourieroux, C. and Monfort, A. (1981). Identification of a mixed autoregressive moving average process: Corner method. In time series analysis. Ed. O.D. Anderson, Pg. 423-436. Amsterdam, Holland.

Benenson, A.S. (1990). Control of communicable diseases in man. American public health association. Washington, D.C. , U.S.A.

Box, G. E. P, and Jenkins, G.N. (1976). Time series analysis. Forecasting and control. Holden day. San Francissco,U.S.A.

Box, G. E.P. and Tia, G. C. (1975). Intervention analysis with applications to economic and environmental problems. *J. Amer. Stat. Assoc.* **70,** 70-79.

Brownlee, J. (1906). Statistical studies in immunity: the theory of an epidemic. *Proc. Roy. Soc. Edin,* **26**, 484-521.

Carrol, N. , Miller, J. and Nash, W. (1976). Health: science of human adaptation. *Wm.C.* Brown company. Iowa, U.S.A

Central bureau of statistics. (1952-1994). Statistical abstracts. Kenya government printers. Nairobi, Kenya.

Collet , D. (1991). Modeling binary data. T.J.press (Padstow) ltd, Britain.

Cox, D. R. and Snell, E.J. (1970). Analysis of binary data. (2^{nd} edition). Chapman and Hall, London.

Diggle, P. J. (1990). Time series: A Biostatistical introduction. Oxford University Press, London.

Diggle, J.P. , Liang, K. and Zeger, L. S. (1996). Analysis of longitudinal data. Clarendon press. Oxford.

Everitt B.S. (1994). Statistical methods for medical investigations.University press, London.

Farr, W. (1840). Progress of epidemics. *Second report of the registrar General of* England. Pg. 91-8.

Godambe, V.P. (1985). The foundations of finite-sample estimation in stochastic processes. *Biometrika, Vol* **72**, 414-428.

Godambe, V.P. (1991). Estimating functions. Oxford university press. Newyork.

Granger, C. W. J. and Newbold, P. (1976). Forecasting transformed time series. *JRSS, Ser. B. Vol.* **38**, 189-203.

Hannan, E. J. and Quinn, B.G. (1979). The determination of the order of autoregression. *J. R. Statist. Soc.* **B-41**, 190-195.

Hamer, W. H. (1906). Epidemic disease in England. *The Lancet.* **I**, 733-9

Kenya medical annual reports (KMAR). (1952-1958, 1962, 1982, 1994, 1995, 1998). Kenya government printers. Nairobi, Kenya.

Kibuga, D. (1995). The National Leprosy and Tuberculosis programme manual. (NLTP). Ministry of health: Kenya.

Klimko, L.A. and Nelson, P.I. (1978). Conditional least squares estimation for stochastic processes. *Ann. Statist.* Vol. **6,** 629-642.

McCullagh, P. and Nelder, J.A. (1983). Generalized linear models. Chapman and Hall, London.

Mochkovskii, S. D. (1950). Basic laws of the epidemiology of malaria. AMN, Moscow.

Nelder, J.A. and Wedderburn, R.W.M. (1972). Generalized linear models. *J.R. Statist. Soc.* **A135**, 370-384.

Newbold, P. and Granger, C.W. (1974). Experience with forecasting univariate time series and conbination of forecasts. *J. R. Statist. Soc.* **A137**, 131-146.

Njogu, A.R. and Tukei, P.M. (1982). Tropical diseases : Trends in research. Government printers. Nairobi, Kenya.

Ross, R. (1911). The prevention of malaria, (2^{nd} edition). John Murray, London.

Ross, R. (1915). Some a priori pathometric equations. *British medical journal.* **1**,546-7.

Ross, R. (1916). An application of the theory of probabilites to the study of a priori pathometry,II. *Proc. R. Soc* **A93**, 212-25.

Ross, R. (1917). An application of the theory of probabilities to the study of a priori Pathometry, II. *Proc. R. Soc.,* **A93**,212-25.

Schartz, G. (1978). Estimating the dimension of a model. *Ann. Statist.* **6**, 461-464.

Shumway, H.R. (1988). Applied statistical time series analysis. Prentice Hall. New Jersey, U.S.A.

Slutzky, H. (1937). The summation of random causes as the source of cyclic process; *Econometricia vol.* **5,** 105-146.

Weddeburn, W.M. (1976). On the existence and uniqueness of MLE for certain generalized linear models. *Biometrika* **63,** 27-32.

Wetheril,G.B (1986). Regression analysis with applications. Chapman and Hall Limited, London.

Witte, J.J. (1977). We are not immunizing enough of children.Today's Health. *Pg. 4-5*

Wold, H. O. A. (1954). A study in the analysis of stationary time series. (2^{nd}, ed. ,1954). Almquist and Wiksell: Uppsala. Sweden.

World Health organization (WHO) (1998). *Bulletin.* **56,** 21-38.

World health organization (WHO) (1994). Annual report.

Yule, G. U. (1927). Methods of investigating periodicities in disturbed series with special reference to wolfer's sunspot numbers. *Phil. Trans. Roy. Soc.* **A226,** 267-298.

Zeger, L.S. (1988). A regression model for time series of counts. *Biometrika* **75**, 621-629.

Zeger, L.S and Qagish, B (1988). Markov regression models for time series: A quasi-likelihood approach. *Biometrics* **44**, 1019-1031.

Zeger, L.S and Moulton, L.H (1989). Analyzing repeated measures on generalized linear models via the bootstrap. *Biometrics* **45**, 381-394.

Zeger, l. S ; Liang and Albert, P.S (1988). Models for longitudinal data. A generalized estimating equation approach. *Biometrics* **44,** 1049-1060.