UNIVERSITY OF NAIROBI

COLLEGE OF BIOLOGICAL AND PHYSICAL SCIENCES

SCHOOL OF MATHEMATICS

ANALYSIS OF TIME TO EVENT DATA TO ESTIMATE VACCINE EFFECTIVENESS

Geoffrey Odiwuor Arunga

I56/68510/2011

A PROJECT REPORT SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF A DEGREE OF MASTER OF SCIENCE IN BIOMETRY OF THE UNIVERSITY OF NAIROBI.

JULY 2013

DECLARATION

This research project is an original work and has never been presented for examination at any other learning institution in Kenya or elsewhere.

STUDENT

ARUNGA GEOFFREY ODIWUOR

SIGN: ---------- **DATE: ----------**

This project has been submitted for examination with the approval of the following as the University Supervisor.

SUPERVISOR

DR. NELSON OWUOR

SIGN: ---------- **DATE: ----------**

ACKNOWLEDGEMENT

My heartfelt gratitude to God almighty for his grace and love that has enabled me to see this project to the end; it would have not been possible without him. My most sincere and profound gratitude to my project supervisor Dr. Nelson Owuor for his nurture, guidance and support during the entire course of the project without which the success of the project would have not been possible.

I also give my heartfelt appreciation to my classmates, colleagues and lecturers for their encouragement and guidance during the entire course, which has helped to make all this possible.

DEDICATION

I dedicate this project and its success to my family for their unwavering and bountiful support both materially and otherwise during the duration of the course and especially their words of encouragement and guidance which have made this possible.

Table of Contents

Page Number

EXECUTIVE SUMMARY

There has been a continued undervaluation and underutilization of vaccination as a way of preventing vaccine preventable diseases and thus consequently reduce morbidity and mortality associated with the particular diseases. Currently vaccination is the most potent tool to prevent diseases and prevent deaths because there has been a dramatic decrease or wiping out altogether of diseases attributed to increased vaccination against the diseases. The undervaluation and underutilization of vaccines stems from the underestimation of the severity of the vaccine preventable diseases, underestimating the benefits of vaccination and concerns regarding the side effects of vaccines.

The undervaluation and underutilization of vaccines has resulted in a drop in the investment in research and development of new vaccines to combat diseases that are considered preventable in humans especially amongst the developing countries like Kenya; thus worsening an already worse situation. Though there are costs incurred when eliminating a disease through vaccination, there are great amounts of savings to be made if vaccination is undertaken to prevent a disease amongst a population. The savings made are counted in human lives saved, cost of vaccines not needed hence not bought and the cost of surveillance activities not carried out pertaining to the disease. A population can be free of a disease because of high or increasing vaccination rates against a particular disease, which can also result in the development of herd immunity within the population under consideration, where a member of the population, whether vaccinated or not vaccinated, is at a lowered risk of getting infected with a particular disease because there are fewer members of the population who are susceptible or infected due to increased vaccination within the population.

Typically vaccines contain the same antigens that cause diseases but the antigens in vaccines are either killed or greatly weakened. Vaccine antigens are not strong enough to cause disease but they are strong enough to make the immune system produce antibodies against them. Memory cells prevent re-infection when they encounter that disease again in the future. Through vaccination, humans develop immunity without suffering from the actual diseases that the vaccinations are designed to prevent.

There has been widespread interest in the level of protection offered by vaccines. This level can be assessed using two ways namely, estimates of efficacy from randomized control trials (RCTs)

and also as estimates of effectiveness from observational studies. Sometimes the efficacy estimates from the RCTs are higher than the effectiveness estimates from observational studies and this is mainly due to the fact that RCTs are conducted in controlled environments. Observational studies are normally more ethical and feasible to carry out as compared to RCTs and would thus be better in a population to evaluate the effectiveness of a vaccine. Vaccine effectiveness (VE) is a measure of how well the vaccine works to protect against infections and illness when they are used in routine circumstances in a population, and it is normally easily estimated using the observational studies which are conducted in community settings (populations) and the researchers involved have no control over those who chose to be vaccinated or not to be vaccinated during the study. Vaccine effectiveness studies are subject to various forms of bias, which are typically much, more than RCTs; thus caution should be exercised when analyzing and evaluating the results of these studies. Typically the three types of biases that affect the interpretation of results from VE studies are confounding bias, selection bias and information bias. Factors that raise or lower the apparent attack rates in either the vaccinated or non-vaccinated groups will bias the vaccine effectiveness estimates and should be considered and taken care of in the study design and analysis as any potential biases which are typically inherent in observational studies. Despite the above challenges of using observational studies to estimate the effect of vaccines, they are the best way of estimating the effects hence when one considers and takes care of the potential biases that may arise then the treatment effects (effectiveness) of the vaccines will be accurately estimated using observational studies.

From the results obtained in this study it's seen that those who got vaccinated had massively higher hazards as compared to those who were not vaccinated before the vaccination campaign and that might be the reason why they got vaccinated and thus consequently their hazards decreased greatly after the vaccination campaign. Hence it was observed that the vaccine offered some level of protection to those who were vaccinated but this wasn't statistically significant. Going forward there should be greater encouragement and research geared towards the use of observational studies to estimate effectiveness of vaccines, this would make this information readily available to the population.

With more accurate and available information on the efficacy, safety, potential side effects and effectiveness associated with a vaccine, a population would be more willing to consider

receiving the vaccination under consideration. Overall the role of vaccination in the greater realm of public health cannot be underestimated in the world, be it the developed or developing world.

CHAPTER ONE: INTRODUCTION

1.1 Background of the study

With the need for an effective way to prevent people from getting infected with diseases; vaccination has come up as the one of the most appropriate ways of achieving this. Vaccination is also a cost effective way of dealing with diseases as compared to treatment of those infected with the same diseases (vaccine-preventable diseases); hence dealing with the morbidity associated with the disease and the accompanying mortality. Though some vaccines are universally offered, the receipt of the vaccines is voluntary, and would thus depend on the attitude and perception of the population considered eligible for vaccination towards the vaccination. It has been noted, that there has been an undervaluation and underutilization of vaccines stemming from the lack of information regarding the efficacy, safety, the side effects and the effectiveness associated with the vaccines under consideration.

There have been two ways to estimate the effects of a vaccine; this could by obtaining the estimates of the efficacy from randomized control trials (RCTs) and also estimates of effectiveness from observational studies. Sometimes the efficacy estimates from the RCTs are higher than the effectiveness estimates from observational studies and this is mainly due to the fact that RCTs are conducted in controlled environments unlike observational studies that are conducted within a typical population setting where there is no control from the research regarding the settings. RCTs are typically carried out during the stage before licensing of the vaccines but once the vaccine has been licensed then the vaccination with the vaccine has to be voluntary and thus observational studies would be ideal in the post licensing stage for a vaccine. Observational studies to assess effectiveness of vaccines are normally affected by various forms of bias, which are typically much, more than RCTs; thus caution should be exercised when analyzing and evaluating the results of these studies. Typically the three types of biases that affect the interpretation of results from VE studies are confounding bias, selection bias and information bias. When in a typical population setting to estimate the effects of a vaccine then

the observational studies would be better suited than RCTS because they are more ethical and feasible.

Thus the proper utilization of vaccines would enable the greater public health aim of decreasing or eliminating diseases to become a reality. Thus great savings would be made in terms of human lives saved, cost of vaccines not needed hence not bought and the cost of surveillance activities not carried out regarding the particular disease. Hence clearly vaccination is a public health imperative and its role in the greater realm of public health cannot be underestimated in the world, be it the developed or developing world.

1.2 Statement of problem

The estimation of the effectiveness of vaccines has been a challenge especially when done using observational studies. While randomized control trials are able to give more accurate estimates of the effect of a treatment (vaccine) due to the elimination of bias and confounding, they are only permissible before the licensing of a vaccine, to test efficacy, once this has been done then only the effectiveness of the vaccine can be estimated. Hence the effectiveness can only be estimated using other types of studies apart from randomized control trials (experimental trials), this is mainly observational studies. To make observational studies more accurate in the estimation of the effects of the vaccine, minimization of the different forms of bias that would arise during the course of the study would be essential. With more accurate results from the observational studies then we would have a quality way of assessing the effectiveness of a vaccine in a population.

Hence if time to event data (survival data) analysis techniques can be used to estimate the effectiveness of a vaccine using data from an observational study; we would be able to create another avenue through which vaccine effectiveness could be accurately estimated.

1.3 Study objectives

Compare the survival distribution functions for both the fully vaccinated and unvaccinated groups before the vaccination campaign.

Compare the survival distribution functions for both the fully vaccinated and unvaccinated groups after the vaccination campaign.

Estimate the effectiveness of the vaccine using time to event data analysis

1.4 Justification of the study

Vaccination over the last two centuries has been the most effective way of preventing diseases and thus minimizing the morbidity and mortality associated with the diseases. In spite of this, vaccination as a tool has been undervalued and underutilized, this has stemmed from the apathy towards it by the populations who are supposed to get vaccinated; this has been mainly because of the lack of accurate or easily available information on the safety, efficacy, side effects and effectiveness of the vaccines. Thus remedying this and making the above information available would have more people willing to get vaccinated, and with more getting vaccinated then herd immunity effect would be achieved and thus make even the ones who are not vaccinated less susceptible to infections due to reduced numbers of susceptible and infections.

Observational studies are the most ethical and feasible way to estimate treatment effects in a population setting for a vaccine that has already been licensed like in the case of the vaccine in the study. Data from observational studies are easily available as opposed to those from randomized control trials hence their utilization as an avenue through which the estimation of the effectiveness of the vaccine can be estimated, would be prudent going forward in vaccine effectiveness studies. Though they are biases and challenges inherent in using observational studies, thess can be dealt with through the proper set up of the study by minimizing the biases that may crop up and also the proper analysis of the data. This would produce quality and reliable information about the estimates of effectiveness of a vaccine.

Thus with information on the safety, efficacy, side effects and effectiveness of the vaccines, then more members of a population would be more than willing to get vaccinated and this would lead

to increased vaccination coverage. Thus the proper utilization of vaccines would enable the greater health imperative of preventing diseases and thus minimizing morbidity and mortality associated with the diseases (vaccine preventable diseases).

Thus this study that applies time to event data analysis on data collected from an observational study to assess (estimate) effectiveness of a vaccine would enable the above and thus would justifiable and necessary to conduct for the benefit of public health.

CHAPTER TWO: LITERATURE REVIEW

2.1 Literature review

Vaccination is currently the most effective way of preventing diseases, and there has been a dramatic decrease or wiping out of diseases altogether due to the inoculations given as part of a vaccination campaign. In the twentieth century there have been many new vaccines developed and under development which have impacted greatly on the occurrence of diseases. The role of vaccination in the greater realm of public health cannot be underestimated in the world, be it the developed or developing world. The Centers for Disease Control and Prevention declared vaccinations to be one of the ten greatest achievements in the field of public health during the twentieth century, because vaccines have spared millions of people from infectious diseases.

Vaccines have continued to be undervalued and most vaccine-preventable diseases still pose a threat to the human population [1]. Moreover given the undervaluation there may be room for a drop in investment in research and development of new vaccines to combat diseases that are considered preventable in humans especially amongst the developing countries like Kenya. These diseases include but are not limited to the "big five" namely: diarrheal diseases, malaria, tuberculosis, pneumonia and HIV/AIDS. The undervaluation and underutilization of vaccines stems from the underestimation of the severity of the vaccine preventable diseases, underestimating the benefits of vaccination and concerns regarding the side effects of vaccines.

Generally when there is no longer an imminent fear of contracting a disease, the public may forget about the limitations of cures and can become apathetic about prevention. Disease severity, vaccine availability and the quality of vaccination programs differ greatly throughout the world. A child in a developing country is at a higher risk of dying of a disease preventable disease as compared to a child in an industrialized country. Vaccination is a community (population) activity because when one person get vaccinated it can lead to the protection of the entire community and from the effect of migration between boundaries, countries and continents there may be a lasting effect globally. High vaccination rates in one country may benefit the neighbouring countries, high rates in one generation may benefit the next generation. Hence clearly high vaccination rates benefit all as the spread of infection drops.

Though there are costs incurred when eliminating a disease through vaccination, there are great amounts of savings to be made if vaccination is undertaken to prevent a disease amongst the population. The savings made are counted in human lives saved, cost of vaccines not needed hence not bought and the cost of surveillance activities not carried out. By 2002, the annual savings as a result of vaccination campaigns such as the smallpox one have resulted in the prevention of 350 million new infections and some 40 million deaths from the disease worldwide. The benefits of getting the measles vaccine are also not to be underestimated as the cost of treating a measles infection is 23 times the cost of getting the measles vaccine. Measles still remains one of the most contagious diseases and a major killer of children in developing countries. There is an enormous economic burden of failing to vaccinate against vaccine-preventable diseases. The reasons why developing countries like Kenya should be investing in immunization programs are because they are low-risk investments with a great payoff and impact on the population. They are highly cost-effective, have economies of scale and can thus be sustained by developing countries.

A population can be free of a disease because of high vaccination rates and increased vaccination rates against a particular disease can result in the development of herd immunity within the population under consideration, where a member of the population is at a lowered risk of getting infected with a particular disease because there are fewer members of the population who are susceptible or infected due to increased vaccination within the population [2]. While the unvaccinated in such a case may be said to also enjoy the benefit of also being spared the vaccine by avoidance of the adverse effects associated with the vaccine. When the number of the unvaccinated is small then the herd immunity is not compromised, but as more and more people in the population choose not be vaccinated the effect of the herd immunity diminishes, consequently again there shall be need for people to get vaccinated so as to avoid being susceptible to the disease and then getting infected. Hence the herd immunity's special aura emanates from the extension of protection impacted by an immunization program beyond the vaccinated individuals and to the unvaccinated individuals in the same population. Herd immunity as a term refers to "The resistance of a group to attack by a disease when a large proportion of the members of the population are immune, thus lessening the possibility of a patient with a disease coming into contact with a susceptible individual". Herd immunity is achieved through sustained vaccination campaigns in a population. If a population's vaccination

rates drop below a certain threshold of the herd immunity, then widespread disease outbreaks can occur. Hence if the vaccination rates drop significantly below the herd immunity threshold, the level of the population protection may not be enough to prevent the levels of infections increasing within the community due to increased number of susceptible and infected.

Mass vaccination campaigns allowed health authorities to achieve herd immunity; they believed in the development of "herd immunity". In diseases that can be passed from person to person, it is more difficult to pass the diseases on when there are those who are immune to it. The more immune individuals there are, then the less likely it is that a susceptible person will come into contact with someone who has the disease; high vaccination coverage is needed to merit herd immunity threshold [2]. Vaccination still remains one of the few preventive health measures that directly save money, in developing countries the annual cost of immunizing millions of children against the six infectious diseases is equivalent to the cost of a single day of health care in the U.S.A. Thus vaccines are unquestionably one of the most cost effective public health measures available yet they are undervalued and underestimated throughout the world. It's thus imperative that international agencies, governments, and health policy makers keep this preventive measure in the spotlight. Ultimately, it is the global society and future generations that benefit when all countries make the effort to protect their population from vaccine preventable diseases. Vaccines are less profitable than medicines, and pharmaceutical firms understandably have been apprehensive about making investments in the research and development of vaccines against diseases. The main reasons for that is the unprofitability associated with manufacturing vaccines. This unprofitability is due to the inability of the manufacturers to predict the demand for the vaccines; vaccines are normally bought dependent on the status of the disease severity and prevalence and the attitudes of the population in general about vaccination particularly its need, efficacy, safety and effectiveness. Hence when the demand is unknown the manufacturer will produce a small batch of vaccines to ensure there is reduced loss if the demand doesn't suffice. The small batch sizes make the cost per unit to be quite high hence not viable. A remedy to this would be the more recent approach of guaranteeing a large market and a reasonable price in advance to pharmaceutical firms thus encouraging them to produce the vaccines that are needed by nations in the world. WHO has played a critical role in managing the vaccination activities worldwide; to achieve this it has worked with governments, UN agencies, development partners, and other international organizations. Its main activities have been centered on support and

facilitation of research and development, ensuring the quality and safety of vaccines, developing policies and strategies for maximizing the use of vaccines, reducing financial and technical barriers to the introduction of vaccines and technologies, and also supporting countries in acquiring the skills and information needed to achieve disease control and eradication.

In 1796 Edward Jenner demonstrated that immunity to small pox could be produced by inoculating a human with a material from a lesion on the udder of a cow. This infectious material was named vaccine and the procedure named vaccination by Jenner. By 2006, there were vaccines available for a myriad of infectious diseases namely Measles, Rubella, Mumps, Diphtheria, Tetanus, Pertussis, Polio, Hib (Infant), Hepatitis B, Varicella, Pneumococcal disease, Influenza, Meningococcal diseases, Hepatitis A and Rotavirus A. Numerous vaccines are under development with major potential for impact by improving health in developing countries and in the world as a whole. Vaccines function by providing the immune system with harmless copies of an antigen. An antigen is a portion of the surface of the bacterium or virus that the immune system recognizes as foreign. An antigen often plays a role of in curing disease, i.e. by enabling a virus or bacterium to attach to cells. A vaccine may also provide a non-poisonous version of a toxin-(a poison produced by a bacterium) so that the body can derive a defense against it. Once an antigen is detected by the immune system, white blood cells called B-lymphocytes create a protein called an antibody that is precisely designed to attach to that antigen. Many copies of this antibody are produced and if a true infection of the same disease occurs, still more antibodies are created and as they attach to their target they may block the activity of the virus or bacterial strain directly, thus curbing infection. Once the antibodies are in place they make it easier for other components of the immune system to recognize and destroy the invading agents. The immune system is designed to remember, thus once it has been exposed to a particular bacterium or virus it retains immunity against it for years, decades and even at times for life. Consequently when there are later infections the immune system will be able to defeat these infections and to do so quickly. This ability and the speed with which the immune system will be able to do its work is a huge benefit, but it is due to the induced immunity from the vaccination; because for a body exposed for the first time to a germ, the body may need from 7 to 12 days, which may be too long and the body may be infected by that germ and even it may lead to a death. Humans are born with an immune system composed of cells, glands, organs, and fluids located throughout the body. The immune system recognizes germs that enter the body as “foreign invaders” or

antigens, and produce proteins substances called antibodies to fight them. A normal, healthy immune system can produce millions of antibodies to aid in the fight against germs during the thousands of attacks the body will face every day. Antibodies often disappear once they have destroyed the invading antigens, but the cells involved in antibody production remain in and become “memory cells”. Memory cells remember those original antigens and will always defend the body in case of an attack to the body by the same antigens in attempt to infect the person again; this protection is known as immunity.

Normally vaccines contain the same antigens that cause diseases but the antigens in vaccines are either killed or greatly weakened. Vaccine antigens are not strong enough to cause disease but they are strong enough to make the immune system produce antibodies against them. Memory cells prevent re-infection when they encounter that disease again in the future. Through vaccination, humans develop immunity without suffering from the actual diseases that the vaccinations are designed to prevent.

Vaccines used routinely for immunization are very effective in preventing diseases, although no vaccine can obtain 100% effectiveness. A single dose of a vaccine generally gives increased immunity though it is not recommended as one should get the full dose as recommended. Vaccines generally are safe though they have side effects they are a bit milder especially as compared to the effects of the diseases they are designed to prevent. Parental attitudes about vaccinations affect their willingness to have their children vaccinated. Some of the concerns that parents have are the efficacy, safety and effectiveness of the vaccines, worry about the side effects of vaccines and the need to get vaccinated as a whole. Unvaccinated members of a population pose a threat to other members of the population, and most importantly with dire consequences for members of the population who are unable to get vaccinated due to one reason or another. Vaccines protect against diseases by inducing immunity, and they have been administered across the world to much gain. Vaccination campaigns work on the idea that it is better to keep people protected and from falling ill than it is to treat them once they are ill.

Vaccination is a way of inducing immunity against a disease in a given population. Hence vaccination becomes a public health imperative and is thus an important component of public health because it prevents many people from becoming sick with a communicable disease, consequently reducing the risk to themselves and others in the same population.

There are studies that are used to determine how well a vaccine works; these are namely the randomized control trials (RCTs) and observational studies. The RCTs are used to measure the vaccine efficacy by comparing the frequency of infections in the vaccinated and unvaccinated groups. RCTs are normally required before a new vaccine is licensed for routine use by a national regulatory authority. The observational studies are used to measure the vaccine effectiveness by comparing the frequency of infections among the vaccinated and unvaccinated groups while controlling for other factors that differ across the two groups and may affect the results.

RCTs are a gold standard for estimating treatment effects because they minimize the potential for confounding, co-interventions, and thus bias, thus maximizing the strength of the causal inference, typically they are double blinded to enable the study be unbiased. While observational studies like the one described in this paper are not comparable to randomized blinded trials, they are more feasible and ethical. Hence for feasibility the observational studies are easier to carry out as compared to the RCTs because one does not require a lot of preparation to conduct it. While for ethical reasons the observational studies are more preferred because all participants in the study have a chance to get the vaccine if they desire to, which is not the case in RCTs where participants are randomly assigned to treatments arms. During a RCT, to minimize bias and thus invalid results, the investigator recruits participants into the study who have predetermined characteristics they then administer the treatment(s) and collect data on the participants' health for a defined time period. These participants are normally volunteers and they are not paid anything to participate in the study. The data collected includes measurements like vital signs, concentration of the study drug in the blood, and whether the patient's health has improved or not. The data is then analyzed using statistical tests to be able to determine the effects of the treatment. In an observational study the investigators observe the participants and take measurements on outcomes and other characteristics but they do not manage the study while for a randomized control trial the investigators give the study participants a particular intervention usually they compare the treated subjects to subjects who did not receive the treatment or received a standard treatment; then the researcher measures how the health of the study participants changes.

There has been widespread interest in the level of protection offered by vaccines. This level can be assessed using two ways; estimates of efficacy from randomized control trials and also as estimates of effectiveness from observational studies. Sometimes the efficacy estimates from the RCTs are higher than the effectiveness estimates from observational studies and this is mainly due to the fact that RCTs are conducted in controlled environments.

Factors that may results in the differences in efficacy estimates from RCTs and observational studies may be the study design, different age or risk groups being enrolled for the study, waning immunity within a season or between seasons, the vaccine strain and circulating virus strains and the effect of repeated vaccination [3]. The estimates of how well a vaccine works can be obtained by comparing the proportion of infections in the fully vaccinated and unvaccinated groups. The twin words efficacy and effectiveness are differentiated by the time and scope of their application, while the former is the prevention of infections among vaccinated persons in controlled trials and essentially ideal settings the latter is the prevention of infection among vaccinated populations. Vaccine efficacy and effectiveness studies have used different possible outcomes such as laboratory confirmed-virus infections. Vaccine effectiveness studies are normally more ethical and feasible to carry out as compared to RCTs and would thus be better in a population to evaluate the effectiveness of a vaccine. Vaccine effectiveness (VE) is measure of how well the vaccine works to protect against infections and illness when they are used in routine circumstances in a population, and it is normally easily estimated using observational studies which are not conducted in community settings (populations) and the researchers have no control over those who chose to be vaccinated or not to be vaccinated. Vaccine effectiveness studies are subject to various forms of bias, which are typically much, more than RCTs; thus caution should be exercised when analyzing and evaluating the results of these studies. Typically the three types of biases that affect the interpretation of results from VE studies are confounding bias, selection bias and information bias [5, 8]. The confounding bias occurs when the effect of vaccination on the risk of the outcome of interest is distracted by other factors associated with both vaccination and disease infection. Most likely those who developed the outcome of interest had an underlying medical condition and thus were more likely to get vaccinated to prevent the exacerbation of their condition. This may lead to an estimate of effectiveness of a vaccine that is too low, due to confounding bias. The selection bias arises from the errors introduced by the differences between those people who were enrolled into the study compared with the people

who are not enrolled. Typically people who'd usually get vaccinated are those who seek healthcare sooner routinely and also strive to live a healthy lifestyle as compared to those who do not participate in the studies. Thus the participants who got vaccinated (participated in the study) cannot be representative of the general population and the study results may be biased towards finding higher vaccine effectiveness [6]. The last form of bias, the information bias emanates from the quality or accuracy of measuring vaccination status or the outcome of interest (disease infection) in the two groups being compared in the study. Hence during the design of studies to observe the effectiveness of vaccines there should be tailoring to minimize the above biases. There is another form of bias that emanates from the biases above, this is the healthy vaccinee bias which typically afflicts most observational studies; it is normally defined as the systematic willingness to get vaccinated by healthy study participants while the frail and elderly will tend to be averse to getting vaccinated [7]. Though some vaccines are universally offered, the receipt of the vaccines is voluntary and thus may be preferentially sought by healthier individuals this is known as the healthy adherer effect. If not measured and not taken for care of this kind of bias or confounding by functional status could well yield a bias large enough to account for estimates in existing studies and thus produce artifactual vaccine effectiveness. This healthy vaccinee bias can be overcome if relevant confounders are measurable and measured well. Primary methods to overcome these biases would be, obtaining more accurate information on confounders, including functional status and then evaluating the extent to which the bias has been reduced by the methods selected.

Typically the outcomes measured in a VE study affect the interpretation of the VE estimates, these outcomes may include laboratory confirmed disease infection or preventing a medical attended infection [5]. Generally the more specific the outcomes in a VE study, the more accurate the measurement of the effect of the vaccination, but when less specific outcomes are used then the vaccine effect estimates (vaccine effectiveness) are lower. Vaccine effectiveness estimates are typically more dependent on the vaccine strain developed matching the circulating strain in the general population; thus vaccine relatedness and vaccine effectiveness can support real-time risk communication and mitigation.

Observational methods are important in the measurements of vaccine effectiveness as randomized control trials (RCTs) cannot be used for measurement of vaccines already on the

vaccination schedule [8]. Normally at times, the efficacy measured in RCTs under ideal conditions may differ with the effectiveness estimated in the field normally in non-ideal conditions and in different populations. Observational methods of estimating vaccine effectiveness are particularly important when disease incidence does not decrease as expected with increased vaccine coverage. There are many potential biases inherent in observational methods to estimate vaccine effectiveness which should be considered in the study design and analysis stage. Hence any factor that raises or lowers the apparent attack rates in either the vaccinated or non-vaccinated groups will bias the vaccine effectiveness estimates. Biases may arise in case-ascertainment, which is normally achieved through the randomization in the RCTs, ascertainment of vaccination status which may be also subject to classification errors in the vaccination status, these may reduce the vaccine effectiveness estimates. Hence there is a bias towards misrepresenting vaccinated cases as unvaccinated cases there might be an underestimation of the vaccine coverage and thus hinder the estimation of its effectiveness. During the comparison of the two groups (vaccinated and unvaccinated) using the two study types there could be a potential bias because while for the RCTs potentially confounding variables are randomly assigned among the groups under study, in the observational studies the groups may differ in many ways, only some of which may be recognized by the investigator. Typically the efficacy estimates obtained from RCTs, measured in ideal conditions, may differ from effectiveness estimates measured from not so ideal field conditions and population settings.

There are infections that are seasonal by nature infection, thus an estimation of the protective effect of vaccination would be limited to periods of circulation of the infection. Consequently the assessment of the vaccine effectiveness during periods of circulation and non-circulation of the infection could really help to distinguish a true vaccine effect from an effect of bias due to differences in the underlying characteristics of the vaccinated and unvaccinated groups [4].

Vaccination is a potent tool to reduce morbidity associated with a vaccine preventable disease and also the mortality emanating from the cases of morbidity. Typically for every vaccination campaign there are groups that are considered priority as compared to other sections (groups) in the population under study and are thus targeted for vaccination. These may normally include children, pregnant women, healthcare sector practitioners, the elderly and also those living with underlying conditions. Typically these groups would be encouraged to receive the vaccination

and the vaccines would be offered free of charge to them thus ensuring that they are able to obtain the risk, this may result in herd immunity. Herd immunity was defined as the concept of protection offered by having more and more people in a population getting vaccinated, to the members of the same population who've not been vaccinated.

Typically vaccine schedules (list of recommended vaccines) continually change as new vaccines are introduced, booster doses are added and the timing of the doses is changed. Thus to assure the public and foster their confidence in the vaccines in the schedule, it is essential that the effectiveness of the new vaccines and effects of the changes in the schedules be evaluated. The evaluation of current vaccine schedules should also be monitored to enable the detection of variation in effectiveness of the vaccines over time which may result from changes in the target population or in the epidemiology of the diseases under consideration [8]. In the case of new vaccines, effectiveness studies may be added as a part of surveillance after licensing of the vaccine. Though it should be noted that when most of the population under study is vaccinated, most cases of disease infections might be from vaccine failure, and such high proportions of vaccines failure might not necessarily point towards declining vaccine effectiveness.

Vaccination may produce more than one effect, both at the individual and population level [8]. The individual effects include the production of an immunologic response, protection against infection or in some cases only against disease or severe disease, a reduction in the degree or duration of infectivity, or even behavioral effects such as changes in the rate of contact with potentially infectious sources. Population effects include a reduction in transmission of disease and low infections with the disease. Accurately estimating the public health benefits of vaccination strategies is paramount to enable pandemic planning and to judge the need for alternative preventive approaches during non-pandemic periods such as increasing the range and intensity of universal immunization (vaccination) which will provide herd immunity within the population under study. This will in turn benefit the whole population whether vaccinated or not.

CHAPTER THREE: METHODOLOGY

3.1 Research Design

The study was a prospective observational study, carried out between June and September 2010. Study participants were part of a Population Based Demographic Surveillance (PBDS) system in Kibera. Children aged between 6months and less than 11 years from households enrolled in the PBDS were eligible for vaccination with the seasonal influenza vaccine and thus could be enrolled into the study. There were three vaccination centres during the vaccination campaign; once a participant arrived at the vaccination centre they were requested for their consent to participate in the study. Data for the time to event of interest (influenza infection) was collected before the start of the vaccination campaign and after the end of the vaccination campaign.

3.2 Description of the Methods

Those parents who consented to have their children participate in the study had the children's temperature taken and were also asked if children had allergies to eggs and any other allergies. For those eligible children who had fever or had allergies, vaccination was not considered; this was to avoid any side effects due to the vaccination. Vaccination forms were administered as the vaccination exercise was carried out, whether a participant was vaccinated or not; the form had questions related to the participant and the vaccination exercise thus they typically collected participant identification data and vaccination data. The form was administered using a netbook and thus all the data was available immediately once a participant had been vaccinated.

The dataset used for the study consisted of children who were eligible for vaccination and consented to participate in the study; it had the demographic variables, household and individual identification variables plus the vaccination related variables. The dataset used for this study was a Right Censored survival dataset more specifically a Type I right censored dataset. The vaccination status variable was categorical with two categories those who were fully vaccinated and those who not vaccinated at all.

3.3 Theory of survival distribution functions

The variable of interest (**response variable**) was the time to influenza infection or censoring time, with a censoring variable which indicated whether a participant was censored or experienced the event of interest as another key variable.

Hence let

Y_j Be the time to influenza infection or censoring time for the **j**th participant.

δ_j Be the censoring variable which indicates whether the **j**th participant was censored or experienced the event or interest, **value 1** when **j**th participant was censored and **0** otherwise.

T_j Be the time to event of interest (influenza infection) for the **j**th participant.

C_j Be the censoring time for the **j**th participant.

$$Y_j = \begin{cases} T_j, & \delta_j = 0 \\ C_j, & \delta_j = 1 \end{cases}$$

$\boldsymbol{Y_j \geq 0}$**,**

$\boldsymbol{T_j \geq 0}$

$\boldsymbol{C_j \geq 0}$

Hence for the dataset there were **k** happenings of the event of interest (**Influenza Infection)**.

Thus let

t_i, be the time of the happening of the **i**th event of interest. Where $\boldsymbol{i = 1, 2, \ldots, k}$

To achieve our objective of comparing the survival distribution functions for the both the fully vaccinated and unvaccinated groups we would need to estimate their respective survival functions. We could use the Kaplan-Meir estimator or the Nelson-Aalen estimator to achieve this described objective.

The Kaplan-Meir estimator is the method for estimating the **Survival function, $S(t)$.**

$S(t) = \mathbf{Pr}(T > t)$, the probability of survival beyond time t i.e. the time to event (T) occurring after time t.

Hence let

n_i, be the number of participants at risk of event of interest (Influenza Infection) at time t_i.

d_i, be the number of events of interest (**Influenza Infection**) observed at time t_i.

By description the conditional probability of surviving past time t_i given survival to that time is estimated by $\frac{(n_i - d_i)}{n_i}$.

Thus overall, then the unconditional probability of surviving beyond any time t is estimated by the below equation;

$$\widehat{S}(t) = \prod_{t_i \leq t} \frac{(n_i - d_i)}{n_i}$$

(Equation 3a)

The above equation is known as the **Kaplan-Meir estimator**. Recall $i = 1, 2, \ldots, k$

To obtain the **Kaplan-Meir estimator** one needs to construct a table like the one below, or from the use of statistical software one would produce a graph of the estimator.

i	t_i	d_i	c_i	n_i	$(n_i - d_i)$	$\frac{(n_i - d_i)}{n_i}$	$\hat{S}(t)$
0	t_0	d_0	c_0	n_0	$(n_0 - d_0)$	$\frac{(n_0 - d_0)}{n_0}$	$\frac{(n_0 - d_0)}{n_0}$
1	t_1	d_1	c_1	n_1	$(n_1 - d_1)$	$\frac{(n_1 - d_1)}{n_1}$	$\frac{(n_0-d_0)}{n_0} * \frac{(n_1-d_1)}{n_1}$
.							
.							
.							
k	t_k	d_k	c_k	n_k	$(n_k - d_k)$	$\frac{(n_k - d_k)}{n_k}$	$\frac{(n_0-d_0)}{n_0} * \frac{(n_1-d_1)}{n_1} * \ldots * \frac{(n_k-d_k)}{n_k}$

(**Table 3a)**

The Nelson-Aalen estimator is the method for estimating the **Cumulative hazard function,** $H(t)$**.**

The relationship between $H(t)$ and the $S(t)$ **the cumulative hazard function and the survival function respectively is illustrated below.**

$$S(t) = exp(-\int_0^t h(u)du$$

Where $h(u)$ is the hazard function.

$$S(t) = exp(-H(t))$$

Thus the above is the relationship between the cumulative hazard function and the survival function.

$S(t) = \Pr(T > t)$, the probability of survival beyond time t i.e. the time to event (T) occurring after time t.

Hence let

n_i, be the number of participants at risk of event of interest (Influenza Infection) at time t_i.

d_i, be the number of events of interest (**Influenza Infection**) observed at time t_i.

By description the estimate of the cumulative hazard function is given by

$$\widehat{H}(t) = \sum_{t_i \leq t} \frac{d_i}{n_i}$$

(Equation 3b)

The above equation is known as the **Nelson-Aalen estimator**. Recall $\boldsymbol{i = 1, 2, \ldots, k}$

To obtain the **Nelson-Aalen estimator** one needs to construct a table like the one below, or from the use of statistical software would produce a graph of the estimator.

$\boldsymbol{i}$	$\boldsymbol{t_i}$	$\boldsymbol{d_i}$	$\boldsymbol{c_i}$	$\boldsymbol{n_i}$	$\boldsymbol{\frac{d_i}{n_i}}$	$\widehat{H}(t)$
0	$\boldsymbol{t_0}$	$\boldsymbol{d_0}$	$\boldsymbol{c_0}$	$\boldsymbol{n_0}$	$\boldsymbol{\frac{d_0}{n_0}}$	$\boldsymbol{0}$
1	$\boldsymbol{t_1}$	$\boldsymbol{d_1}$	$\boldsymbol{c_1}$	$\boldsymbol{n_1}$	$\boldsymbol{\frac{d_1}{n_1}}$	$\boldsymbol{0 + \frac{d_1}{n_1}}$
.						
.						
.						
k	$\boldsymbol{t_k}$	$\boldsymbol{d_k}$	$\boldsymbol{c_k}$	$\boldsymbol{n_k}$	$\boldsymbol{\frac{d_k}{n_k}}$	$\boldsymbol{0 + \frac{d_1}{n_1} + \frac{d_2}{n_2} + \cdots + \frac{d_k}{n_k}}$

(Table 3b)

3.4 Comparison of two Survival distribution functions

The objective of comparison of two Survival distribution functions would be achieved using the Log Rank Test.

To achieve our objective of comparing whether the survival distribution functions for the fully vaccinated and unvaccinated groups are statistically significantly different. We would have to test the hypothesis below

$$H_o: S_{(a)} = S_{(b)}$$

Versus

$$H_1: S_{(a)} \neq S_{(b)}$$

Where $S_{(a)}$ is the survival function for the fully vaccinated group while the $S_{(b)}$is the survival function for the unvaccinated group. Hence we have

Let n_i **and** t_i be as defined earlier,

n_i, be the number of participants at risk of event of interest (**Influenza Infection**) at time t_i.

d_i, be the number of events of interest (**Influenza Infection**) observed at time t_i.

Then we shall define

n_{mi}, be the number of participants at risk of event of interest (Influenza Infection) at time t_i, for the $\mathbf{m}^{\mathbf{th}}$ group.

d_{mi}, be the number of events of interest (**Influenza Infection**) observed at time t_i, for the $\mathbf{m}^{\mathbf{th}}$ group.

Let **X** be the random variable that takes the values of d_{mi}.

Recall $m = 1, 2$; there are two groups fully vaccinated and unvaccinated groups.

A log-rank test needs 2* 2 contingency table with the variables described in the slide before to obtain the test statistic.

Group	**No. Infected**	**No. Not-Infected**	**No. At Risk**
Fully Vaccinated	d_{1i}	$n_{1i} - d_{1i}$	n_{1i}
Unvaccinated	d_{2i}	$n_{2i} - d_{2i}$	n_{2i}
Total	d_i	$n_i - d_i$	n_i

The log-rank test statistic follows a hyper-geometric distribution hence the probability of the random variable **X** taking the value d_{1i}, would be given by

$$Prob(X = d_{1i}) = \frac{\binom{n_{1i}}{d_{1i}}\binom{n_{2i}}{d_{2i}}}{\binom{n_i}{d_i}}$$

While the expectation and variance of the random variable **X** are given as below

$$E(X) = \frac{d_i n_{1i}}{n_i} \quad \text{and} \quad Var(X) = \frac{n_{1i} n_{2i} d_i (n_i - d_i)}{n_i^2 (n_i - 1)}$$

Let **Y**, the total number of infections in group 1 (Fully vaccinated group).

$$Y = \sum_{t_i} d_{1i}$$

Then

$$E(Y) = \sum_{t_i} E(d_{1i}) \text{ and } Var(Y) = \sum_{t_i} Var(d_{1i})$$

Let **Z** be defined as below,

$$Z = \frac{Y - E(Y)}{\sqrt{Var(Y)}}$$

$E(Y) = 0$ and $Var(Y) = 1$ consequently $Z \sim N(0,1)$.

After standardization then **Z** is given as below

$$Z = \frac{\sum_{t_i} d_{1i} - \sum_{t_i} E(d_{1i})}{\sqrt{\sum_{t_i} Var(d_{1i})}}$$

(Equation 3c)

Then from probability distribution if consequently $Z \sim N(0,1)$ then $Z^2 \sim \times^2 (1)$.

Thus Z^2 is known as the log-rank test statistic and it follows a chi-square distribution with one degree of freedom.

$$Z^2 = \frac{\left\{\sum_{t_i}\left[d_{1i} - \frac{d_i n_{1i}}{n_i}\right]\right\}^2}{\sum_{t_i} \frac{n_{1i} n_{2i} d_i (n_i - d_i)}{n_i{}^2 (n_i - 1)}}$$

(Equation 3d)

Thus to conduct the hypothesis test for the hypotheses below

$$H_o: S_{(a)} = S_{(b)}$$

Versus

$$H_1: S_{(a)} \neq S_{(b)}$$

You obtain a critical value from the chi-square distribution tables at $\propto$ level of significance. If the **Test Statistic (Z^2) is greater than the Critical Value** then reject $H_o: S_{(a)} = S_{(b)}$**.**

3.5 Estimating the effectiveness of the vaccine using time to event data analysis

To achieve our objective of estimating the effectiveness of the vaccine given we would have to conduct a cox proportional hazards model regression analysis. We would have to test whether the regression coefficients in the model pertaining to each of the predictor variables including the vaccination status are statistically significantly different from zero. The hypotheses to be tested would be

$$H_o: \beta_l = 0$$

Versus

$$H_1: \beta_l \neq 0$$

The cox proportional hazards regression model is the most popular approach to modeling the predictors in order to estimate their effects on the time to event. It is a semi-parametric model because it takes no assumption of the probability distribution function for the time to event; the baseline hazard takes any form and the predictors enter the model through the linear predictor, n_l, . It is based on the assumption that the hazards are proportional.

The cox proportional hazards regression model is based on the hazard function, $h(t, X)$ at time t

If we let $x_1, x_2, \ldots, x_p$ be the values taken by the following covariates $X_1, X_2, \ldots, X_p$. Where **p** is the total number of predictor variables in a model.

Then the linear predictor, n_l, is given as

$$n_l = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p$$

Consequently the cox proportional hazards regression model is expressed as follows

$$\boldsymbol{h(t,X) = h_0(t)exp\left(\sum_{l=1}^{p}\beta_l X_l\right)}$$

(Equation 3e)

$\boldsymbol{h_0(t)}$Is the baseline hazard function at time t.

β_l Is the **l**$^{\textbf{th}}$ regression parameter for the cox proportional hazards model, while X_l is **l**$^{\textbf{th}}$ predictor variable

Where

$\boldsymbol{h_0(t) \geq 0}$

and

$$\sum_{l=1}^{p}\beta_l X_l \geq \mathbf{0}$$

Recall$\boldsymbol{l = 1, 2, \ldots, p}$**,** is the number of predictors in the cox proportional hazards model

One advantage of the cox proportional hazards model is that it takes care of censored observations aside from it being semi-parametric.

The proportionality of the hazards is an important assumption of the cox proportional hazards model, it means that the hazard ratio is constant over time, or more precisely the hazard for an individual observation is proportional to the hazard for any other individual observation.

The hazard ratio (HR) is obtained as shown from the ratio of two hazard functions.

Let $\boldsymbol{h(t,X) = h_0(t)exp\left(\sum_{l=1}^{p} \beta_l X_l\right)}$ be the hazard function for individual A

And

Let $\boldsymbol{h(t,X^*) = h_0(t)exp\left(\sum_{l=1}^{p} \beta^*{}_l X^*{}_l\right)}$ be the hazard function for individual B

Then the hazard ratio (HR) is given by **HR**=$\frac{h(t,X)}{h(t,X^*)}$=$\exp\left[\sum_{l=1}^{p}(x_l{}^* - x_l)\right]$

Hence when the value above for the individuals A and B is constant then the proportional hazards assumption is satisfied.

To estimate β_l the **l**$^{\textbf{th}}$ regression parameter for the cox proportional hazards model we'll use the partial likelihood method. We shall only estimate for the **l**$^{\textbf{th}}$ covariate.

Steps to obtaining the Partial Likelihood

- Arrange the event times in order i.e. $t_1 < t_2 <, \dots, < t_k$.
- Obtain the risk sets $R(t_1), R(t_2), \dots, R(t_k)$
- Then obtain the probability of an individual dying at time t_i given risk set $R(t_i)$, $t_i / R(t_i)$, which is given by $= \frac{exp(\beta X_i)}{\sum_{r \in R(t_i)} exp(\beta X_r)}$
- Then finally obtaining the **Partial Likelihood** $\boldsymbol{L(\beta)}$

We shall assume that there are k different event of interest times such that there are no ties at the

$\boldsymbol{t_i}$, be the time of the happening of the **i**$^{\textbf{th}}$ event of interest. Where $\boldsymbol{i = 1, 2, \dots, k}$

Then the partial likelihood $\boldsymbol{L(\beta)}$ can be obtained as shown below

$$\boldsymbol{L(\beta) = \prod_{i=1}^{k} \frac{h_i(t_i)}{\sum_{r \in R(t_i)} h_r(t_r)}}$$

$$\boldsymbol{L(\beta) = \prod_{i=1}^{k} \frac{exp(\beta X_i)}{\sum_{r \in R(t_i)} exp(\beta X_r)}}$$

(Equation 3f)

Solving the equation above enables us to obtain $\boldsymbol{exp(\beta)}$ and thenβ.

The steps to solving the equation are listed below.

- Obtain the partial likelihood $\boldsymbol{L(\beta)}$
- Obtain the$Log\big(\boldsymbol{L(\beta)}\big)$**,** natural log
- Then obtain the first derivative of the $Log\big(\boldsymbol{L(\beta)}\big)$ with respect to β**,** $\boldsymbol{\frac{d\ Log(L(\beta))}{d\ \beta}}$

With the below derivative

$\boldsymbol{\frac{d\ Log(L(\beta))}{d\ \beta} = 0}$ **,** we can obtain $\boldsymbol{exp(\beta)}$ and then β

Hence we can estimate the B_l the $\mathbf{l^{th}}$ regression parameter for the cox proportional hazards model.

To test the significance of the parameters obtained we can either use the Wald, Partial Likelihood ratio and Score Test as means of doing that.

We shall use the Wald to test significance of the parameters, using the hypotheses test below

$$H_o: \beta_l = 0$$

Versus

$$H_1: \beta_l \neq 0$$

The Wald test statistic denoted by W, is obtained as below

$$W = \frac{\bar{\beta}_l}{se(\bar{\beta}_l)}$$

(**Equation 3g)**

$$W \sim H_o \sim N(0,1)$$

Thus to conduct the hypothesis test for the hypotheses below

$$H_o: \beta_l = 0$$

Versus

$$H_1: \beta_l \neq 0$$

You obtain a critical value from the normal distribution tables at $\propto$ level of significance. If the **Test Statistic (W) is greater than the Critical Value** then reject $H_o: \beta_l = 0$

Hence one would be able to obtain the significance for the regression parameter.

CHAPTER FOUR: DATA ANALYSIS AND RESULTS

4.1 Data analysis and results

The data analysis was conducted using STATA S/E version 9.2 statistical software; the techniques used for analysis of the data were survival data analysis techniques. The techniques used for the analysis were geared towards the achievement of the objectives outlined in the objectives section of this paper.

This data analysis and result section is split into two; the first section deals with the period before the vaccination campaign and the second section for the period after the campaign.

Objective: Compare the survival distribution functions for both the fully vaccinated and unvaccinated groups before the vaccination campaign

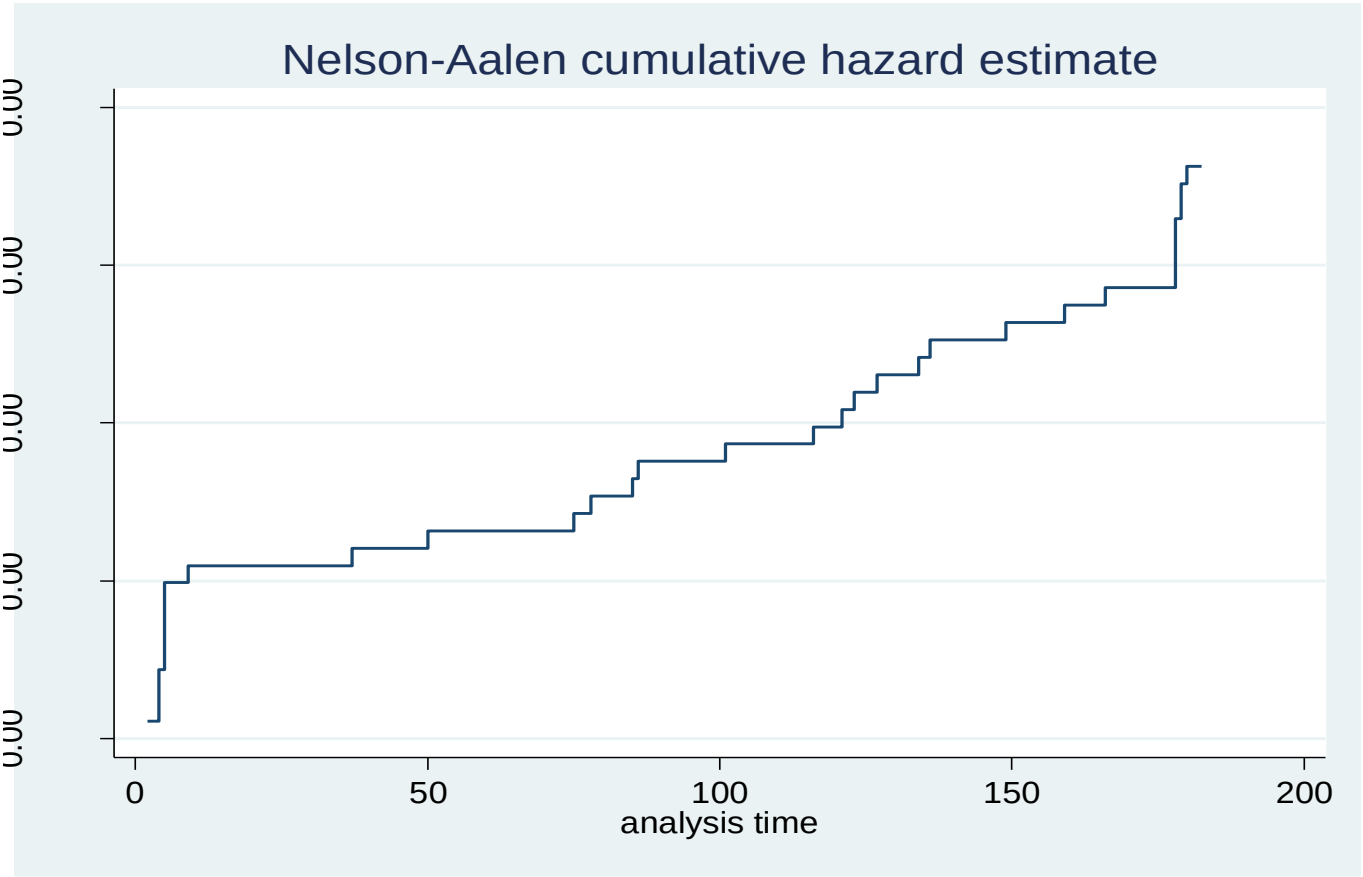

Figure 4a: A graph with the cumulative hazards curve for all the participants before the vaccination campaign.

Objective: Compare the survival distribution functions for both the fully vaccinated and unvaccinated groups before the vaccination campaign

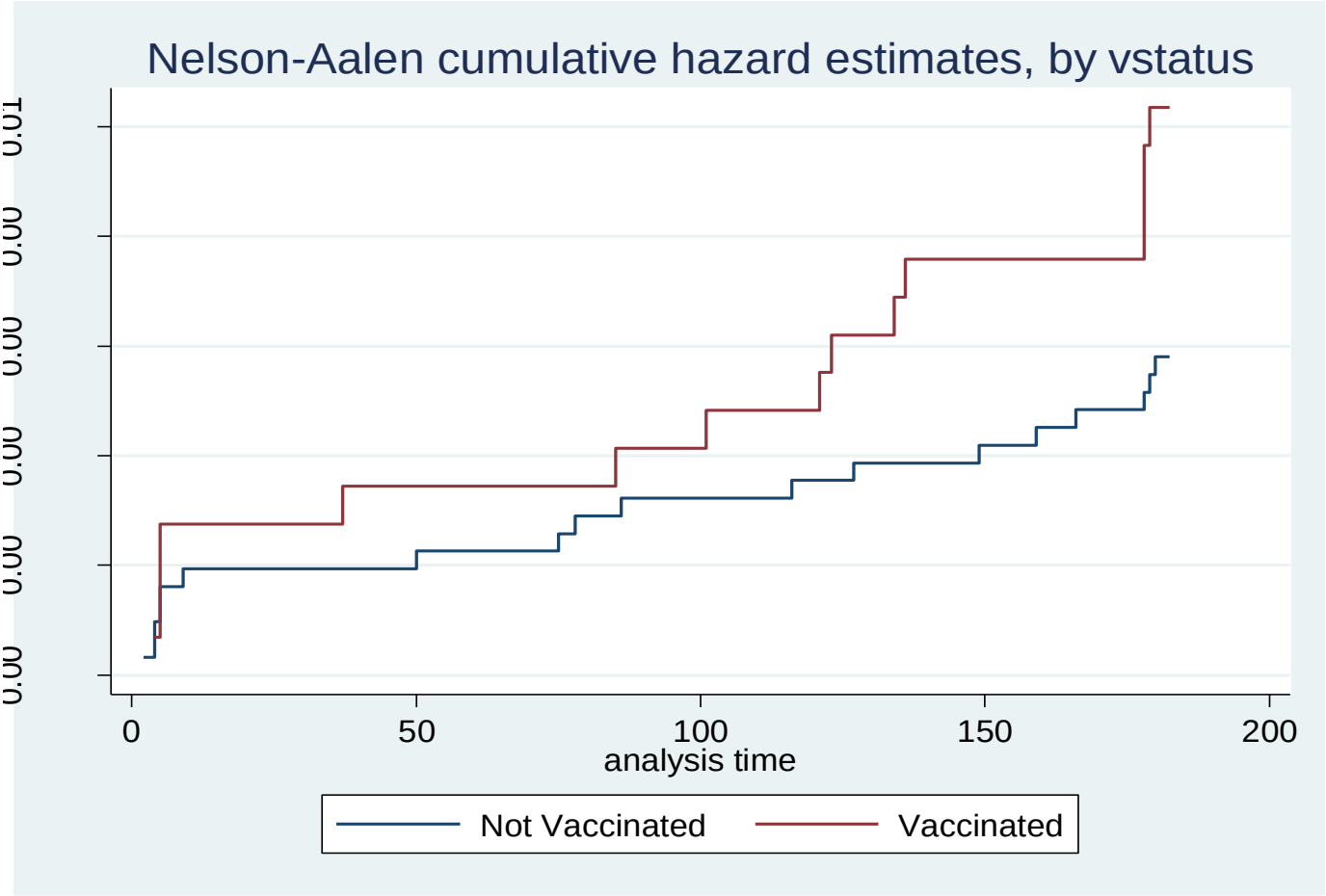

Figure 4b: A graph with the cumulative hazards curves for the Fully vaccinated and the Unvaccinated groups before the vaccination campaign.

The Nelson-Aalen estimator using the cumulative hazards function was used to plot the above graphs of the **cumulative hazards versus the time to event of interest (influenza infection) during the period before the vaccination campaign**; both for the two groups together and also for the two groups individually on one plot. This was to characterize the cumulative hazards for both the groups individually and together as a population.

Objective: Compare the survival distribution functions for both the fully vaccinated and unvaccinated groups before the vaccination campaign

Variable	Rho	chi2	Df	Pr>chi2
Gender	0.07229	0.17	1	0.6779
Vaccination Status	0.07741	0.20	1	0.6565
Global Test		0.37	2	0.8308

Table 4a: Proportional hazards assumption test before the vaccination campaign

The table above is a test for the assumption of proportional hazards for the main predictor variable vaccination status, and also for the available confounder gender for the period before the vaccination campaign.

Vaccination Status	Events observed	Events expected
Unvaccinated	18	22.49
Fully vaccinated	15	10.51
Total	33	33
Test Statistic	chi2(1)	2.82
P-value	Pr>chi2	0.0933

Table 4b: The log rank test before the vaccination campaign

The table **(Table 4b)** is a test for the equality of survival distribution functions for the two groups for the period before the vaccination campaign using the log rank test.

Objective: Estimate the Vaccine Effectiveness using the time to event data analysis

Variable	Hazard Ratio	Std. Err.	Z	P>z	[95% Conf.	Interval]
Gender	**1.146282**	**0.4007439**	**0.39**	**0.6960**	**0.5777086**	**2.274439**
Vaccination Status	**1.783206**	**0.6234144**	**1.65**	**0.0980**	**0.898708**	**3.538215**

Table 4c: The Cox Proportional Hazard Regression Analysis before the vaccination campaign

The above table contains the results of Cox Proportional Hazards Regression Analysis **during the period before the vaccination campaign**, it includes the main predictor variable vaccination status, and also for the available confounder gender in the model.

We used the Cox proportional hazards regression analysis to estimate the hazard ratio of the event of interest (**influenza infection**) for the fully vaccinated and unvaccinated study participants after the assumption of the proportional hazards was guaranteed using the Proportional Hazards test.

Objective: Compare the survival distribution functions for both the fully vaccinated and unvaccinated groups after the vaccination campaign

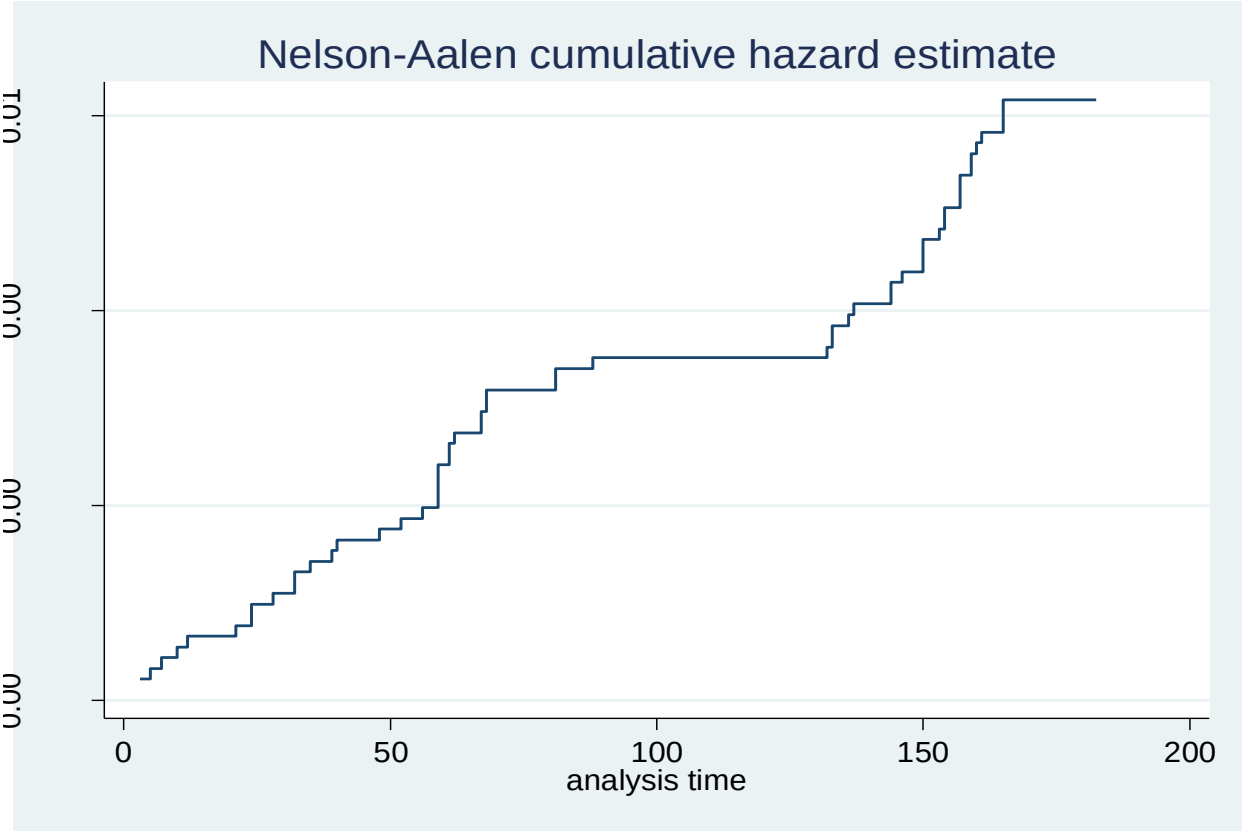

Figure 4c: A graph with the cumulative hazards curve for all the participants after the vaccination campaign.

Objective: Compare the survival distribution functions for both the fully vaccinated and unvaccinated groups after the vaccination campaign

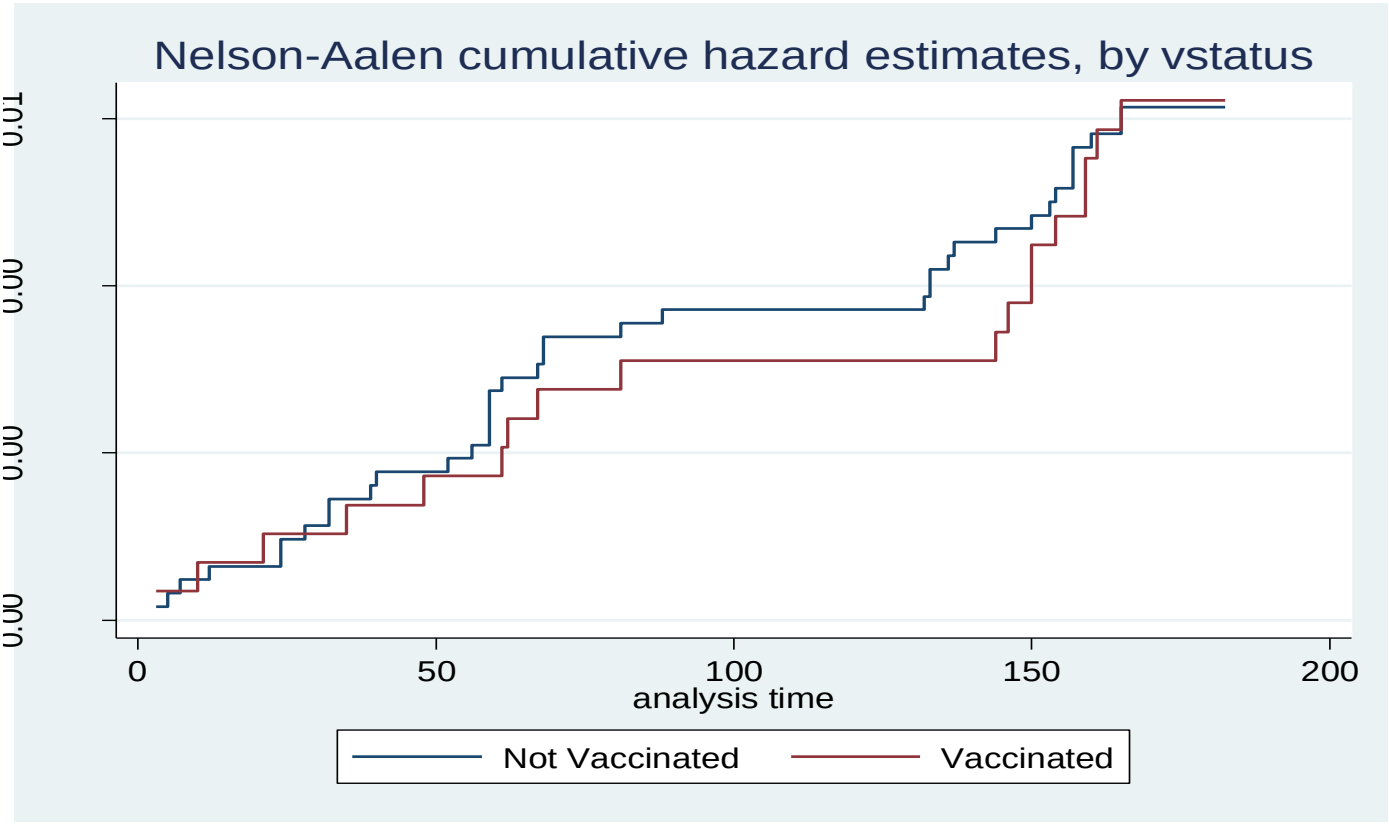

Figure 4d: A graph with the cumulative hazards curves for the Fully vaccinated and the Unvaccinated groups after the vaccination campaign.

The Nelson-Aalen estimator using the cumulative hazards function was used to plot the above graphs of the **cumulative hazards versus the time to event of interest (influenza infection) during the period after the vaccination campaign**; both for the two groups together and also for the two groups individually on one plot. This was to characterize the cumulative hazards for the both the groups individually and together as a population.

Objective: Compare the survival distribution functions for both the fully vaccinated and unvaccinated groups after the vaccination campaign

Variable	rho	chi2	df	Pr>chi2
Gender	0.03772	0.08	1	0.7777
Vaccination Status	0.10730	0.64	1	0.4220
Global Test		0.73	2	0.6959

Table 4d: Proportional hazards assumption test after the vaccination campaign

The table above is a test for the assumption of proportional hazards for the main predictor variable vaccination status, and also for the available confounder gender for the period after the vaccination campaign.

Vaccination Status	Events observed	Events expected
Unvaccinated	38	38.15
Fully vaccinated	18	17.85
Total	56	56
Test Statistic	chi2(1)	0.00
P-value	Pr>chi2	0.9667

Table 4e: The log rank test after the vaccination campaign

The table **(Table 4e)** is a test for the equality of survival distribution functions for the two groups for the period after the vaccination campaign using the log rank test.

Objective: Estimate the Vaccine Effectiveness using the time to event data analysis

Variable	Hazard Ratio	Std. Err.	z	P>z	[95% Conf.	Interval]
Gender	**0.889379**	**0.2378486**	**-0.44**	**0.6610**	**0.5265607**	**1.502193**
Vaccination Status	**1.012686**	**0.2897617**	**0.04**	**0.9650**	**0.5779878**	**1.774314**

Table 4f: The Cox Proportional Hazard Regression Analysis after the vaccination campaign

The above table contains the results of Cox Proportional Hazards Regression Analysis **during the period after the vaccination campaign**, it includes the main predictor variable vaccination status, and also for the available confounder gender in the model.

We used the Cox proportional hazards regression analysis to estimate the hazard ratio of the event of interest (**influenza infection**) for the fully vaccinated and unvaccinated study participants after the assumption of the proportional hazards was guaranteed using the Proportional Hazards test.

4.2 Discussion

From **figure 4b**, it's seen that the cumulative hazards curve for the fully vaccinated group is above (**higher than**) that of the unvaccinated group; thus showing that the fully vaccinated group experienced higher hazards as compared to the unvaccinated ones during the **period before the vaccination campaign**. Thus the fully vaccinated were at a higher risk of experiencing the event of interest (influenza infection) as compared to the unvaccinated ones during the period before the vaccination campaign started.

It's seen in **Table 4a** that the key assumption of the hazards assumption cox proportional hazards regression model was met by each of the variables included in the model individually and also globally as a model, thus the results from the model are valid given that the data was collected during the period **before the vaccination campaign.**

The comparison of the survival distribution functions, by conducting a test of equality of survival functions, was achieved using the log rank test especially after the proportional hazards assumption was met. From **Table 4b,** it's seen that survival distribution functions for the two groups, the fully vaccinated and the unvaccinated groups are not statistically significantly different as shown from the data collected during the **period before the vaccination campaign, p-value=0.0933.**

From the cox proportional hazards regression model analysis for the data collected during the **period before the vaccination campaign**, as shown in **Table 4c,** it's seen that females experienced hazards that were 15% higher than those for males, while also the fully vaccinated experienced hazards that were 78% higher than those who were not vaccinated. From **Table 4c** it's seen that the relationships were not statistically significant, with the p-values being equal to **0.6960** and **0.0980** for gender and vaccination status variables respectively.

From **figure 4d**, it's seen that the cumulative hazards curve for the fully vaccinated group is now below (**lower than**) that of the unvaccinated group; thus showing that the fully vaccinated group experienced lower hazards as compared to the unvaccinated ones during the **period after the vaccination campaign**. Thus the fully vaccinated were at a lower risk of experiencing the event of interest (influenza infection) as compared to the unvaccinated ones after the vaccination campaign.

It's seen in **Table 4d** that the key assumption of the hazards assumption cox proportional hazards regression model was met by each of the variables included in the model individually and also globally as a model, thus the results from the model are valid given that the data was collected during the **period after the vaccination campaign.**

The comparison of the survival distribution functions, by conducting a test of equality of survival functions, was achieved using the log rank test especially after the proportional hazards assumption was met. From **Table 4e,** it's seen that survival distribution functions for the two groups, the fully vaccinated and the unvaccinated groups are not statistically significantly different as from the data collected during the **period after the vaccination campaign, p-value=0.9667.**

From the cox proportional hazards regression model regression analysis for the period before the vaccination campaign, as shown in **Table 4f,** it's seen that females experienced hazards that were **11% lower** than those for males, while also the fully vaccinated experienced hazards that were **only 1% higher** than those who were not vaccinated. From **Table 4f** it's seen that the relationships were not statistically significant, with the p-values being equal to **0.6610** and **0.9650** for gender and vaccination status variables respectively.

CHAPTER FIVE: CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

Those who got vaccinated during the vaccination campaign had massively higher hazards of experiencing the event of interest (influenza infection) as compared to those who were not vaccinated this was during the period before the vaccination campaign. While during the period after the vaccination campaign those who got vaccinated had lower hazards as compared to those who were not vaccinated of experiencing the event of interest (influenza infection). This may be explained partly by the view that those who got vaccinated were encouraged to get vaccinated during the vaccination campaign by their experiences (higher hazards) during the period before the vaccination campaign started. Thus these experiences of higher hazards spurred those who got vaccinated to get the vaccinated so as to prevent them from getting infected. That's why the hazards for the two groups switched after the end of the vaccination campaign. This was evidenced by the Nelson-Aalen Cumulative Hazards plots in **figure 4b** and **figure 4d**.

The above has also been supported by the results of the cox proportional hazards regression models which show that the hazards experienced by the vaccinated group reduced between the period before the vaccination campaign and the period after the vaccination campaign, thus showing a protective effect of the vaccine for those who were erstwhile not vaccinated but became vaccinated during the vaccination campaign. Though the results of the cox proportional hazards models regression analyses were not statistically significant, this may be due to the lack of other confounders and predictors related to the main predictor variable and the outcome of interest and the time to event of interest to enrich the model.

The non-significance of the results from the cox proportional hazards models regression analyses may lead to the need to add more predictor variables into the model thus enriching it. This would be possible by obtaining the profiles of the participants including demographic details and also their functional statuses to augment the model and thus the analyses too.

The vaccine offered a level of protection to those who were vaccinated but this wasn't statistically significant as seen in the results section.

5.2 Recommendations

The profiles of the participants including demographic details and also their functional statuses were not present in the dataset used during the analysis, hence were not taken care of in the model. Obtaining these profiles would have been beneficial to the study in improving the multivariable model used for determining the effectiveness of the vaccine by estimating the hazards ratios. Thus a wider range of potential predictors of vaccination could have been added to the model.

Stratification of the vaccine effectiveness analysis should have been considered as this would have also been prudent in minimizing and eliminating bias and confounding on the estimates of the vaccine effectiveness.

The above analysis would also be enriched if carried out at the end of the entire vaccination campaign when all the data from the entire rounds of the vaccination campaign are available, these data may offer a better avenue to carry out the stratified analysis described above.

Since influenza is a seasonal infection, thus an estimation of a protective effect of vaccination should be limited to periods of influenza viral circulation; there should a pre-vaccination study during the influenza virus circulation period to establish whether the two groups are comparable before the administration of the vaccination during the campaign. When the two groups are comparable, the vaccine effect will not be present in the pre-vaccination period hence one would be able to accurate estimate the effect of the vaccine using vaccine effectiveness estimates after the vaccination campaign ends. Thus this approach, of using a control period, would be better suited for observational studies like the one in the study to estimate vaccine effectiveness. Hence in a nutshell, a two phase study design, an assessment of the vaccine effectiveness during influenza and non-influenza virus circulation periods could really help to distinguish a true vaccine effect from an effect of bias due to differences in the underlying characteristics of the vaccinated and unvaccinated groups.

One should also note that influenza viruses are also subject to frequent antigenic changes, thus for these reasons the influenza vaccine is typically reformulated each year to optimize antigenic match between the vaccine and circulating virus strains. Hence the vaccine effectiveness studies would be needed yearly to estimate the effect of the vaccines, thus there is need to obtain ways

of obtaining the estimates of the vaccine effectiveness using observational studies while minimizing the problems that afflict observational studies. This was the thrust of the above research project to show how time to event data (survival data) analysis can be used to estimate vaccine effectiveness using data from an observational study.

Given the evidence of bias in existing observational studies then obtaining a better understanding of the failure of current adjustment methods applied in observational studies to guard against the effects of confounders and thus minimize bias would be an important area of future further research.

REFERENCES

[1] Ehreth J. The global value of vaccination: Vaccine volume 21, 2003:596-600.

[2] Fine PEM. Herd Immunity (History, Theory, and Practice): The John Hopkins University School of Hygiene and Public Health, Epidemiologic Reviews volume 11, No 2, 1973.

[3] Kelly H, Steffens I. Complexities in assessing the effectiveness of inactivated Influenza Vaccines: Eurosurveillance volume 18, No 7, 2013.

[4] Jackson LA, et al. Evidence of bias in estimates of influenza vaccine effectiveness in seniors: International Journal of Epidemiology volume 35, 2006:337-344.

[5] Skowronski DM, et al. Interim estimates of influenza vaccine effectiveness in 2012/13 from Canada's sentinel surveillance network: Eurosurveillance volume 18, No 5, 2013.

[6] Valenciano M, et al. I-MOVE: a European network to measure the effectiveness of influenza vaccines: Eurosurveillance volume 17, No 39, 2012.

[7] Nelson J.C, et al. New Strategies are needed to improve the accuracy of influenza vaccine effectiveness estimates among seniors: Journal of Clinical Epidemiology volume 62, 2009:687-694

[8] Torvaldsen S, McIntyre BP. Observational methods in epidemiologic assessment of vaccine effectiveness: National Centre for Immunization Research and Surveillance of Vaccine Preventable Diseases, 2002.