Modelling of Smear Positive Tuberculosis (PTB+) Treatment Outcome Data in Kenya for the period 2002-2007

BY

KIPRUTO HILLARY KIPCHUMBA (B.Sc, M.Sc)

School of Mathematics

University of Nairobi

A dissertation submitted in partial fulfilment of the requirements for the award of the degree of Master of Science (Biometry) in the School mathematics of University of Nairobi

UNIVERSITY of Nairobi

August, 2009

University of NAIROBI Library

0378976 5

DECLARATION

I declare that this research project is my original work and has not been presented for a degree in any other university or for any other award.

HILLARY KIPCHUMBA KIPRUTO ---------- 25/08/2009

Signature Date

This research project is submitted for examination with my approval as university supervisor.

Dr. Rosemary Nguti
School of Mathematics,
Mathematics department,
University of Nairobi,

---------- 26/08/2009

Signature Date

Table of Contents

List of Tables

List of Figures

Abbreviations

AIDS	Acquired Immunodeficiency Syndrome
ARV	Anti-retroviral
CNR	Case Notification Rate
DLTLD	Division of Leprosy, Tuberculosis and Lung Disease
DOTS	Directly-Observed Treatment, Short-course
DTLC	District Tuberculosis and Leprosy Coordinator
GoK	Government of Kenya
HIV	Human Immunodeficiency Virus
PTB	Pulmonary Tuberculosis
PTB SS-	Pulmonary TB, sputum smear negative
PTB PS+	Pulmonary TB, sputum smear positive
TB	Tuberculosis
WHO	World Health Organization
EPTB	Extra-Pulmonary Tuberculosis
MDR-TB	Multi Drug Resistant TB
XDR-TB	Extremely Drug Resistant TB

Acknowledgement

First and foremost I thank God. Secondly, immeasurable gratitude to my supervisor Dr. Rosemary Nguti for her advice, patient counsel and support throughout this research project. Special appreciation to Dr. Joseph Sitienei Head of The Division of Leprosy, Tuberculosis and Lung Disease, Ministry of Public Health and Sanitation for giving consent to carry out my research work at the Division.

Special appreciation and thanks goes to colleagues who provided assistance in many ways and many friends who walked with me, encouraging me in every step. To my class mates Aiban Rono, Alex Wambua, Joan Thiga, Festus Kiogora and Gideon Emukule you were the best team to study with and for that thank you very much.

Finally, to my beloved family, thank you for your constant encouragement and support, your prayer took me an extra mile. Many blessings and I dedicate this work to all of you.

Abstract

In this research project, we review and use two statistical methods namely poisson regression and multinomial logit regression are used to model rates of outcomes of count data.

The study reviews the research work on modelling of count data particularly infectious diseases. There has been wide use of the two methods in literature but little application to Tuberculosis modelling. The outcome of smear positive tuberculosis cases namely cure, treatment completed, out of control, failures, transfer outs and deaths were modelled using poisson regression and multinomial regression. The results showed that the two methods can be used to model epidemiological behaviour of the infectious Tuberculosis and the two methods can be adopted at the national and provincial levels to track the occurrence of the outcomes over the years and across the provinces.

The results can be used to track performance of country in terms of the occurrence of treatment outcomes and the models can be used in provinces and districts. The models can act as strong epidemiological analysis tools.

Operational definitions

Smear-positive pulmonary tuberculosis case: At least one sputum examination AFB+ (graded 1+ or greater)

Culture-positive pulmonary tuberculosis case: At least one culture-positive for M. tuberculosis.

Bacteriological confirmed pulmonary tuberculosis case: Smear-positive or culture-positive pulmonary tuberculosis case

Smear-negative pulmonary tuberculosis

Sputum culture positive for *M. tuberculosis* and 2 sputum smear examinations negative for AFB.

Asymptomatic TB case: A person diagnosed with smear-positive or smear-negative tuberculosis, but without suggestive symptoms of cough, hemoptysis, weight loss, fever and/or night sweats.

Tuberculosis suspect: Any person who presents with symptoms or signs suggestive of TB, in particular cough for 3 weeks or longer or has abnormalities on their X-ray.

New case: Patient who has never previously been treated for tuberculosis or who has taken anti tuberculosis drug for less than a month.

Relapse case: Patient who was previously declared cured or who has completed treatment but with a new episode of bacteriologically positive (smear or culture) tuberculosis.

Default case: Patient whose treatment was interrupted for two consecutive months or more after at least one month of treatment.

Failure case: Patient who completed five months or more of treatment but who remained, or again became, bacteriological positive (smear or culture).

CHAPTER ONE

1.0 Introduction

1.1 Background:

Globally, the World Health Organization (WHO) estimates that 2 billion people, or 1/3 of the world's population, are infected with (*Mycobacterium tuberculosis*) the bacillus that causes Tuberculosis (TB). In 2007, 1.7 million people died from TB, making it the leading infectious cause of death worldwide, (WHO, 2008).

Of those infected with TB, only 10% will ever develop the disease. Healthy immune systems can contain TB bacilli in the vast majority of cases. Individuals at higher risk of developing TB are those with lowered immunity – notably people who are HIV+ and those living in poverty. While HIV attacks the body's immune system directly, characteristics of poverty such as poor nutrition, inadequate housing, limited access to clean water and bad sanitation reduces resistance to TB.

Those who develop TB will develop one of two types. Pulmonary TB (PTB), which accounts for about 80% of TB cases. This is highly infectious and attacks the lungs. There are two sub-categorizations of PTB: one that shows up in a sputum sample (PTB SS+) and one that requires further tests, such as a chest x-ray, to diagnose (PTB SS-). The second type, Extra-pulmonary TB (EPTB) affects organs other than the lungs, such as the spine, lymph nodes or abdomen. EPTB is not infectious, unless it is accompanied by PTB.

To combat all forms of TB, a global strategy for diagnosis and treatment was adopted by the WHO, for implementation by national TB programmes, (WHO, 2003). The strategy, called Directly-Observed Treatment, Short-course or DOTS, is simple, based on identifying TB cases in the community, and treating these cases by directly observing

that patients take the correct drug treatment for 6 to 8 months (depending on the type of treatment). The objectives of this strategy are to ensure that the patient is cured, thus minimizing the chances that the patient will relapse, and prevent the development of drug-resistant TB.

Under DOTS, TB patients require a guardian to ensure and document that they have taken their treatment – often this is a family member, community health worker or neighbour. Their role is particularly crucial during the first two months of treatment, known as the intensive phase, when the strongest drugs are used. After two months, patients undergo a sputum examination to determine the concentrations of TB bacilli in their lungs. Depending on the results, the drug therapy is changed and TB patients enter the continuation phase of treatment for 4 or 6 more months (see Figure 1 for treatment process outline).

In the intensive phase, most patients or their guardians are required to pick up their drugs from the designated health facility once per week. During the continuation phase, this changes to once per month.

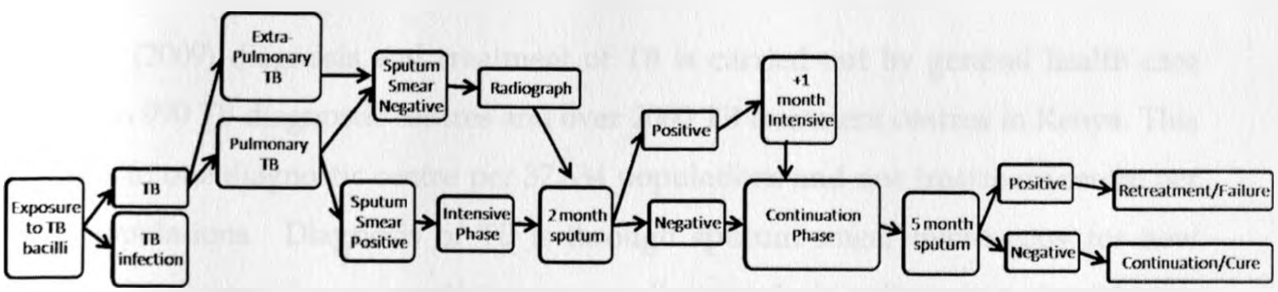


Figure 1 TB Treatment Process

The length of treatment, frequency of drug collection, and human resources required by DOTS are all onerous aspects of TB treatment, involving direct financial and indirect opportunity costs. As it is widely held in the discourse that poor socioeconomic standing and TB are closely linked, the susceptibility of those who are already poor to a

disease that leads to further impoverishment due to illness and treatment demands constitutes a double burden of the most economically vulnerable.

Furthermore, where TB is widespread, an undiagnosed TB case can lead to between 10 and 15 new infections, (WHO, 2008). One of the most widely-associated symptoms of TB is haemoptysis (coughing up blood), but in fact, this is a late-developing symptom - WHO guidelines recommend that patients seek medical attention after two weeks of a chronic cough, (WHO, 2003). As such, the pre-diagnosis phase is a crucial time period for the individual as well as for TB control.

Kenya is among the 22 TB high burden countries in the world. The absolute number of TB cases notified has increased more than tenfold since 1990 while the TB incidence has increased from below 50 per 100,000 in 1990 to 329 per 100,000 population in 2008. The HIV epidemic is the single most contributing factor for this massive increase in the burden of TB in Kenya. From the Kenya AIDS Indicator Survey 2007 the prevalence of HIV in Kenya currently stands at 7.4% (NASCOP, 2008) while the TB HIV co-infection rate is at 45% in 2008 (DLTLD, 2008).

Currently (2009) diagnosis and treatment of TB is carried out by general health care workers in 990 TB diagnostic centres and over 2000 TB treatment centres in Kenya. This translates to one diagnostic centre per 37,634 populations and one treatment centre per 18,411 populations. Diagnosis of TB is through sputum smear microscopy for new smear positive cases. Smear negative cases are diagnosed via a diagnostic algorithm as per the national TB guidelines. The diagnosis of extra pulmonary TB is based on clinical suspicion and the collection of appropriate specimens for TB bacteriology where this is feasible.

Kenya has developed a medium term TB control strategic plan covering the period 2006-2010 which is modelled along the WHO Stop TB Strategy. The global targets of TB

control are to achieve at least 70% TB case detection rate and 85% treatment success rate (WHO, 2008)

Kenya has made good progress in the fight against TB; TB Case detection rate (CDR) reached the global target of 80% in 2007, the treatment success rate was 85% in 2007. By the end of 2008 83% of TB patients were tested for HIV against a national target of 80%. Despite this progress major problems still remain: low access to anti-retroviral treatment for HIV infected TB patients, significant delays in TB diagnosis which facilitates TB transmission and is associated with a higher frequency of the poor sequel of TB and the emerging problem of drug resistant tuberculosis. There are concerns that inadequate infection control measures in health care settings may be facilitating the transmission of TB to health care workers, patients and their visitors which has significant consequences for vulnerable groups such as HIV infected persons

The emergence of multidrug-resistant, Mycobacterium tuberculosis (MDR-TB) worldwide poses a serious problem to the treatment of tuberculosis. These MDR strains are at least resistant to two primary chemotherapeutic agents rifampicin and isoniazid (WHO, 1997) and require treatment with more costly, more toxic second-line drugs.

1.2 Statement of the Problem

The aim of this study is to model the trends of the treatment outcomes of smear positive TB cases notified in Kenya between the years 2002-2007 and its dynamics per province. The information generated will give insights into the rates of change of the treatment outcomes within the year and over the different years. The relative occurrence of treatment outcomes is important in understanding the occurrence of treatment outcomes in the country.

1.3 Objectives

1. To model the trends of treatment outcomes of smear positive TB patients over time.
2. To model the rate of occurrence of the treatment outcomes and factors associated with the occurrence of the outcomes.

1.4 Justification

This study seeks to understand the dynamics of treatment outcomes of the infectious forms of tuberculosis over time.

Given that treatment for TB has reduced from 8 to 6 months, understanding the rates of occurrence of treatment outcomes is important for management measures which are being adopted by different provinces in Kenya. Understanding the inter quarterly rates of occurrence is important since this will provide information on seasonal variations. This epidemiological analysis of tuberculosis treatment outcome will act as a tool for advising TB control policies in Kenya.

CHAPTER TWO

2.0 Literature Review

Cameron et.al (1998), Kleinbaum et.al (1998), describe Poisson regression analysis as a technique which allows to the modeling of dependent variables that describe count data It is often applied to study the occurrence of counts or events as a function of a set of predictor variables, in experimental and observational study in many disciplines, including Economy, Demography, criminology Psychology, Biology and Medicine Gardner et.al (1995). In particular, in the last twenty years, Poisson regression model has been extensively applied in many contexts in bio-medical studies, including Epidemiology, to investigate the occurrence of selected diseases in exposed and unexposed subjects in observational prospective studies, or to evaluate the clinical course of patients in experimental and observational investigations in human clinical setting.

Poisson regression has been used extensively in epidemiological studies Heinzl et.al (2005) used poisson regression to examine 5,255 still births among a total of 1,342,993 births in Austria. The aim of the study was to evaluate the prognostic effect of mothers' age on still birth outcome by applying poisson regression and further examining the use of pseudo R squared measure.

Chen (1988) used poisson regression model to analyse survival data and demonstrated how parsimonious model fitting, estimating and testing methods can be handled to analyse life table data with covariates.

Poisson regression and multinomial regression has also been applied in veterinary science to study the risk factors associated with tuberculosis in livestock In the first study, Andreasen *et al* (2001) carried out a prospective cohort study to evaluate the association between the time elapsed from seroconversion to slaughter and the extent of lung lesions in 830 pigs in Denmark. Examination of thoracic organs at slaughter is commonly used for surveillance and estimation of the economic burden or respiratory disease in pigs. Moreover, longitudinal serological testing allows to estimate at which age seroconversion occurs in herds infected by *Mycoplasma hyopneumoniae and*

Actinobacillus pleuropneumoniae and has been proposed to be associated with the severity of the disease in a previous investigation Sitjar et al (1996).

To evaluate the association between the extent of lung lesions at slaughter and the time elapsed since seroconversion, the Authors recruited 830 pigs from 8 herds and checked for seroconversion by ELISA test every 3 weeks until slaughter. The time elapsed from seroconversion to the death was categorized in 4 groups and considered as the main factor of interest (*e.g.*, similarly to an exposure factor in other cohort studies). The number of lung lesions was modelled by a log-linear Poisson regression model, including the main factor of interest and the following possible confounders: gender, age at slaughter and litter. Moreover, interaction terms between litter and any each other predictor were inserted into the model. A negative relationship between the time elapsed from seroconversion to *M. hyopneumoniae* and mycoplasma-like catharral pneumonia emerged in each herd. Such an association was expressed as mean change percent (MCP) for each 4-week interval elapsing from seroconversion. Such a parameter was easily estimated by the corresponding estimate of the coefficient $\beta 1$ in the Poisson regression model.

Roche et al. (2006), investigated the role of climatic factors on the risk of developing milk fever in grazing cows. Milk fever is a metabolic disorder due to a drop in blood calcium (Ca) concentration, especially occurring in the period between the late pregnancy and the early lactation. Apart from nutritional factors, which are known to be associated to the risk of developing this disease, many other variables, including climatic ones, are suspected to play some role on its incidence. Some indirect evidence came from study on metabolic disorders in sheep, whereas only few studies in bovine had been conducted before, and they did not include dairy cows. The Authors analyzed the relationship by the number of milk fever episodes and many climatic factors (relative humidity, screen temperature, soil and grass temperature, wind, sunlight, rainfall, etc.) by a log-linear Poisson regression model, using the logarithm of the number of days at risk as the offset. All the predictors were categorized on the quintiles distribution, except for rainfall which had a high frequency of 0 values and was

categorized using 0 rainfalls as the first category and other four strata based on the quartiles of the remaining data. Effect of variables recorded at individual (*i.e.*, the cow) level were analyzed by another model (Generalized Estimated Equation, GEE). Univariable Poisson regression analysis revealed an association between evaporation, difference between maximum and minimum screen temperature and rainfall in the previous 48h. Results from multivariable analyses confirmed these effects.

Travier et al (2003) investigated the risk of cancer among male veterinarians in Sweden. Veterinarians were exposed to many carcinogens (radiation, pesticides, anaesthetics, zoonotic viruses), but previous studies about the risk of this occupational group were rare, and based on mortality data only. The Authors identified a large cohort of male veterinarians or workers in veterinary industry, using data from the Swedish Cancer Environment Registry III, which collects information about two Swedish national population censuses (in 1960 and in 1970), the National Register of Causes of death and the Swedish Cancer Registry. The follow up period ranged between 1 January 1971 and 31 December 1989. A total of 1178 subjects (701 veterinarians and 477 other workers) were recruited and classified according the relevant code of occupation or industry. They were categorized into three groups: a) veterinarians working in the veterinary industry; b) veterinarians in other industries; c) workers other than veterinarians in the veterinary industry. The rest of Swedish male population, after the exclusion of subjects working at an extensive contact with animals (*i.e.*, breeders, hunters and butchers), was considered as the unexposed sub cohort. The number of incident cancer cases was modelled as a function of the exposure and of many confounders (namely, five-year age groups, four-year calendar periods, residence regions and urbanization level) by a log-linear Poisson regression model. A reanalysis restricting the unexposed group to persons with a high socio-economic level was also performed. An excess risk of melanoma was observed in all the three groups (RR = 2.77, 1.84 and 3.12 in group a, b and c, respectively). The first group experienced an excess risk of colon cancer (RR = 2.36, 95%CI:1.42-3.91), which was not observed in the other two groups, and an excess

risk of esophageal (RR = 3.78, 95%CI:1.42-10.09), pancreatic (RR = 2.10, 95%CI: 0.94-4.68), and brain (RR = 2.51, 95%CI: 1.04-6.03) cancer, which was present also in the group b, but not in the group c. Finally the group c (not veterinarians) experienced an excess risk of prostate cancer and, even if not significant, of oral and stomach cancer and of multiple myeloma. Analysis restricted to the high socio-economic level confirmed the excess risk for melanoma, esophageal, colon, pancreatic and brain cancer among veterinarians.

Ruger and Kim (2006) used multinomial regression to study global health inequalities in 29 countries and measuring mortality in children and adults in their study the authors carried out multivariate analyses using multinomial logistic regression. They stratified adult mortality into three levels per 1000: 80–250; 258–449; and 460–725. The strata for under-five mortality for 1000 were 3.9–60; 66–156; and 160–316. Multinomial logistic regression analyses were used to estimate associations between the mortality groups and independent variables; to avoid potential collinearity, they used the stepwise procedure of entering several variables at a time. Separate sets of models were estimated for under-five and adult mortality. In each case, the healthiest group, group 1, was the reference group. For each predictor in the model, they estimated one parameter that represented the effect of a 1-unit increase on the logit (log odds) scale (a 10-unit increase produced little change in the log odds). Their first model included only income, education, health expenditure, disease prevention and health risk factor variables. In the second model, they added indicators on the environment, economy, monetary, communication, and information technology. Multinomial regression was validated using the Wald test and log likelihood ratios. Two-tailed p values or 95% CIs are reported for all analyses.

Clerk et al (2009) in their study examining the demographic, health, and social characteristics of mobility device users in long-term care settings, they used multinomial logistic regression to examine the factors associated with the use of different mobility devices (cane, walker, or wheelchair). From their study they

presented logistic regression coefficients and odds ratios for the independent variables as they related to the use of a walker or cane (part A) and to the use of a wheelchair (part B). The need factors, in the form of underlying health problems and difficulty with activities of daily living were associated with an increased use of mobility devices (Model B). The odds of using any mobility device (walker, cane, or wheelchair) were more than 4 times higher among those with breathing difficulties (e.g., asthma, pneumonia, emphysema, bronchitis). Vision impairment (not relieved by glasses) was associated with an increased odds of using a wheelchair (but not a walker or cane). Fractures also increased the likelihood of using mobility devices,with the odds of using a walker/cane higher than that for wheelchair use (OR = 34.8 for walker or cane; OR = 12.4 for wheelchair). The strongest factor associated with mobility device use was self reported difficulty walking. The odds of using a walker or cane were almost 90 times higher among those who had difficulty walking compared to those who were independent in walking. The odds of using a wheelchair were 97 times higher for those who reported difficulty walking.

Schorr et al (2009) in their study examining the association between daily smokers' mental health according to the five-item Mental Health Inventory and the core constructs of the transtheoretical model (TTM): stage of change, processes of change, smoking cessation self-efficacy, and decisional balance they found out that Smokers with lower levels of mental health had increased odds to contemplate quitting within the next 6 months compared to not intending to quit at all. In addition, they reported an elevated use of change processes as well as an enhanced endorsement of positive and negative aspects of nonsmoking. However, in a subsample analysis performed on smokers in contemplation stage, low mental health was related to lower self-efficacy expectancies in negative affect situations.

CHAPTER 3: METHODOLOGY

3.1 Introduction

Modelling is basically about describing the relationship between a response variable and one or more other variables (explanatory variables) in a simplified way.

First you explore the data to detect patterns and relationships and you visualize the relationship by fitting a curve to the data. This is the empirical approach; you choose a model based on the data. Alternatively, if there exist a well-established theory, you choose the model based on the context and see if it fits the data well. You can use for instance, the deterministic growth models in biology, models of consumer behaviour in economics and the majority of models in physics and chemistry.

Regression analysis may be broadly defined as the analysis of relationships among variables. It is one of the most widely used statistical tools because it provides a simple method for establishing functional relationships among variables. The relationship is expressed in the form of an equation connecting the response or dependent variable y, and one or more variables, x_1, x_2,...,x_p. The equation , or, to be more precise, the regression equation takes the form

$$y = b_o + b_1 x_1 + b_2 x_2 + ... + b_p x_p, \qquad (1.0)$$

where $b_o, b_1, b_2, ..., b_p$, are called the regression coefficients, which are determined from the data. A regression equation containing only one independent variable is called a

simple regression equation. An equation containing more than one independent variable is referred to as a multiple regression.

When faced with a functional relationship problem, the estimation of the parameters becomes a primary problem.

The variance function for standard homoscedastic regression model is a constant. However for heteroscedastic regression models, variance is not constant. Heteroscedastic regression models are acceptable as appropriate in a wide variety of fields. Statisticians often look at residual plot to investigate whether the model is heteroscedastic or not and what the appropriate behaviour of the variance function

The linear regression line is probably the best known of all statistical models. It is also the easiest model for demonstrating the steps in the analysis of research data. During the phase of descriptive and exploratory analysis, for instance, you could visualize the data in a scatter plot

Poisson regression as a special case of the generalized linear model, whether characterized as a causal model or not. The Poisson formulation has obvious appeal, as it is relatively simple to interpret because the right hand side is the familiar linear combination of predictors and because when exponentiated, the regression coefficients can interpreted as multipliers. In addition, tests and regression diagnostics available for the normal regression special case carry over, at least in look and feel (Cook and Weisberg: chapter 22). The negative binomial distribution has been suggested by some as an alternative to the Poisson when there is evidence of "overdispersion" (Paternoster and Brame, 1997; Osgood, 2000). Stated loosely for the moment, "overdispersion" implies that there is more variability around the model's fitted values than is consistent with a Poisson formulation. The negative binomial is proposed as a means to correct for this problem, and some say that it automatically does so (Osgood, 2000). There is a parameter for the negative binomial distribution whose estimated value inflates the Poisson dispersion as needed.

The Poisson regression model may be used as an alternative to the Cox model for survival analysis, when hazard rates are approximately constant during the observation period and the risk of the event under study is small (*e.g.*, incidence of rare diseases). For example, in ecological investigations, where data are available only in an aggregated form (typically as a count), Poisson regression model usually replaces Cox model, which cannot be easily applied to aggregated data. Furthermore, using rates from an external population selected as a referent, Poisson regression model has often been applied to estimate standardized mortality and incidence ratios in cohort studies and in ecological investigations Breslow NE, Day NE. (1987) and Estève J et al (1994). Finally, some variants of the Poisson regression model have been proposed to take into account the extra-variability (overdispersion) observed in actual data, mainly due to the presence of spatial clusters or other sources of autocorrelation Cameron AC et.al (1998).

The importance of the cohort study in observational Epidemiology derives to the fact that such a study design represents the investigation which more resembles to an experiment Kleinbaum DG et.al (1987).

In a closed cohort design, the individuals are completely identified in a specific (ideal) instant time and the main factors of interest for the outcome (*e.g.*, cure, failure, treatment completion, death or out of control and different types of medical interventions in routine disease control) are usually measured at the same time. All the subjects are followed up for the same period. In a closed cohort design, the follow-up period usually corresponds to a calendar period or duration treatment.

In a closed cohort, each subject may leave the care only either because the end of the study itself or due to the occurrence of the event of interest. In such a situation, the probability of the occurrence of the event under investigation may be directly estimated at specific time points by the corresponding frequency observed. Such a probability, conditional to the observation time, represents a central estimator of event occurrence in

epidemiological investigations and it is called Risk. The ratio between such estimates is called Relative Risk (RR) and is employed as a measure of an association between the occurrence of an outcome and the probability of the taking medication in a given province. On average, RR will tend to 1 if no association between the exposure and the risk exists, and will be higher, the higher the effect of the exposure is. A value between 0 and 1 will be observed for factors inversely associated to the risk (which should be more correctly called "protective factors", instead of "exposures").

In an open cohort, a subject may withdraw before the end of follow up for factors other than the occurrence of the event under study, for example: migration to other areas (subject "lost to follow up") or death due to cause's different from the disease of interest. In this case the corresponding observation time is "censored" (and its occurrence is named "censoring").

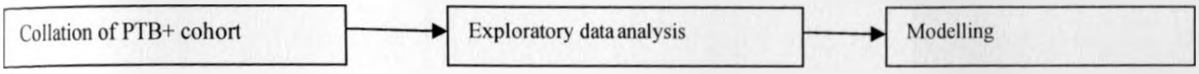
In general, the presence of censoring prevents from directly obtaining risk estimates, because the follow up time is different among the subjects that have not experienced the outcome. RR estimates may be obtained by some different ways in an open cohort study. One of the most important is by using another fundamental measure of occurrence in Epidemiology, *i.e.*, the rate.

In spite of its recent wide application, Poisson regression model remains partly poor known, especially if compared with other regression techniques, like linear, logistic and Cox regression models.

The aim of this study was to introduce the Poisson regression model in the framework of modeling outcomes of cohort of infectious tuberculosis (Smear positive pulmonary Tuberculosis cases) in Kenya in the years (2002-2007). Controlling for confounding and the assessment of interaction will be illustrated. Finally, applications of Poisson model in studying tuberculosis epidemiology will be presented and discussed.

The two methods that will be used are poisson regression and multinomial regression.

3.2 Research Design



3.3 Sampling Method

There was no sampling procedure used since the entire cohort data for the Tuberculosis new smear positive cases notified in Kenya in the years 2002-2007 was used.

The data used related to the counts of different treatment outcomes namely cure, treatment completed, failures, deaths, out of control and transfer outs. The count data was observed over the years 2002-2007.

3.4 Description of Methods

3.4.1 Poisson Regression

A poisson random variable Y has probability density function, $f(y) = P(Y = y)$ given as,

$$P(Y_i = y_i) = \frac{e^{-\lambda_i}\lambda^{y_i}}{y_i!} \quad y_i = 1, 2, 3...$$

Where Y_i are the outcomes namely cure, died, failure, treatment completed (TC), out of control (OOC) and transferred out (TO).

Where the parameter λ is the mean value of the random variable Y such that for the i^{th} outcome, $E(Y_i) = \lambda_i$ which is also referred to as the rate parameter.

Larger values of the mean parameter λ will produce higher number of outcome i.e. large counts.

The poisson random variable in this context relates to counts of pulmonary TB case notified after enrolment into treatment i.e. the observed TB cases for different treatment outcomes.

The response variable in this situation is a quantitative variable, but has the property that it is discrete, taking on only integer values. The basic idea for this model is that the predictor information is related to the rate or susceptibility of the response to increase or decrease in counts.

3.4.2 Poisson Regression Model

The basic model formulation is that the mean of the poisson random variable is a function of predictor information,

$$E(Y_i) = \mu_i = \lambda_i = \exp\{\alpha + \beta_1 x_{1i} + ... + \beta_p x_{pi}\}$$

Because the log of this function produces a linear combination of the predictors, this model is said to have a log link function; the function that links the mean to the linearized predictor is the log function.

Consider a simple model with a single predictor x, we have,

$$E(Y) = \mu = \lambda_i = \exp\{\alpha + \beta\ x_i\}$$

And this function can be re-written as $\exp\{\alpha\}(\exp\{\beta\})^x$.

When we consider a one unit increase in the predictor x we now have a mean function,

$$X_{x+1} = E(Y \mid X+1) = \exp\{\alpha\}(\exp\{\beta\})^{x+1} = \exp\{\alpha\}(\exp\{\beta\})^x(\exp\{\beta\}) = E(Y \mid X)(\exp\{\beta\}),$$

Thus $\exp\{\beta\}$ is the rate of change in the mean response per unit increase in the predictor.

When a response count Y has an index t like population size or some other risk measure then,

$$\log(\mu / t) = \log(\mu) - \log(t) = \alpha + \beta x$$

by moving the $\log(t)$ term to the right side of the equation, we get $\log(\mu) = \log(t) + \alpha + \beta x$, in this expression, the $\log(t)$ is known as the offset. This tells us that the mean is proportional to the index t. Thus for a fixed x, doubling the population size would double the response Y of say the number of deaths in a particular cohort.

As assumed for a poisson model our response variable was the observed count of each treatment outcome. It was assumed that the dependent variable is not over-dispersed and does not have an excessive number of zeros.

3.4.3 Goodness of fit tests

The statistics used in the interpretation of poisson model are as indicated below:

Log Likelihood (LR=-2logL) - This relates to the log likelihood of the fitted model. It is used in the calculation of the Likelihood Ratio (LR) chi-square test of whether all predictor variables' regression coefficients are simultaneously zero and in tests of nested models.

LR chi2 (α) - This is the LR test statistic for the omnibus test that at least one predictor variable regression coefficient is not equal to zero in the model. The degree of freedom (the number in parenthesis) of the LR test statistic is defined by the number of predictor variables (α). LR chi2 (α) is calculated as -2*[ll(null) - ll(model)].

Pseudo R2 - This is McFadden's pseudo R-squared. It is calculated as 1 - ll(model)/ll(null) Poisson regression does not have an equivalent to the R-squared found in OLS regression. There are a variety of pseudo-R-square statistics, it is important to note that this statistic does not have the same interpretation as the R-

square in OLS regression (the proportion of variance of the response variable explained by the predictors) interpretation of this statistic will be done with caution.

3.5 Multinomial Logistic Regression

Logistic regression can easily be extended to outcomes with more than two categories. Initially we consider outcome Y levels 0, 1, ..., J. We will consider Y=0 our referent or non-case group, and beyond we do not need to make any assumptions about order of severity for the remaining outcome categories. For simplicity as in our study we will assume that J=6 or a total of 6 possible outcomes namely cure, died, failure, treatment completed (TC), out of control (OOC) and transferred out (TO). The model is based on the generalized logit function. In our case it results in five equations. Suppose that X=Year, then for the respective logits for death and failure relative to cure are given as

$$g_1(x) = \log\left[\frac{p(y=1\mid x)}{p(y=0\mid x)}\right] = \beta_{01} + \beta_{11}X \qquad (1)$$

$$g_2(x) = \log\left[\frac{p(y=2\mid x)}{p(y=0\mid x)}\right] = \beta_{02} + \beta_{12}X \qquad (2)$$

Where $\beta_{01}, \beta_{11}, \beta_{02} and \beta_{12}$ are unknown parameters to be estimated, these parameters are interpreted as log of odds ratio such that $\exp^{(\beta_{11})}$ will be the change in risk for death relative to cure for unit change in year.

The conditional probabilities for each outcome category are:

$$p(y=0\mid x) = \frac{1}{1+\exp(g_1(x))+\exp(g_2(x))} \qquad (3)$$

$$p(y=1\mid x) = \frac{\exp(g_1(x))}{1+\exp(g_1(x))+\exp(g_2(x))} \qquad (4)$$

$$p(y=2\mid x) = \frac{\exp(g_2(x))}{1+\exp(g_1(x))+\exp(g_2(x))} \qquad (5)$$

CHAPTER 4: RESULTS

4.1 Introduction

In this chapter we give the results of the exploratory data analysis which are presented in form of tables and figures and multivariate data analysis which are presented in form of tables.

From the exploratory data analysis a total of 218,063 smears positive TB cases were evaluated. This data relates to cohorts notified between the years 2002-2007. There was however and increase in the number of cases notified and evaluated over the period under review but a decline was first seen in the year 2007.

4.2 Exploratory Data Analysis

Province	Cure	Died	Failure	TC	OOC	TO	Total
Central	16753	1022	27	1584	1506	1080	219
Coast	17051	1007	118	1945	2882	1605	246
Eastern North	4760	159	1	50	457	55	54
Eastern South	19184	1094	81	1271	2083	982	246
Nairobi	28958	1146	206	4488	5213	3022	430
North Eastern	1907	27	10	3	130	6	20
Nyanza	27872	3267	83	3311	7830	1956	443
Rift Valley North	11392	884	49	1368	1472	896	160
Rift Valley South	15157	1020	31	2330	3162	890	225
Western	9712	717	90	967	1215	519	132
Kenya	**152746**	**10343**	**696**	**17317**	**25950**	**11011**	**218**

Table1: Summary of Treatment outcomes for different provinces for the period: 2002-2007

From table 1 above it shows the distribution of treatment outcomes for the different provinces and shows that Nairobi, Nyanza, Eastern South and coast provinces had the highest burden of smear positive Tuberculosis with North Eastern and Eastern North having lower number Smear positive TB cases.

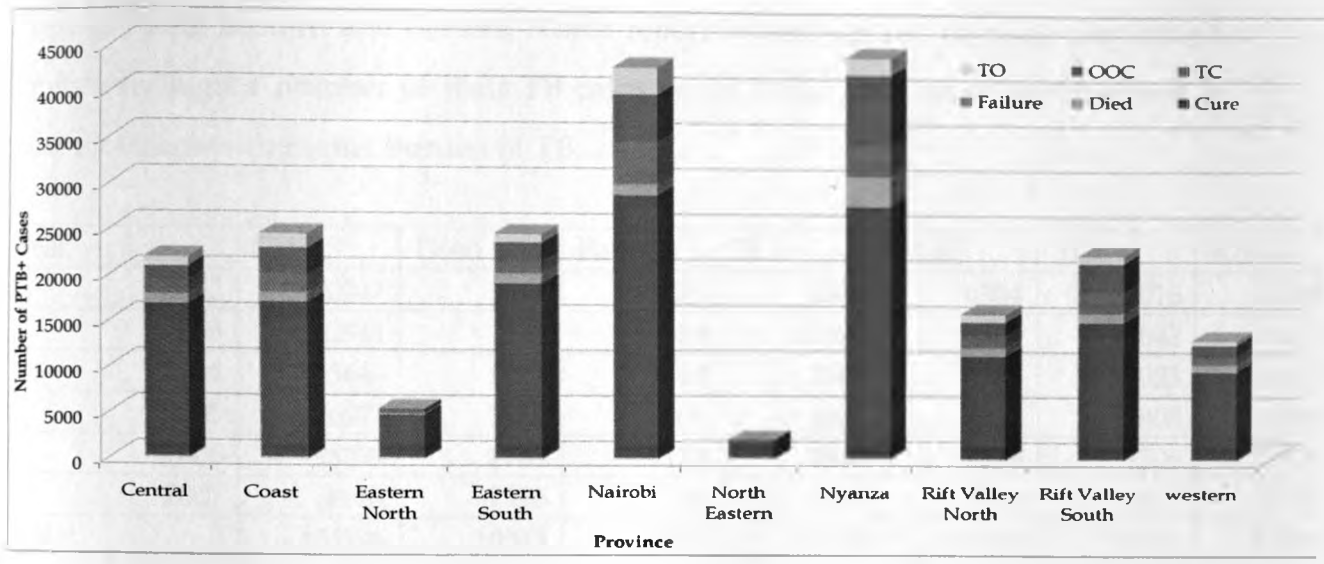


Figure2: Summary of treatment outcomes for different provinces

Figure 2 above shows where as Nairobi and Nyanza are able to cure most of the TB patients there is a significant number with adverse effects. While for Eastern North and North Eastern they are notifying significantly few numbers of smear positive TB patients.

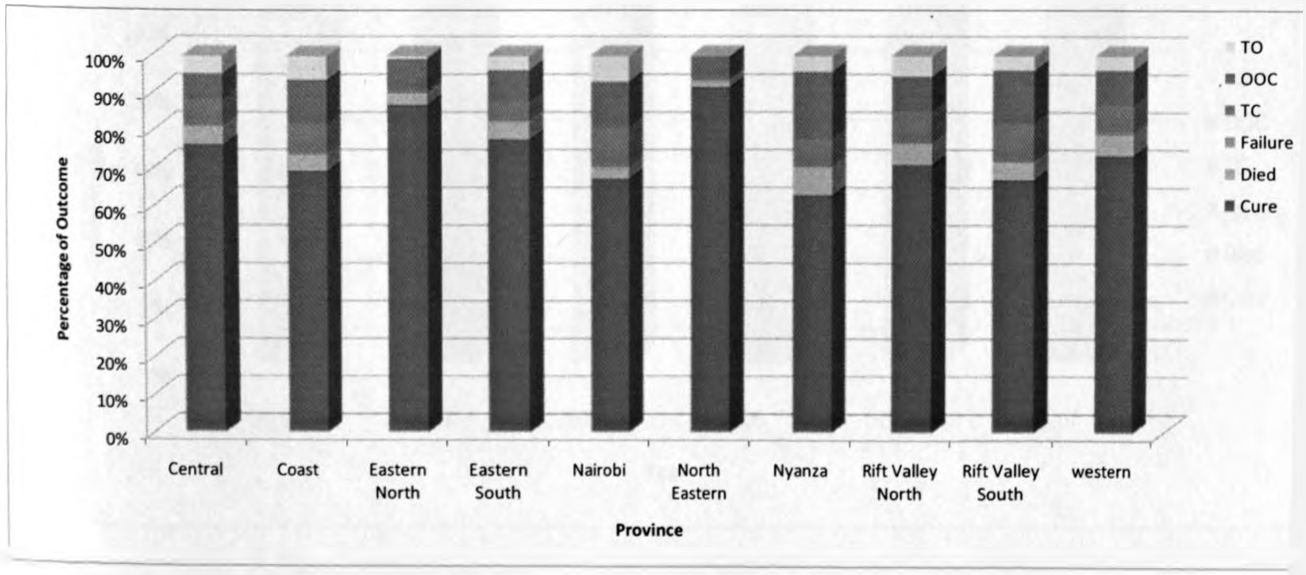


Figure3: Percentage of treatment outcomes for different provinces

Figure 3 above shows the proportion in terms of percentage of outcomes, it shows that where as North Eastern and Eastern North report lower number of cases they cure a significantly higher number of their TB cases while lower cure rates are observed in those provinces with higher burden of TB.

Year	Cure	Died	Failure	TC	OOC	TO	Kenya
2002	20247	1517	135	2746	4304	2016	30965
2003	22941	1725	79	2999	4282	2042	34068
2004	25644	1959	94	2747	4388	2023	36855
2005	26974	1982	105	3005	4332	1808	38206
2006	28279	1765	129	2808	4598	1706	39285
2007	28661	1395	154	3012	4046	1416	38684
Total	**152746**	**10343**	**696**	**17317**	**25950**	**11011**	**218063**

Table2: Summary of Treatment outcomes for different years for the period: 2002-2007

From Table 2 above it shows that the numbers of smear positive TB cases have been increasing but decline was first observed in the year 2007. Notable is that the number of cases cured have been on the increase which is indicative of good TB control efforts but worry is the increase in the number of failures .

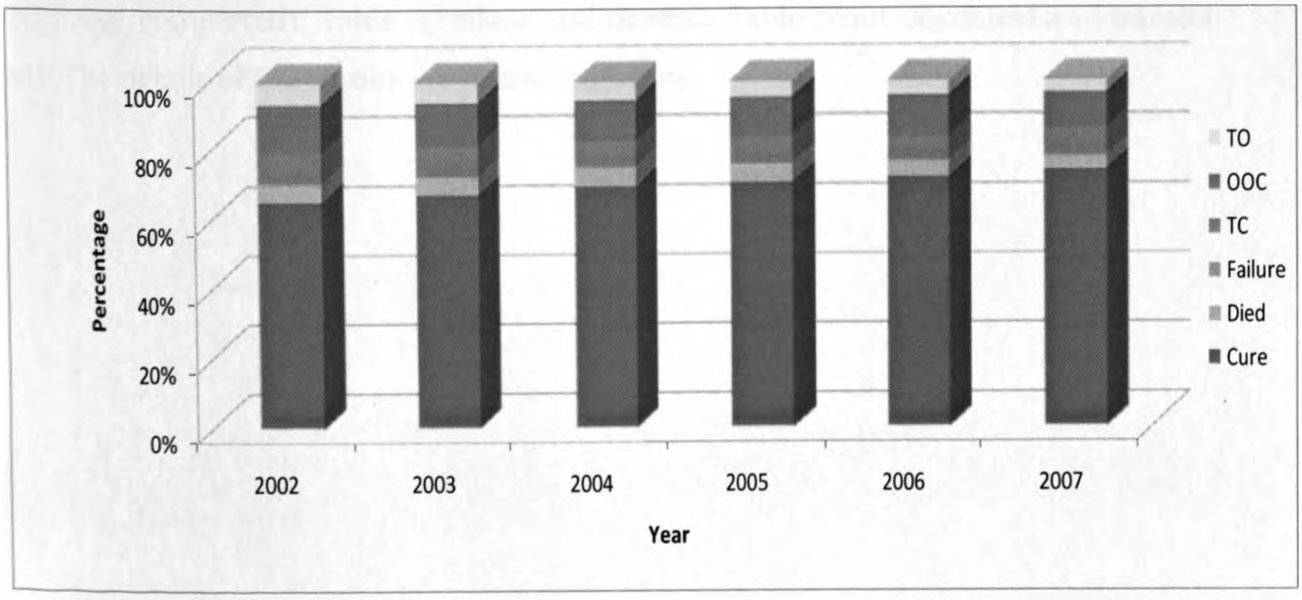


Figure4: Summary of treatment outcomes for years

From figure 4 above it shows that the numbers of TB patients have been on the rise through the years and importantly the numbers of deaths have continued to come down this is indicative of the quality of care that TB patients might be receiving.

4.3 POISSON REGRESSION MODELS

The poisson regression modelling was obtained using the counts of the outcomes notified from the years 2002 to 2007, in the poisson model the predictors included in the model were quarter, year and province. The reference predictors used were quarter one, year 2002 and Nairobi province, the interpretation is based on the exponential value of the coefficients and the 95% confidence interval of the parameter estimates, if the confidence interval contains unity then the parameter estimate is not significantly different from the reference predictor. The presentation of results was based on the different treatment outcomes namely cure, treatment completed, failures, deaths, out of controls and transfer outs.

The results of poisson regression models are presented in tables, Table 3b(cure and Treatment completed), Table 4(Failure and deaths), Table 5(out of control and transfer out). The details of the results are presented below.

	Cure			Treatment Complete(TC)		
	exp($\hat{\beta}$)	[95% Conf.Interval]		exp($\hat{\beta}$)	[95% Conf.Interval]	
Quarter						
Quarter 1	-	-	-	-	-	-
Quarter 2	1.003	(0.989,	1.017)	0.98	(0.947,	1.014)
Quarter 3	1.009	(0.995,	1.024)	0.994	(0.961,	1.028)
Quarter 4	0.995	(0.981,	1.009)	0.915	(0.884,	0.948)
Year						
2002	-	-	-	-	-	-
2003	1.133	(1.112,	1.155)	0.995	(0.954,	1.038)
2004	1.267	(1.243,	1.290)	1.02	(0.978,	1.063)
2005	1.332	(1.308,	1.357)	1.007	(0.965,	1.050)
2006	1.397	(1.372,	1.422)	1.068	(1.025,	1.114)
2007	1.415	(1.390,	1.441)	0.94	(0.900,	0.981)
Province						
Nairobi	-	-	-	-	-	-
Central	0.579	(0.568,	0.590)	0.289	(0.273,	0.306)
Coast	0.589	(0.578,	0.600)	0.553	(0.528,	0.579)
Eastern South	0.662	(0.650,	0.675)	0.4	(0.380,	0.420)
Eastern North	0.164	(0.159,	0.169)	0.088	(0.080,	0.096)
North Eastern	0.066	(0.063,	0.069)	0.025	(0.021,	0.030)
Nyanza	0.962	(0.947,	0.978)	1.502	(1.450,	1.556)
Rift Valley South	0.523	(0.513,	0.534)	0.607	(0.580,	0.634)
Rift Valley North	0.393	(0.385,	0.402)	0.282	(0.266,	0.299)
Western	0.335	(0.328,	0.343)	0.233	(0.219,	0.248)
Constant	957.95	(939.63,	976.63)	111.18	(106.427,	116.146)
			LR chi2(17) = 53603.78	LR chi2(17) = 18056.75		
			Prob > chi2 = 0.0000	Prob > chi2 = 0.0000		
			Pseudo R2 = 0.8581	Pseudo R2 = 0.7685		
			Log likelihood = -4430.5924	Log likelihood = -2719.4048		

Table3: Effects table for the rates of outcomes: Cure and Treatment completed

Poisson Model: Cure

From Table 3 the poisson regression model for cure; the rates of outcome for cure in all quarters were not significantly different when compared with quarter 1, whereas when compared across the years the rates of occurrence were different with the rates of occurrence increasing over the years and in the year 2007 41.5% better outcomes observed as compared to the year 2002. When outcome cure in Nairobi as reference was compared with central the rates of occurrence of cure in central was 42% lower whereas in North Eastern it was 93% while in western it was 64% lower and in Nyanza province it was 4% lower than the rates of occurrence of cure in Nairobi. The LR test statistic indicates that the regression coefficients in the model are not equal to zero hence the significance of the regression coefficients.

This model is quite indicative that it does not matter which time of the year (quarter) when treatment is given; the rates are the same and it clearly indicates that more TB patients are cured in Nairobi followed by Nyanza province. Despite the high incidence of the disease in these two provinces they still cure a significantly higher number of TB patients.

The model also indicates that more TB patients are since 2002 have subsequently been cured and this is indicative of the TB control and management initiatives that have been put in place by the government and partners in TB control.

Poisson Model: Treatment completed

From Table 3 the poisson regression model for treatment completion, the rates of occurrence of the outcome treatment completed were not significantly different for quarter 2&3 when compared with quarter one since the confidence interval includes unity whereas quarter 4 has a 8% lower rates of occurrence when compared with quarter 1,. Further, when compared across the years, the rates of occurrence are not different in the years 2003,2004 and 2005 when compared with the year 2002. When compared with 2002, the year 2007 had a 6% lower rates of occurrence of the outcome.

When outcome treatment completed in Nairobi as reference was compared with coast province, rates of occurrence in coast were lower than in Nairobi by 45% whereas in North Eastern it was 98% lower and 50% higher in Nyanza. The likelihood ratio statistic indicates that the regression parameters are different from zero. This model indicates that there are more TB patients completing treatment in Nairobi, Nyanza and Rift valley south. however this outcome needs to be minimized and outcomes transition to cure. The model indicates that the rates of occurrence of this outcome was 6% lower in 2007 compared to the year 2002, this would be good if the transition is towards cure as an outcome.

	Failure			Deaths		
	exp($\hat{\beta}$)	[95% Conf.Interval]		exp($\hat{\beta}$)	[95% Conf.Interval]	
Quarter						
Quarter 1	-	-	-	-	-	-
Quarter 2	0.734	(0.591,	0.912)	0.957	(0.906,	1.011)
Quarter 3	0.944	(0.774,	1.151)	0.967	(0.916,	1.021)
Quarter 4	0.828	(0.673,	1.019)	0.942	(0.892,	0.995)
Year						
2002	-	-	-	-	-	-
2003	0.585	(0.443,	0.772)	1.137	(1.061,	1.218)
2004	0.696	(0.535,	0.906)	1.291	(1.208,	1.381)
2005	0.778	(0.603,	1.004)	1.307	(1.222,	1.397)
2006	0.956	(0.751,	1.216)	1.163	(1.086,	1.246)
2007	1.127	(0.895,	1.421)	0.919	(0.855,	0.989)
Province						
Nairobi	-	-	-	-	-	-
Central	0.131	(0.088,	0.196)	0.892	(0.820,	0.970)
Coast	0.573	(0.457,	0.718)	0.879	(0.807,	0.956)
Eastern South	0.393	(0.304,	0.508)	0.955	(0.879,	1.037)
Eastern North	0.005	(0.001,	0.035)	0.139	(0.118,	0.164)
North Eastern	0.049	(0.026,	0.092)	0.024	(0.016,	0.035)
Nyanza	0.403	(0.312,	0.520)	2.851	(2.665,	3.049)
Rift Valley South	0.15	(0.103,	0.220)	0.89	(0.818,	0.968)
Rift Valley North	0.235	(0.172,	0.321)	0.771	(0.706,	0.842)
Western	0.431	(0.337,	0.553)	0.625	(0.570,	0.687)
Constant	3.809	(3.010,	4.819)	10.869	(10.022,11.787)	
	LR chi2(17) = 532.74			LR chi2(17) = 6337.34		
	Prob > chi2 = 0.0000			Prob > chi2 = 0.0000		
	Pseudo R2 = 0.3703			Pseudo R2 = 0.7775		
	Log likelihood = -452.99578			Log likelihood = -906.912		

Table4: Effects table for the rates of outcomes: Failure and Deaths

Poisson Model: Failures

This model relates the outcome failure which refers to those TB patients who are non responsive to the potent TB drugs. From Table 4: the poisson regression model for failure, the rates of outcome for failures in quarter 3 & 4 were not significantly different when compared with quarter 1.notwithstanding, quarter 2 rates of occurrence of failure is 26% lower when compared to quarter 1, whereas when compared across the years the rates of occurrence were different for the years 2003 and 2004 when compared to the year 2002 in fact 42% and 30% lower respectively but for the years 2005, 2006 and 2007 the rates of occurrence were not significantly different from the failures observed in the year 2002 although the rates of occurrence have been increasing over the years.

When outcome failure in Nairobi as reference was compared with coast the rates of occurrence of failure in coast was 43% lower whereas in North Eastern it was 95% while in western it was 67% lower and in Nyanza province it was 60% lower than the rates of occurrence of cure in Nairobi. The LR test statistic indicates that the regression coefficients in the model are not equal to zero hence the significance of the regression coefficients.

This model is quite indicative that it does matter which time of the year (quarter) when treatment is given with more failures likely to be seen during the second quarter of the year. From the model it indicates that you are more likely to see more failures in Nairobi, Nyanza and coast provinces while the North Eastern province and Eastern North have the lowest failures across the provinces.

The model also indicates that there has been increasing number of failures over the years though not significantly different from the failures in the year 2002. This is indicative of the TB control and management initiatives that have been put in place by the government and partners in TB control to contain the number of failures.

Poisson Model: Deaths

From Table 4 the poisson regression model for deaths the rates of outcome for deaths in quarters 2&3 were not significantly different when compared with quarter 1, whereas death rates occurrence was significantly different in quarter 4 compared to quarter 1. When compared across the years the rates of occurrence were different with the rates of occurrence being higher and increasing over the years and beginning to decline in the years 2006 and 2007 as compared to the year 2002. When outcome death in Nairobi as reference was compared with Eastern south there was no significant difference while the occurrence of deaths was 2.85 times higher in Nyanza compared to Nairobi. Significantly, the model indicates that occurrence of deaths is 98% lower in North Eastern and 86% lower in Eastern North when compared to Nairobi province.

The LR test statistic indicates that the regression coefficients in the model are not equal to zero hence the significance of the regression coefficients.

This model indicates that the death rates began to decline from the year 2006 from a high point 1.30 compared to the year 2002 and it further indicates that there are more deaths occurring in Nyanza province, central and coast compared to Nairobi. These provinces (findings) relates to the provinces with the highest burden of tuberculosis.

From the model there is need to find out why there are more deaths in the high burden provinces and what are the causes of the deaths.

	Out of control (OOC)			Transferred Out (TO)		
	exp($\hat{\beta}$)	[95% Conf.Interval]		exp($\hat{\beta}$)	[95% Conf.Interval]	
Quarter						
Quarter 1	-	-	-	-	-	-
Quarter 2	0.904	(0.866,	0.943)	0.989	(0.939,	1.042)
Quarter 3	0.957	(0.918,	0.997)	0.956	(0.907,	1.007)
Quarter 4	0.958	(0.919,	0.998)	0.891	(0.845,	0.940)
Year						
2002	-	-	-	-	-	-
2003	1.092	(1.037,	1.150)	1.013	(0.952,	1.077)
2004	1	(0.949,	1.055)	1.003	(0.943,	1.067)
2005	1.094	(1.039,	1.152)	0.897	(0.842,	0.956)
2006	1.023	(0.970,	1.078)	0.846	(0.793,	0.903)
2007	1.095	(1.040,	1.153)	0.703	(0.657,	0.753)
Province						
Nairobi	-	-	-	-	-	-
Central	0.353	(0.333,	0.374)	0.357	(0.333,	0.383)
Coast	0.433	(0.411,	0.457)	0.531	(0.500,	0.564)
Eastern South	0.283	(0.266,	0.301)	0.325	(0.302,	0.349)
Eastern North	0.011	(0.008,	0.015)	0.018	(0.014,	0.024)
North Eastern	0.001	(0.000,	0.002)	0.002	(0.001,	0.004)
Nyanza	0.738	(0.705,	0.772)	0.647	(0.611,	0.685)
Rift Valley South	0.519	(0.494,	0.546)	0.295	(0.273,	0.317)
Rift Valley North	0.304	(0.286,	0.323)	0.297	(0.275,	0.320)
Western	0.215	(0.201,	0.230)	0.172	(0.157,	0.189)
Constant	37.291	(35.417,	39.264)	24.037	(22.597,	25.569)
	LR chi2(17) = 11493.96			LR chi2(17) = 7565.36		
	Prob > chi2 = 0.0000			Prob > chi2 = 0.0000		
	Pseudo R2 = 0.8064			Pseudo R2 = 0.8025		
	Log likelihood = -1379.5441			Log likelihood = -931.1265		

Table5: **Effects table for the rates of outcomes: Out of Control (OOC) and Transferred Out (TO)**

Poisson Model for treatment outcome: out of Control (OOC)

From Table 5 : the poisson regression model for out of control, the rates of outcome for out of control in all quarters are significantly different when compared with quarter 1 with the rates being between being 5-10% lower compared to the quarter one, whereas

when compared across the years the rates of occurrence were not significantly different for the years 2004 and 2006 but significantly different for the other years with the rates of occurrence being 1.1 times higher in the year 2007 compared to the year 2002. When outcome out of control in Nairobi as reference was compared with other provinces most of the provinces were lower ranging between (36-100%) lower compared to Nairobi.

The LR test statistic indicates that the regression coefficients in the model are not equal to zero hence the significance of the regression coefficients.

Out of control being an adverse outcome and the need to reduce this to bare minimum, lower rates of out of control is indicative of a good TB control program being implemented.

The model also indicates that more TB patients were out of control in the year 2007 compared to the year 2002. This underscores the need for the government to find out why there were consistently higher rates of out of control being observed across all the years.

Poisson Model for treatment outcome: Transferred Out (TO)

From Table 5 the poisson regression model for transferred out the rates of outcome for transferred out in quarters 2&3 were not significantly different when compared with quarter 1 but quarter 4 was 11% lower compared to quarter 1, whereas when compared across the years the rates of occurrence were not significantly different for the years 2003 and 2004 while the rates of occurrence were different for the years 2005, 2006 and 2007 when compared with the year 2002, the rates of outcomes have been on the decline over the years 2005-2007.

When outcome transferred out with Nairobi as reference was compared with other provinces, the rates of occurrence of transferred out in central was 64% lower whereas in North Eastern it was 99% while in western it was 83% lower and in Nyanza province it was 35% lower than the rates of occurrence of transferred out in Nairobi. The LR test

statistic indicates that the regression coefficients in the model are not equal to zero hence the significance of the regression coefficients.

This model indicates that efforts have been made to contain this adverse effects over the years though efforts now need to be directed to Nairobi, Nyanza and coast provinces. Though these regions are majorly urbanized, there is need for the formulation of the strategies to bring down the high rates of transfer outs or mechanisms established to obtain the true status of the outcomes of these TB patients transferred out.

3.3 MULTINOMIAL LOGISTIC REGRESSION MODELS

In multinomial logistic regression the reference categories used were: base outcome counts of deaths, base year=2002 and the base quarter=Quarter 1.

The presentation of the discussion of the models will be based on the treatment outcomes cure, treatment completed, failures, transferred out and out of control. The odds of outcomes relative to outcome death are taken not to be significantly different if the 95% confidence interval includes unity. The outcome death is used as a reference point because it is an adverse outcome.

The results of multinomial logistic regression models are presented in tables, Table 6 (cure and Treatment completed), Table 7(Failure, out of control and transfer out). The details of the results are presented below.

	Cure			Treatment Completion		
	exp($\hat{\beta}$)	[95% Conf.Interval]		exp($\hat{\beta}$)	[95% Conf.Interval]	
Province						
Nairobi	-	-	-	-	-	-
Central	0.65	(0.596,	0.709)	0.325	(0.294,	0.360)
Coast	0.677	(0.620,	0.738)	0.631	(0.573,	0.695)
Eastern South	0.703	(0.646,	0.765)	0.42	(0.381,	0.463)
Eastern North	1.17	(0.989,	1.386)	0.634	(0.523,	0.767)
North Eastern	2.348	(1.597,	3.453)	0.982	(0.645,	1.496)
Nyanza	0.342	(0.319,	0.366)	0.531	(0.492,	0.572)
Rift Valley South	0.583	(0.534,	0.635)	0.679	(0.618,	0.748)
Rift Valley North	0.509	(0.465,	0.558)	0.368	(0.332,	0.409)
Western	0.536	(0.487,	0.591)	0.372	(0.332,	0.416)
Quarter						
Quarter 1	-	-	-	-	-	-
Quarter 2	1.077	(1.018,	1.140)	1.038	(0.973,	1.107)
Quarter 3	1.088	(1.028,	1.150)	1.049	(0.984,	1.118)
Quarter 4	1.106	(1.045,	1.170)	1.003	(0.940,	1.070)
Year						
2002	-	-	-	-	-	-
2003	1.002	(0.933,	1.077)	0.879	(0.811,	0.953)
2004	0.992	(0.925,	1.064)	0.793	(0.733,	0.859)
2005	1.026	(0.957,	1.100)	0.777	(0.717,	0.841)
2006	1.191	(1.109,	1.279)	0.918	(0.847,	0.995)
2007	1.458	(1.353,	1.573)	0.995	(0.914,	1.083)
Constant	21.469	(19.754,	23.333)	5.02	(4.579,	5.503)

Table6: Multinomial Regression Effects table for the outcomes: Cure and Treatment completed

Multinomial Logistic Model: Cure

From table 6, thus the odds cure of relative to death was 35% lower in central as compared to Nairobi. In Nyanza province the odds of cure relative to death was 66% lower than in Nairobi, in Eastern North the odds of cure relative to death as compared to Nairobi was not significantly different, while in North Eastern the odds of cure relative to death was 2.34 times less likely than in Nairobi implying that you are more

likely to see fewer deaths and more cures relative to Nairobi province. There was no significant difference in odds of cure relative to death across all the quarters when compared with quarter 1, whereas the odds of cure relative to death were significantly different in the years 2006 and 2007 with 2007 less likely to experience the outcome death by 46% relative to the year 2002.

From the model it shows that the odds of cure relative to death in the last two years was higher implying that indeed more cures and less deaths were observed.

Multinomial Logistic Model: Treatment Completion

From table 6, thus the odds of treatment completion relative to death ranged between 2-77% lower across provinces when compared with Nairobi with the odds of treatment completion relative to death being 58% lower in Eastern South as compared to Nairobi. In Nyanza province, the odds of treatment completion relative to death was 47% lower than in Nairobi. This implies that you are more likely to see fewer deaths and more TB patients in all the other provinces relative to Nairobi province. There was no significant difference in odds of treatment completion relative to death across all the quarters when compared with quarter 1, whereas the odds of cure relative to death were significantly different in all the years except the year 2007 when compared with the year 2002 this indicates that the odds of treatment completion relative to deaths was almost similar in the two years.

From the model it shows that the odds of treatment completion relative to death across the years was lower but kept increasing implying that the number of TB patients completing treatment has been on the rise.

	Failures			Out of Control (OOC)			Transferred Out (TO)		
	exp($\hat{\beta}$)	[95% Conf.Interval]		exp($\hat{\beta}$)	[95% Conf.Interval]		exp($\hat{\beta}$)	[95% Conf.Interval]	
Province									
Nairobi	-	-	-	-	-	-	-	-	-
Central	0.149	(0.099,	0.225)	0.197	(0.178,	0.218)	0.402	(0.361,	0.449)
Coast	0.666	(0.522,	0.848)	0.254	(0.231,	0.281)	0.602	(0.542,	0.668)
Eastern South	0.424	(0.324,	0.556)	0.15	(0.136,	0.166)	0.338	(0.303,	0.377)
Eastern North	0.035	(0.005,	0.251)	0.038	(0.027,	0.053)	0.133	(0.097,	0.183)
North Eastern	1.592	(0.754,	3.359)	0.009	(0.003,	0.029)	0.091	(0.037,	0.221)
Nyanza	0.146	(0.112,	0.190)	0.134	(0.124,	0.145)	0.227	(0.208,	0.248)
Rift Valley South	0.171	(0.116,	0.252)	0.296	(0.269,	0.326)	0.332	(0.297,	0.372)
Rift Valley North	0.309	(0.224,	0.428)	0.197	(0.177,	0.219)	0.389	(0.347,	0.437)
Western	0.703	(0.539,	0.917)	0.173	(0.154,	0.194)	0.273	(0.240,	0.312)
Quarter									
Quarter 1	-	-	-	-	-	-	-	-	-
Quarter 2	0.826	(0.659,	1.035)	0.522	(0.489,	0.559)	1.058	(0.980,	1.141)
Quarter 3	1.049	(0.853,	1.289)	0.531	(0.498,	0.567)	1.01	(0.936,	1.089)
Quarter 4	0.964	(0.777,	1.196)	0.576	(0.539,	0.616)	0.968	(0.896,	1.045)
Year									
2002	-	-	-	-	-	-	-	-	-
2003	0.515	(0.387,	0.686)	0.948	(0.869,	1.035)	0.9	(0.820,	0.987)
2004	0.557	(0.424,	0.731)	0.761	(0.698,	0.829)	0.78	(0.712,	0.855)
2005	0.618	(0.475,	0.804)	0.849	(0.780,	0.925)	0.699	(0.637,	0.767)
2006	0.86	(0.668,	1.107)	0.905	(0.829,	0.988)	0.76	(0.691,	0.835)
2007	1.143	(0.894,	1.460)	2.813	(2.582,	3.065)	0.735	(0.665,	0.812)
Constant	0.242	(0.189,	0.311)	10.278	(9.370,	11.275)	3.238	(2.923,	3.586)

Table7: Multinomial Regression Effects table for the outcomes: Failures, out of control (OOC) and Transferred Out (TO).

Multinomial Logistic Model: Failures

From table 7, the odds of failures relative to death were lower across all provinces except North Eastern province where the odds of failures relative to deaths was 1.59 times higher when compared to Nairobi. in central, the odds of failure relative to death

/as 85% lower as compared to Nairobi. In Nyanza province the odds of failure relative) death was 85% lower than in Nairobi while in coast the odds of failure relative to eath as compared to Nairobi was 37% lower.

here was no significant difference in odds of failure relative to death across all the uarters when compared with quarter 1, whereas the odds of failure relative to death vere significantly different only in the years 2003, 2004 and 2005. notable however was he rise in the odds of failures relative to deaths from the year 2002-2007.

rom the model, it shows that the odds of failures relative to death has been on an ıpward trend though relative to the year 2002 where the failures were higher and it)egan to decrease the numbers of failures have began to increase particularly in the last wo years this means that the actors in the TB control and management need to devise ways to contain the rise of the failures because this could be indicative of the failures of the potent TB drugs and will give rise to Multi drug resistant tuberculosis.

Multinomial Logistic Model: out of Control (OOC)

From table 7, thus the odds out of control of relative to death were 80% lower in central as compared to Nairobi. In Nyanza province the odds of out of control relative to death was 87% lower than in Nairobi, in Eastern North the odds of out of control relative to death as compared to Nairobi was 96%, while in North Eastern the odds of cure relative to death was 99% lower than in Nairobi. Across all the provinces though, there were lower odds of out of control relative to deaths indicating that Nairobi had the highest number of out of controls.

There was significant difference in odds of out of control relative to death across all the quarters when compared with quarter 1, whereas the odds of out of control relative to death were significantly different in the years 2004, 2005,2006 and 2007 with 2007 having 2.813 times higher odds of out of control relative to deaths.

From the model it presents very interesting findings that the odds out of controls relative to deaths have been rising and it would be desirable to reduce such adverse outcomes to bare minimum.

Multinomial Logistic Model: Transferred Out (TO)

From table 7, thus the odds of transferred out relative to death was lower across the provinces ranging from (40-91%) with North Eastern province registering 91% odds of odds transferred out relative to deaths and coast province being 40% lower as compared to Nairobi.

There was no significant difference in odds of out of control relative to death across all the quarters when compared with quarter 1, whereas the odds of odds transferred out relative to death were significantly different across all the years. Notable however, was the decline in the odds of odds transferred out relative to death across all the years.

From the model it shows that the odds of odds transferred out relative to death in the last two years has been on the decline implying that more deaths than transferred out are being observed. The challenge however is that the true treatment outcome need to be deduced for this category of TB patients.

CHAPTER 5 Discussion, Conclusion and Recommendation

5.1 Discussion

From the modelling framework undertaken in this study it provided insights and collaborative information about the provinces and across the years. From the results presented it is clear that remarkable efforts have been put in place to improve the treatment outcomes over the years, this has involved the change in policy initiatives to combat the seemingly ever rising cases of tuberculosis this have been fuelled by the HIV epidemic since half of the TB patients are HIV positive (DLTLD, 2008) this implies that unless concerted effort are undertaken to contain HIV higher numbers of TB cases are going to be continually observed.

The results presented collaborate the known high burden provinces but what was not clear initially what the fact that despite high burden they were number of TB patients cured was higher compared to other provinces. The high number of TB cases in these provinces is in part attributed to the high HIV prevalence and also it is in these provinces that some of the best facilities in the country are found. Notable also is that the burden of Tuberculosis is higher amongst the poor communities and this is due to the poor living conditions which provides avenues for spread of TB disease (DLTLD, 2008).

The findings in North Eastern and Eastern North indicate they cure a higher percentage of their patients they notify fewer number of patients, but they also have fewer adverse outcomes like deaths, out of control and transfer outs. It can be attributed that the number of cases notified are fewer because the HIV prevalence rate in the two provinces is low (DLTLD, 2008).

From the results of the Poisson model for cure indicates that more TB patients are since 2002 have subsequently been cured and this is indicative of the TB control and management initiatives that have been put in place by the government and partners in TB control.

The model for treatment complete also indicates that the rates of occurrence of this outcome was 6% lower in 2007 compared to the year 2002, this would be good if the transition is towards cure as an outcome.

This model indicates that the death rates began to decline from the year 2006 from a high point 1.30 compared to the year 2002 and it further indicates that there are more deaths occurring in Nyanza province, central and coast compared to Nairobi. These provinces (findings) relates to the provinces with the highest burden of tuberculosis.

From the model there is need to find out why there are more deaths in the high burden provinces and what are the causes of the deaths.

From the poisson model for deaths it indicates that the death rates began to decline from the year 2006 from a high point 1.30 compared to the year 2002 and it further indicates that there are more deaths occurring in Nyanza province, central and coast compared to Nairobi. These provinces (findings) relates to the provinces with the highest burden of tuberculosis.

It is important to find out why there are more deaths in the high burden provinces and what are the causes of the deaths.

When multinomial logit model for cure was considered it showed that the odds of cure relative to death in the last two years was higher implying that indeed more cures and less deaths were observed which confirms the findings from the poisson regression.

From the multinomial logit model for failures, it shows that the odds of failures relative to death has been on an upward trend though relative to the year 2002 where the failures were higher and it began to decrease the numbers of failures have began to increase particularly in the last two years this means that the actors in the TB control and management need to devise ways to contain the rise of the failures because this could be indicative of the failures of the potent TB drugs and will give rise to Multi drug resistant tuberculosis.

5.2 Conclusion

From the discussion above it is evident that Kenya has managed and continues to cure most of the infectious forms of Tuberculosis hence being able to contain the spread of the disease, it shows that it managed to post more cures in the year 2007 than ever before this could be indicative of the policy initiatives that have been put in place in the country to contain tuberculosis.

There has also been significant number of tuberculosis patients who have managed to successfully complete treatment. From the two models Poisson regression and Multinomial logistic regression new insights have been brought to the fore about the intra provincial disparities and it is evident which provinces have significantly higher caseload, despite the fact that they have high adverse outcomes they are curing significantly higher number of TB patients.

From the models, it is indicative that the numbers of those who die have been on a downward trend since the year 2005; however what needs to be established is the true cause of death. There is however concern in the high number of deaths in Nyanza province and urgent efforts needs to be made to establish what are the causes of death ; is it that patients present with advanced forms of tuberculosis?.

From the model the numbers of the out of control seems to have remained higher over the years but more importantly Nairobi, Nyanza, Rift valley South and coast provinces have continually reported higher rates of out of control.

There is thus a need for more concerted efforts to reduce the more adverse effects particularly the out of control and to track those patients who have been transferred out of a province to ensure that they have finished their medication if the transmission of tuberculosis is to be cut down.

Recommendation

he two models used in this study can be adopted by the country to obtain information n the rates of occurrence of the different outcomes across the years and provinces and an give insights into the need for new strategies and management practices.

here is an urgent need to assess what is the true burden of deaths due to tuberculosis ecause as it were the surveillance system is not sensitive enough to deduce the ifference. There is also an urgent need for the country to develop a surveillance system which will be able to capture the treatment outcomes of those patients who are ransferred out from one province to another.

There is also need for the country to undertake a study to understand why up to 6% of the TB patients default on treatment particularly in high burden provinces Nairobi, Nyanza and coast province.

And finally for better TB control and management there is need for the country to undertake a study on the risk factors attributed to Tuberculosis since the cases observed seems to be increasing over the years, once the risk factors have been identified targeted interventions can be formulated.

eferences

ndreasen M, Mousingt J, Thomsen LK. (2001) No simple association between time elapsed from seroconversion until slaughter and the extent of lung lesions in Danish swine. Preventive Veterinary Medicine, 52:147-61.

reslow NE, Day NE. (1987) Statistical Methods in Cancer Research, Vol. 2, The Design and Analysis of Cohort Studies (IARC Scientific Publications No. 82), Lyon (Fr), IARC.

Cameron AC, Trivedi PK. (1998) Regression analysis of count data. Econometrics Society Monographs No. 30. Cambridge University Press. Cambridge (UK).

Cameron AC, Trivedi PK. (1998) Regression analysis of count data. Econometrics Society Monographs No. 30. Cambridge University Press. Cambridge (UK).

Division of Leprosy, Tuberculosis and Lung Disease (DLTLD) (2008) Annual Report 2008. Nairobi: The Ministry of Public Health and Sanitation.

Estève J, Benhamou E, Raymond L. (1994) Descriptive Epidemiology. IARC Scientific Publications No 128, Lyon (Fr).

Gardner W, Mulvey EP, Shaw EC (1995). Regression analyses of counts and rates: Poisson, overdispersed Poisson and negative binomial models. Psychological Bulletin, 118(3):392-404.

Harald Heinz, Thomas Waldöh and Martina Mittlböck (2005). Careful use of pseudo R-squared measures in epidemiological studies.

Kleinbaum DG, Kupper LL, Morgentern H. (1987) Epidemiologic Research. John Wiley & Sons, Inc., New York.

Kleinbaum DG, Kupper LL, Muller KE, Nizam A. (1998) Applied regression analysis and other multivariable methods - Third Edition. Brooks/Cole Publishing Company, Duxbury Press, Pacic Grove (CA).

National AIDS and STI Control Programme, Ministry of Health, Indicator Survey 2007: Preliminary Report. Nairobi, Kenya.

Osgood, W. (2000) Poisson-based Regression Analysis of Aggregate Crime Rates." Journal of Quantitative Criminology 16: 21-43.

Paternoster, R., and R. Brame (1997) Multiple Routes to Delinquency? A Test of Developmental and General Theories of Crime," Criminology 35: 45-84.

Philippa Clarke, Pamela Chan, P. Lina Santaguida and Angela Colantonio: The Use of Mobility Devices Among Institutionalized Older Adults. *Journal of Aging Health* 2009. 21; 611

Qi-GUANG CHEN *(1988)*. The Analysis of Life Table and Follow-up Data with covariates Using Poisson Regression Model

Roche J.R., Berry D.P. (2006) Periparturient climatic, animal, and management factors influencing the incidence of milk fever in grazing systems. Journal of Dairy Science, 89(7): 2775-83.

Ruger J P, Kim and H-J (2006). Comparison Global health inequalities: an international. Journal of Epidemiology and Community Health 2006;60;928-936.

Schorr G, Ulbricht S, Baumeister SE, Rüge J, Grothues J, Rumpf HJ, John U, Meyer C. (2009). Mental Health and Readiness to Change Smoking Behavior in Daily Smoking Primary Care Patients. International Journal of Behavioural Medicine.

Sitjar M, Noyes EP, Simon X, Pijoan C. (1996) Relationship among seroconversion to *Mycoplasma hyopneumoniae*, lung lesions and production parameters in pigs. Swine Health Prod. 4(6):273-77.

The WHO/IUATLD Global Project on Anti-Tuberculosis Drug Resistance Surveillance. *Anti-Tuberculosis Drug Resistance in the World.* The WHO/IUATLD Global Project on Anti-tuberculosis Drug Resistance Surveillance 1994- 1997. Geneva: World Health Organization; 1997.

Travier N., Gridley G., Blair A., Dosemeci M., Boffetta P. (2003). Cancer incidence among male Swedish veterinarians and other workers of the veterinary industry: a record-linkage study. Cancer Causes & Control, 14(6): 587-93.

World Health Organization (2003) Treatment of Tuberculosis: Guidelines for National Programmes, 3rd ed. Geneva: WHO.

Appendix I: Syntax

Poisson Regression

xi: by outcome_, sort : poisson observat i.quarter i.year i.provinc1, exposure(outcome_)

Multinomial Regression

xi: mlogit outcome_ i.provinc1 i.quarter i.year [fweight = observant], base outcome(4)