2011
2012

FACULTY OF SCIENCES

Master of Statistics: Biostatistics

Masterproef

Evaluation of different strategies for weighted average vaccine efficacy

Promotor :
Prof. dr. Marc AERTS

Promotor :
Ms. MARIE-PIERRE DAVID
Dr. FABIAN TIBALDI

Trias Wahyuni Rakhmawati

Master Thesis nominated to obtain the degree of Master of Statistics , specialization Biostatistics

Maastricht University

universiteit hasselt
KNOWLEDGE IN ACTION

De transnationale Universiteit Limburg is een uniek samenwerkingsverband van twee universiteiten in twee landen: de Universiteit Hasselt en Maastricht University

universiteit hasselt
KNOWLEDGE IN ACTION

Maastricht University

Universiteit Hasselt | Campus Diepenbeek | Agoralaan Gebouw D | BE-3590 Diepenbeek
Universiteit Hasselt | Campus Hasselt | Martelarenlaan 42 | BE-3500 Hasselt

2011
2012

FACULTY OF SCIENCES

Master of Statistics: Biostatistics

Masterproef

Evaluation of different strategies for weighted average vaccine efficacy

Promotor :
Prof. dr. Marc AERTS

Promotor :
Ms. MARIE-PIERRE DAVID
Dr. FABIAN TIBALDI

Trias Wahyuni Rakhmawati

Master Thesis nominated to obtain the degree of Master of Statistics , specialization Biostatistics

Maastricht University

universiteit hasselt
KNOWLEDGE IN ACTION

Master Thesis

Evaluation of Different Strategies for Weighted Average Vaccine Efficacy

By:

Trias Wahyuni Rakhmawati

Internal Supervisor:

Prof. Dr. Marc Aerts

External Supervisors:

Ms. Marie-Pierre David

Dr. Fabian Tibaldi

September, 2012

Abstract

Most literatures on estimation of vaccine efficacy assumed random mixing throughout the population. In the practice, the population is somehow stratified into different groups obviously into the baseline status of HPV, naïve and non-naïve sub-population, when the study related to efficacy of HPV vaccine. The present project aims to estimate the efficacy of HPV vaccine in a weighted manner based on baseline HPV status. In this paper, different approaches were proposed for estimating the weighted average vaccine efficacy related to different measurement of relative risk (RR). Simulation study was used to compare the performance of the different proposed methods for weighted average vaccine efficacy.

Acknowledgements

First of all, I would like to thank and grateful to the Almighty Allah who always guiding me in my life.

I would also like to thank to my internal and external supervisors, for their constant helps and guidance. Their advices were proved to be priceless for the completion of this project.

I express my gratitude to the Flemish Interuniversity Council (VLIR) for granting me a scholarship in Hasselt University so that I can gain more knowledge to benefit of my country and myself.

Last but not least, I owe great thanks to my parents, brother and sisters as well as my friends who always given me support within my study periods.

"So which of the favors of your Lord would you deny? " (QS. 55:13).

Contents

1 Introduction

1.1 Background

Human Papillomavirus (HPV) is a virus from the *Papillomavirus* family that infects humans. HPVs develop infections only in the area of the skin, keratinocytes or mucous membranes. The most of the known types of HPV cause no symptoms in most people, some types can cause warts, while others can have a minority of cases leading to cancers. Oncogenic *Human Papillomavirus* (HPV) has a causal role in nearly all cervical cancers and in many vulvar, vaginal, penile, anal, and or pharyngeal cancers (IARC, 2007).

More than 30 to 40 types of HPV are typically transmitted through sexual contact and infect the anogenital region. Some sexually transmitted HPV types may cause genital warts. Persistent infection with "high-risk" HPV types different from the ones that cause skin warts may progress to precancerous lesions and invasive cancer. *Human Papillomavirus* (HPV) infection appears to be a necessary factor in the development of almost all cases of cervical cancer (Schiffman & Castle, 2003).

There are more than 100 types of HPV, with more than 40 anogenital types, of which approximately 15 are oncogenic. About a dozen HPV types (including types 16, 18, 31, and 45) are called "high-risk" types because they can lead to cervical cancer, as well as anal cancer, vulvar cancer, vaginal cancer, and penile cancer. *Human Papillomavirus* 16 is uniquely oncogenic, accounting for approximately one-half of cervical cancers (Muñoz, et al., 2003).

HPV infection is the most frequent sexually transmitted disease in the world (National Institute of Allergy and Infectious Disease, 2010). Methods of prevention include abstinence, condoms, vaccination and microbicides are considered to reduce disease burden. Two vaccines are available to prevent infection of HPV types: Gardasil (Merck) and Cervarix (GlaxoSmithKline). Both protect against initial infection with HPV types 16 and 18, which cause most of the HPV associated cervical cancer cases. Gardasil also protects against HPV types 6 and 11, which cause 90% of genital warts. In order to see how effective the vaccine could be given ideal circumstances, the vaccine efficacy needs to be estimated.

1.2 Introduction of the Problem

In a clinical trial, the total vaccinated cohort (TVC) included all women who were given at least one vaccine dose and were evaluable for efficacy irrespective of other criteria and was intended to represent the general population of women, including those who are sexually active. This condition makes the baseline characteristics of the population diverse. The HPV status of each patient at baseline is assessed by the presence or not of HPV DNA in the cervical sample taken before the administration of the vaccine. The global population can be divided between a naïve sub-population made of women who are HPV DNA negative at baseline (no evidence of oncogenic HPV infection) and a non-naïve sub-population made of women who are HPV DNA positive at baseline (evidence of oncogenic HPV infection). The HPV vaccine works better for naïve population compared to non-naïve population (FDA, 2006).

In order to get better estimates for the overall vaccine efficacy on the global population, the baseline HPV status of the subjects need to be taken into account. Furthermore, adaptations of the classical methods of vaccine efficacy are needed. There is no common agreement for the best strategy to estimate the overall vaccine efficacy on the general population. The different methods of vaccine efficacy that take into account the heterogeneity of the population need to be considered.

1.3 Objective

The heterogeneity of the baseline HPV status needs to be taken into account when estimates overall vaccine efficacy. The heterogeneity can be captured by the weighting factor for each sub-population, HPV naïve and HPV non-naïve cohort. The objective of the study is to explore the different methods of the vaccine efficacy in weighted manner. In this study, the vaccine efficacy was defined based on different approaches of relative risk. The weighted average of vaccine efficacy was applied to those different approaches. Furthermore, simulations were used to compare the performance of the different proposed methods.

2 Statistical Methodology

2.1 General Definition of VE

Vaccine efficacy (VE) is the reduction in incidence of the disease due to the infectious agent if everyone took the vaccine compared to the incidence of the disease due to the infectious agent under the placebo/control (Small, et al., 2010). The general expression for vaccine efficacy (VE) is described by the following expression:

$$\begin{aligned} VE &= \frac{\mathrm{R}\left(\text{Unvaccinated}\right) - \mathrm{R}\left(\text{Vaccinated}\right)}{\mathrm{R}\left(\text{Unvaccinated}\right)}, \\ &= 1 - RR, \end{aligned} \tag{1}$$

where R denotes one of the measures of risk and RR is the relative risk, the ratio of vaccinated risk to unvaccinated risk.

The vaccine efficacy (VE) ranges from (-∞,1]. VE of one indicating that the provides 100% protection against the disease, a VE greater than zero indicating that the vaccine is more effective than the control, VE equal to zero indicating that the vaccine is equivalent to the control and VE less than zero indicating that the vaccine is worse than the control. In practice, vaccines are neither perfectly effective nor totally ineffective. HPV vaccine on TVC against cervical cancer and its lesion, for example, is 0.83-0.93 effective when appropriately administered (HPV Patricia Study Group, 2007).

The measure of risk (R) can be a form of the transmission probability, which conditions of exposure to infection, or the incidence rate, hazard rate, or cumulative incidence (attack rate), which do not condition on exposure to infection (Halloran, et al., 2010). In application, the information of exposure to infection is sometimes difficult or impossible to gain. Moreover, the study is designed to estimate VE from events per person-time of potential rather than actual exposure or simply from the proportion of people who become infected in the vaccinated compared to the unvaccinated/control groups. The assumption is made that the vaccinated and unvaccinated groups are equally exposed to infection, so that any differences in the risk in the two groups are due to the biological effects of the vaccine.

Considering the unconditional approach on exposure to infection, the relative risk (RR) can be described as the ratio of cumulative incidence risk (attack rate) of the vaccinated group relative to the unvaccinated/control group under equal follow up time assumption. Meanwhile, under different follow up time assumptions, RR can be described as the ratio of

incidence rates, hazard or (less often) odds of adverse events in the vaccinated group relative to the control group. Different statistical methods were used for each type of RR.

Different parameters of relative risk (RR) to use for estimating VE depends on different information on the type and duration of the study, the infectious agent and its transmission mode, the resources available, and the assumptions of the distribution of protection within the vaccinated group (Halloran, et al., 2010).

Primary vaccine efficacy studies often report VE based on relative events per person time and based on hazard rates using survival analysis methods. Both of them require the time to event as well as the period of exposure (follow up time) of each person under study. When covariates such as age and gender are added, the analysis is modeled by the covariates or Poisson regression can be used. Under the assumption that the affect of the vaccine is multiplicative, constant, and homogeneous, the Cox proportional hazards model can be used to estimate VE and requires only the ordering of the event time. Meanwhile, the cumulative incidence rate (attack rate) requires only final value data in the end of the study.

2.2 Estimator of Weighted Vaccine Efficacy

The heterogeneity of the vaccine efficacy of the two sub-populations related to baseline HPV status, naïve and non-naïve cohorts, need to be taken into account when estimating global vaccine efficacy of the vaccine group compares to the unvaccinated/control group. Several approaches were used in this study using the baseline HPV status for each group as a weighted factor for estimating VE, namely weighted vaccine efficacy. Different weighted vaccine efficacies are estimated based on different parameters of relative risk (RR).

2.2.1 The Attack Rates Ratio Approach

The attack rate (AR) is the ratio of total number of events (cumulative incidence) over some interval time to the total number of people at risk in the beginning of the study. Estimating VE based on the cumulative incidence or attack rates requires only information about whether persons reported or not the event of interest (that can be HPV infection, cervical lesion associated with HPV, disease, …) by the end of the study, that is, final value data.

Consider a cohort study with N_c individuals in the unvaccinated/control group and N_v in the vaccinated group, $N = N_v + N_c$. Let X_c and X_v denote the observed numbers of events at the end of the study between control and vaccinated group, respectively. Then $AR_v = X_v / N_v$, and

$AR_c = X_c / N_c$ are the attack rates for both groups. A measure of vaccine efficacy in this cohort based on attack rates describes as the following expression:

$$\widehat{VE} = 1 - \frac{AR_v}{AR_c}. \tag{2}$$

This measurement is in general sensitive to the indirect effects of vaccination and may not be a biologically interpretable efficacy parameter (Halloran, et al., 1992). Haber et al. (1991) solved a deterministic epidemic model for the transmission rates as a function of the attack rates and obtained the estimate for vaccine efficacy and the variance based on the relative transmission rates that can be obtained from the attack rate as follows:

$$\widehat{VE} = 1 - \frac{\ln\left(1 - AR_v\right)}{\ln\left(1 - AR_c\right)}. \tag{3}$$

This estimate of vaccine efficacy ($\widehat{VE}$) in equation (3) exceed the traditional estimate described in equation (2) when $AR_v < AR_c$. The difference between these quantities decreases as the fraction vaccinated group increases. While the variance of $\widehat{VE}$ approximates using following expression:

$$\widehat{Var}\left(\widehat{VE}\right) \simeq \left[\frac{X_v}{N_v\left(N_v - X_v\right)} + \left(1 - \widehat{VE}\right)^2 \frac{X_c}{N_c\left(N_c - X_c\right)}\right] \Bigg/ \left[\ln\left(\frac{N_c}{N_c - X_c}\right)\right]^2. \tag{4}$$

The estimate of $(1-\alpha)100\%$ confidence interval for VE can be constructed using standard procedure as defined in Appendix A1.

2.2.1.1 Halloran et. al Approach

The weighted average of vaccine efficacy can be derived from modifying the general weighted average as described by Halloran et all (1992) when the stratification of the population appears in both groups, vaccinated and control group.

Let AR_{vk} and AR_{ck} denote the attack rates for k^{th} sub-population (strata) in the vaccine and the control group respectively with corresponding weight α_{vk} and γ_{ck}. The weights defined as the proportion of patients enrolled in k^{th} sub-population for each group vaccine and control, i.e:

$\alpha_{vk} = N_{vk}/N_v$ and $\gamma_{ck} = N_{ck}/N_c$, with the restrictions, $\sum_{k=1}^{K} \alpha_{vk} = 1$ and $\sum_{k=1}^{K} \gamma_{ck} = 1$.

Then, the weighted average of vaccine efficacy can be derived from following expression:

$$\widehat{VE}_{\bullet} = 1 - \frac{\sum_{k=1}^{K} \alpha_{vk} \ln\left(1 - AR_{vk}\right)}{\sum_{k=1}^{K} \gamma_{ck} \ln\left(1 - AR_{ck}\right)}, \tag{5}$$

where k=1, 2, …,K. In this study, K=2 which corresponding to naïve and non-naïve sub-population. The variance of weighted average of vaccine efficacy as derived in Appendix A2 yield as follows:

$$\widehat{Var}\left(\widehat{VE}_{\bullet}\right) = \left(\frac{\sum_{k=1}^{K} \alpha_{vk} \ln\left(1 - AR_{vk}\right)}{\sum_{k=1}^{K} \gamma_{ck} \ln\left(1 - AR_{ck}\right)}\right)^2 \times \left[\frac{\sum_{k=1}^{K} \alpha_{vk}^2 \left(\frac{1}{\left(1 - AR_{vk}\right)}\right)^2 \left(\frac{AR_{vk}\left(1 - AR_{vk}\right)}{N_{vk}}\right)}{\left(\sum_{k=1}^{K} \alpha_{vk} \ln\left(1 - AR_{vk}\right)\right)^2} + \frac{\sum_{k=1}^{K} \alpha_{ck}^2 \left(\frac{1}{\left(1 - AR_{ck}\right)}\right)^2 \left(\frac{AR_{ck}\left(1 - AR_{ck}\right)}{N_{ck}}\right)}{\left(\sum_{k=1}^{K} \alpha_{ck} \ln\left(1 - AR_{ck}\right)\right)^2}\right] \tag{6}$$

The $(1-\alpha)100\%$ confidence interval can be estimate using the same expression as the equations in Appendix A1.

2.2.1.2 General Weighted Approach

The general structure of weighted average of parameter estimate can be used to estimate the weighted average of vaccine efficacy. The expression of weighted average of vaccine efficacy can be defined as follows:

$$\widehat{VE}_{\bullet} = \sum_{k=1}^{K} w_k \widehat{VE}_k. \tag{7}$$

Where $\widehat{VE}_k$ is the estimate of vaccine efficacy for k[th] sub-population that can be derived using the equation (3) and w_k is a weight applied to k[th] sub-population, $w_k = N_k / N$ and $\sum_{k=1}^{K} w_k = 1$. The variance of vaccine efficacy for each sub-population can be estimated using equation (4). The estimate of vaccine efficacy is weighted by the proportion of the number of patients for each sub-population. The variance of weighted vaccine efficacy is defined as:

$$\widehat{Var}\left(\widehat{VE}_{\bullet}\right) = \sum_{k=1}^{K} w_k^2 \widehat{Var}\left(\widehat{VE}_k\right). \tag{8}$$

The confidence interval of weighted vaccine efficacy can be estimated using previous approach as defined in Appendix A1.

2.2.1.3 Precision Based Approach

Haber et al (1991) applied another method of weighted estimate of vaccine efficacy among sub-population using the precision-based approach as follows:

$$\widehat{VE}_{\bullet} = \frac{\sum_{k=1}^{K} \hat{w}_k \widehat{VE}_k}{\sum_{k-1}^{K} \hat{w}_k}, \tag{9}$$

where $\widehat{VE}_k$ is the estimate of vaccine efficacy for k^{th} sub-population that can be derived using the equation (3) and $\hat{w}_k$ is the estimate of weight applied to k^{th} sub-population, $\hat{w}_k = 1/\widehat{Var}\left(\widehat{VE}_k\right)$. The variance of $\widehat{VE}_k$ can be described using the expression in equation (4) . The variance of weighted average vaccine efficacy as derived in Appendix A6 as follows:

$$\widehat{Var}\left(\widehat{VE}_{\bullet}\right) \approx \left(\widehat{VE}_{\bullet}\right)^2 \times \left[\frac{\left(\frac{1}{\sum_{k=1}^{K} \hat{w}_k \widehat{VE}_k}\right)^2 \widehat{Var}\left(\sum_{k=1}^{K} \hat{w}_k \widehat{VE}_k\right) + \left(\frac{1}{\sum_{k-1}^{K} \hat{w}_k}\right)^2 \widehat{Var}\left(\sum_{k-1}^{K} \hat{w}_k\right) - }{2\sqrt{\left(\left(\frac{1}{\sum_{k=1}^{K} \hat{w}_k \widehat{VE}_k}\right)^2 \widehat{Var}\left(\sum_{k=1}^{K} \hat{w}_k \widehat{VE}_k\right)\right) \times \left(\left(\frac{1}{\sum_{k-1}^{K} \hat{w}_k}\right)^2 \widehat{Var}\left(\sum_{k-1}^{K} \hat{w}_k\right)\right)}} \right]. \tag{10}$$

The equations in Appendix A1 can also be applied to estimate the $(1-\alpha)100\%$ confidence interval of weighted average of vaccine efficacy.

This precision-based approach is usually preferable since it provides a narrow interval estimate when there is approximate uniformity of effect over the strata/sub-population (Kleinbaum, et al., 1982). The estimate, weighted by the reciprocal of the variance of each point estimate, gives the highest weight to the most precise estimates. Therefore, the estimate of weighted vaccine efficacy reflects the number of patients in each sub-population (Haber, et al., 1991). This approach assumes the homogeneity of vaccine efficacy among sub-populations. However, between naïve and non-naïve sub-population, the assumption is not always fulfilled.

2.2.2 The Person Time Approach

Person-time approach account for length of follow-up by accumulating time until subjects are diagnosed with the disease or the trial ends, whichever comes first. The vaccine efficacy can be estimated by following expression:

$$\widehat{VE} = 1 - \frac{X_v / T_v}{X_c / T_c}, \tag{11}$$

where T_v and T_c denote as the total follow up time in years at vaccinated and control group respectively. While X_c and X_v denote the observed numbers of event at the end of the study between control and vaccinated group, respectively. The ratio between total follow up time in vaccinated and control group defines as $r = T_v / T_c$ and *p* denotes as the proportion of event in the vaccinated group, $p = X_v / (X_v + X_c)$, thus the equation (11) simply become:

$$\widehat{VE} = 1 - \frac{p}{r(1-p)} \tag{12}$$

The numbers of events in the study follow a Poisson distribution, while the proportion of event in vaccinated group, *p,* follows Binomial distribution.

Dragalin et al (2002) proposed the following expression to estimate the confidence interval (CI) of vaccine efficacy derived from exact CI for *p* using the Clopper-Pearson interval approach:

$$1 - \frac{UL}{r(1-UL)} < \widehat{VE} < 1 - \frac{LL}{r(1-LL)}, \tag{13}$$

where,

$$LL = \left(1 + \frac{2(X_c + 1)}{2X_v F(\alpha/2; 2X_v; 2(X_c + 1))}\right)^{-1}, \tag{14}$$

and,

$$UL = \left(1 + \frac{2X_c}{2(X_v + 1) F(1-\alpha/2; 2(X_v + 1); 2X_c)}\right)^{-1}. \tag{15}$$

F (q; df1; df2) represents the q^{th} quantile from the F-distribution with numerator degrees of freedom df1 and denominator degree of freedom df2.

The variance of vaccine efficacy can be derived based on the log of the ratio of two binomial random variables as defined in Appendix A3 as follows:

$$\widehat{Var}\left(\widehat{VE}\right) \approx \left(\frac{p}{r(1-p)}\right)^2 \left(\frac{1}{X_v}+\frac{1}{T_v}+\frac{1}{X_c}+\frac{1}{T_c}\right). \quad (16)$$

2.2.2.1 Halloran et. al Approach

The weighted average method to estimate of vaccine efficacy using the Halloran et. al approach described in section 2.2.1.1 was applied in this case:

$$\widehat{VE}_{\bullet} = 1 - \frac{\sum_{k=1}^{K} \alpha_{vk}\left(X_{vk}/T_{vk}\right)}{\sum_{k=1}^{K} \gamma_{ck}\left(X_{ck}/T_{ck}\right)}, \quad (17)$$

where k=1, 2, …,K. In this study, K=2 which corresponding to naïve and non-naïve sub population. The weights defined as the proportion of patients enrolled in k^{th} sub-population for each group vaccine and control, i.e:

$\alpha_{vk} = N_{vk}/N_v$ and $\gamma_{ck} = N_{ck}/N_c$, with the restrictions, $\sum_{k=1}^{K}\alpha_{vk}=1$ and $\sum_{k=1}^{K}\gamma_{ck}=1$.

The variance of weighted average of vaccine efficacy as derived in Appendix A4 yield as follows:

$$\widehat{Var}\left(\widehat{VE}_{\bullet}\right) = \left(\frac{\sum_{k=1}^{K} \alpha_{vk}\left(X_{vk}/T_{vk}\right)}{\sum_{k=1}^{K} \gamma_{ck}\left(X_{ck}/T_{ck}\right)}\right)^2 \times \left[\frac{\sum_{k=1}^{K} \alpha_{vk}^{\,2}\left(\left(X_{vk}/T_{vk}\right)^2 \times \left(1/X_{vk}+1/T_{vk}\right)\right)}{\left(\sum_{k=1}^{K} \alpha_{vk}\left(X_{vk}/T_{vk}\right)\right)^2} + \frac{\sum_{k=1}^{K} \gamma_{ck}^{\,2}\left(\left(X_{ck}/T_{ck}\right)^2 \times \left(1/X_{ck}+1/T_{ck}\right)\right)}{\left(\sum_{k=1}^{K} \gamma_{ck}\left(X_{ck}/T_{ck}\right)\right)^2}\right]. \quad (18)$$

2.2.2.2 General Weighted Approach

The general expression of weighted average of vaccine efficacy as described in equation (7) can be applied. Where $\widehat{VE}_k$ is the estimate of vaccine efficacy for k^{th} sub-population that can be derived using the equation (12) and w_k is a weight applied to k^{th} sub-population, $w_k = N_k / N$ and $\sum_{k=1}^{K} w_k = 1$. In this case, the estimate of vaccine efficacy is weighted by the number of patients for each sub-population. The variance of weighted average VE was estimated using equation (8) with the variance of $\widehat{VE}_k$ can be derived from equation (16).

Another approach of choosing the weight factor w_k based on a total follow up time. Since in this study, the follow up time was considered to be relatively important between sub-population, naïve and non- naïve cohort. In the other words, the vaccine efficacy is weighted by the total follow up time in each sub-population as follows:

$$\widehat{VE}_{\bullet} = \sum_{k=1}^{K} w_k \widehat{VE}_k . \tag{19}$$

The weighted factor w_k can be derived as $w_k = T_k / T, where \sum_{k=1}^{K} T_k = T$ and $\sum_{k=1}^{K} w_k = 1$.

The variance of weighted average VE was estimated using equation (8) with the variance of $\widehat{VE}_k$ can be derived from equation (16).

2.2.2.3 Precision Based Approach

The weighted average of vaccine efficacy based on a precision approach described in section 2.2.1.3 can also be applied in this case. The variance of weighted average precision- based approach derived in Appendix A7 as follows:

$$\widehat{Var}\left(\widehat{VE}_{\bullet}\right) = \left(\widehat{VE}_{\bullet}\right)^2 \widehat{Var}\left(\ln \widehat{VE}_{\bullet}\right),$$

$$\approx \left(\widehat{VE}_{\bullet}\right)^2 \left[\left(\frac{1}{\sum_{k=1}^{K} \hat{w}_k \widehat{VE}_k} \right)^2 \widehat{Var}\left(\sum_{k=1}^{K} \hat{w}_k \widehat{VE}_k \right) + \left(\frac{1}{\sum_{k-1}^{K} \hat{w}_k} \right)^2 \widehat{Var}\left(\sum_{k-1}^{K} \hat{w}_k \right) - 2\sqrt{\left(\left(\frac{1}{\sum_{k=1}^{K} \hat{w}_k \widehat{VE}_k} \right)^2 \widehat{Var}\left(\sum_{k=1}^{K} \hat{w}_k \widehat{VE}_k \right) \right) \times \left(\left(\frac{1}{\sum_{k-1}^{K} \hat{w}_k} \right)^2 \widehat{Var}\left(\sum_{k-1}^{K} \hat{w}_k \right) \right)} \right]. \tag{20}$$

2.2.3 The Hazards Ratio

Under the assumption that the effect of the vaccine is multiplicative, constant, and homogeneous, the Cox proportional hazards model can be used to estimate VE and requires only the ordering of the event time (Halloran, et al., 2010). In this case, it is not necessary to estimate the hazard rate in the unvaccinated group, but only the relative hazard rate. The proportional hazards model with covariates can be used to investigate possible confounding factors.

The hazard function is the instantaneous potential of the disease, i.e., at a specified moment in time after vaccination, *t*, given that a patient has been disease free up to time *t*; the time

between vaccination and the events. Monto et al (2001) applied a Cox proportional hazard model to estimate vaccine efficacy of influenza. The hazard ratio can be described as: $\widehat{HR} = \exp\left(\hat{\beta}\right)$,thus the estimate of vaccine efficacy is defined as:

$$\widehat{VE} = 1 - \widehat{HR} = 1 - \exp\left(\hat{\beta}\right), \tag{21}$$

where $\hat{\beta}$ is the partial likelihood estimate of the log hazard ratio of the group effect. The variance of vaccine efficacy can be estimated from the variance hazard ratio as follows:

$\widehat{Var}\left(\widehat{HR}\right) = \left(\widehat{HR}\right)^2 \widehat{Var}\left(\beta\right)$, thus yield $\widehat{Var}\left(\widehat{VE}\right) = \widehat{Var}\left(1 - \widehat{HR}\right) = \widehat{Var}\left(\widehat{HR}\right)$.

The confidence interval for VE can be constructed using parameter estimate and its variance.

The weighted average of vaccine efficacy is estimated using the precision based approach by Haber et al (1991) as described in equation (9). The variance of weighted average VE based on precision approach for proportional hazard model was not derived in this study due to numerical complexity.

Moreover, the general weighted average approach discussed in section 2.2.2.2 can also be applied. The vaccine efficacy is weighted by the proportion number of patients for each sub-population, naïve and non-naïve cohorts. The weight factor w_k based on a total follow up time is also considered.

2.2.4 The Incidence Rates Ratio- Poisson regression

This approach was the extension of the person-time approach from previous list. When covariates such as age and gender are added, the analyses are stratified by the covariates or Poisson regression can be used (Halloran, et al., 2010).The number of events assumed to follow Poisson distribution and model the mean of the number of events (expected value) described as follows:

$$\log\left(\lambda\right) = \mathbf{X}^{'}\boldsymbol{\beta} + \log(T) \tag{22}$$

Where λ is the mean of the number of adverse events, **X** are covariates (vaccine/control, age, etc), T is the follow up time and **β** are coefficients estimated by the model. The offset variable log(T) is needed to account for possible different observation periods for different patients. Poisson regression model with quasi likelihood adjustment of the standard error was used in order to handle over dispersion issue. In this case, the relative risk (RR) is the incidence rate ratio, $\widehat{RR} = \exp\left(\hat{\beta}\right)$, where $\hat{\beta}$ correspond to the parameter estimate for the covariate of

vaccine status (vaccine/control). The vaccine efficacy can be directly estimated from following equation:

$$\widehat{VE} = 1 - \widehat{RR} = 1 - \exp\left(\hat{\beta}\right) \tag{23}$$

The variance of vaccine efficacy can be estimate from variance hazard ratio as follows:

$\widehat{Var}\left(\widehat{RR}\right) = \left(\widehat{RR}\right)^2 \widehat{Var}\left(\hat{\beta}\right)$, thus yield $\widehat{Var}\left(\widehat{VE}\right) = \widehat{Var}\left(1 - \widehat{RR}\right) = \widehat{Var}\left(\widehat{RR}\right)$.

The confidence interval for VE can be constructed using parameter estimate and its variance. The weighted average of vaccine efficacy is estimated using the precision-based approach by Haber et al (1991) as describe in equation (9). However, due to numerical complexity the variance of weighted average VE based on precision approach for Poisson regression model was not derived in this study. Another weighted average approach using weighted factor (w_k) as the proportion number of patients for each sub-population as well as total follow up time were applied.

2.3 Example

The overview of the application of weighted average VE approaches that described in the previous section show in the following example:

The HPV vaccine study for 5000 patients was randomly selected among the population. The patients were divided into two treatment groups with the same proportion. The baseline HPV status, naïve and non-naïve HPV, was checked before patients enter the study. The following table shows the classification of the patients in the study.

Table 1. Classification of the Patients

HPV Status		Naïve				Non-Naïve				Total Patients
		# Patients			Total Follow Up Time (Year)	# Patients			Total Follow Up Time (Year)	
		Event	No Event	Total		Event	No Event	Total		
Treatment Group	Vaccine	12	1711	1723	6867.891	10	733	743	2958.936	2466
	Control	36	1683	1719	6822.150	38	777	815	3187.665	2534
Total Patients				3442				1558		5000

The estimate of vaccine efficacy using the methods from previous section shows in the following table:

Table 2. The estimate of vaccine efficacy

RR Approach	VE Approach	VE	95% CI				coverage
Attack Rate (AR)	raw	0.857	(	0.752	;	0.922	)	95.700
	we_N	0.868	(	0.766	;	0.929	)	95.700
	we_hal	0.857	(	0.752	;	0.922	)	96.800
	we_pres	0.883	(	0.776	;	0.941	)	97.684
Person Time Approach	raw	0.857	(	0.742	;	0.926	)	96.400
	we_N	0.867	(	0.691	;	0.951	)	99.496
	we_t	0.867	(	0.691	;	0.951	)	99.496
	we_hal	0.856	(	0.752	;	0.921	)	96.878
	we_pres	0.883	(	0.720	;	0.959	)	98.389
Proportional Hazard Model	raw	0.857	(	0.743	;	0.919	)	94.400
	we_N	0.867	(	0.691	;	0.941	)	99.295
	we_t	0.867	(	0.691	;	0.941	)	99.295
	we_pres	0.884	(	0.713	;	0.950	)	99.200
Poisson Regression	raw	0.857	(	0.816	;	0.888	)	60.900
	we_N	0.867	(	0.808	;	0.908	)	79.859
	we_t	0.867	(	0.808	;	0.908	)	79.557
	we_pres	0.890	(	0.833	;	0.923	)	63.400

VE approach: raw: Raw of vaccine efficacy, we_hal: weighted average vaccine efficacy based on Halloran approach, we_N: general weighted average vaccine efficacy based on sample size for each sub-population , we_pres : weighted average vaccine efficacy based on precision approach, we_t: general weighted average vaccine efficacy based on total follow up time for each sub-population.

3 Simulation Study

Different methods of estimating the weighted average vaccine efficacy (VE) related to different relative risk (RR) described in previous sections was applied in the simulation. The simulation was done in a population partitioned into two sub-populations of baseline HPV status, naïve and non-naïve cohorts. The estimate of weighted average vaccine efficacy, together with standard error, confidence interval (CI) and CI coverage as well as CI length was calculated from simulated data. The comparison among estimator of different methods was done under specific scenario.

3.1 Data Generation Method

The simulation was conducted under 3×3×2×2 conditions to evaluate the performance of different methods of weighted average vaccine efficacy. Most clinical trials for HPV vaccine contained a balanced design of two treatment groups, vaccinated and unvaccinated/control group. For each group of treatment, the different baseline HPV status, naïve and non-naïve cohort, was appearing in the population.

The duration of the follow-up was fixed at 4 years for all subjects. No dropped-out were considered in this study. The follow-up time for the estimation of vaccine efficacy was defined as the minimum between the time to the event and maximum follow up time (4 years). If there is no event during entire follow-up time period for a particular patient, the follow up time for that of patient was defined as the maximum follow up time (4 years). Only one event per subject was considered during the entire follow-up of the study.

The time to event was defined follows exponential distribution. Since it describes the time between events in a Poisson process, i.e. a process in which events occur continuously and independently at a constant average rate and assumed that the random external event are causing the adverse event. The finding mean of follow up time in previous study was 34.9 months (2.9083 year) (HPV Patricia Study Group, 2009). Thus, in this study the time of event was generated follows exponential distribution with parameter λ equal 0.34384, $\lambda = 1/mean(time)$.

Other scenarios for generating the data are considered in this study describe in the following setting:

Data was randomly assign to *n* patients for balance design of two arms study, vaccinated and unvaccinated group, where n=2,500 patients , n=5,000 patients, and n=10,000 patients.

1. For each treatment group, the different proportion of the baseline HPV status, naïve and non-naïve cohort was considered in the study. The ratio between sub-population naïve and non-naïve defined as follow: 50/50, 70/30 and 90/10.
2. The incidence rate (IR) of events in the naïve sub-population was assumed to be lower than those in the non-naïve sub-population, since HPV vaccine works better for naïve population compared to non-naïve population (FDA, 2006). The incidence rate of events in the control group (IRc) was defined in the beginning. The incidence for naïve sub-population in the control group is made to be fixed (0.03), while in the non-naïve sub-population was defined equal to 0.05 and 0.1.
3. The vaccine efficacy in the naïve sub-population was assumed to be higher than those in the non-naïve sub-population. The vaccine efficacy (VE) in the vaccine group for naïve sub-population was made to be fixed (0.9), while for non-naïve sub-population was defined equal to 0.8 and 0.5. Thus, the incidence rate of events in the vaccinated group (IRv) simply derive from IR control group and VE vaccine group as defined in equation (1) as follows:

$$VE = 1 - \frac{IR_v}{IR_c}, \text{ thus } IR_v = (1 - VE)\, IR_c$$

One thousand simulations were run under 36 conditions above for each weighted average method. The simulation process was done using SAS 9.2 software. The codes are given in the Appendix C. The generated data structure shows in the following table:

Table 3. The Structure of generated data

J	Group	Naive	Case	Censored	Time
1	1	1	0	1	4.000
2	0	1	1	0	0.751
3	0	1	0	1	4.000
4	0	0	1	0	1.356
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
n	0	0	1	0	0.043

The generated data contained variable group with indicator 1 defines as the vaccinated group, while 0 defines as unvaccinated/control group. The variable naïve with indicator 1 defines as

naïve sub-population and 0 for non-naïve. The variable case with indicator 1 defines if the event occurs during the study and 0 when the event doesn't occur. Moreover, the variable time is the event time if the j^{th} patient has the event during study and variable censored is defined as 0, otherwise the time is the maximum follow up time if the j^{th} patient doesn't has the event during study and variable censored is defined as 1.

3.2 Methods Comparison

Thousand simulations were run to completion for combinations of criteria on each weighted average VE methods. The sample mean and standard deviation of vaccine efficacy, mean standard error, mean squared error (MSE), mean of length for CI as well as the coverage for CI of vaccine efficacy were computed from these simulations.

The standard deviation of sample VE was computed in order to see how closely individuals within the sample differ from the sample mean of vaccine efficacy, while standard error is to see how close to the population mean, the sample mean of VE is likely to be.

MSE was computed to quantify the different between values implied by an estimator and the true value of quantifying being estimated. MSE assesses the quantify of an estimator in term of its variation and biasedness. The closer MSE to zero the better parameter estimate is likely to gain.

The coverage of CI was computed in order to check the validity of CI. Since in the study the confidence interval was calculated based on nominal 95% CI, thus the actual coverage was expected to be around 95%. In other words, about 95% intervals were computed during the simulation process should contain the true value.

3.2.1 The Attack Rate Ratio Approach

The first sets of simulation when relative risk (RR) as attack rates approach present in Appendix B1 (Table B1.1-B1.3). The estimate of vaccine efficacy based on attack rate (AR) approach with different set criteria of sample size (n), the proportion of the number of patients in naïve and non-naïve sub-population (Ratio naïve: non-naïve), the incidence rate in the control group for naïve and non-naïve (IRc naïve: non-naïve) and the vaccine efficacy for naïve and non-naïve (VE naïve: non-naïve).

The tables describe the raw vaccine efficacy (Equation (3)), the estimate of vaccine efficacy that do not take into account the heterogeneity of the population; the weighted average vaccine efficacy using Halloran approach (Equation (5)); general weighted average vaccine

efficacy based on sample size for each sub-population (Equation (7)); and weighted average vaccine efficacy based on a precision approach (Equation (9)). It shows that different criteria give different estimates of vaccine efficacy. On average, under raw VE the estimate of vaccine efficacy is less close to the true value compared to weighted average vaccine efficacy. Among three approaches of weighted average vaccine efficacy, the estimate of weighted average VE based on Halloran approach and general weighted average VE based on sample size seems closer to the real value compared those on precision approach. The true values were defined in equations (32) and (33) (Appendix A5).

The standard deviation of the sample of estimate VE (std VE) is very close to the mean standard error VE (se VE), meaning that the deviation of estimate VE is close to the true value of standard deviation. The raw VE, weighted average VE based on Halloran approach and weighted average VE based on sample size produce a standard deviation for sample estimate VE closer to the mean standard error VE compared to those on precisions based approach. Moreover, the MSE value in average close to zero, raw VE has the highest MSE value compared to weighted average VE approach. While among weighted average VE, the weighted average VE precision based approach has the highest MSE. Meanwhile, general weighted average VE based sample size for each sub-population gave the smallest MSE. These results were in line with the bias measurement results.

Table 4. Percent Coverage CI for Attack Rate (AR) Ratio Approach

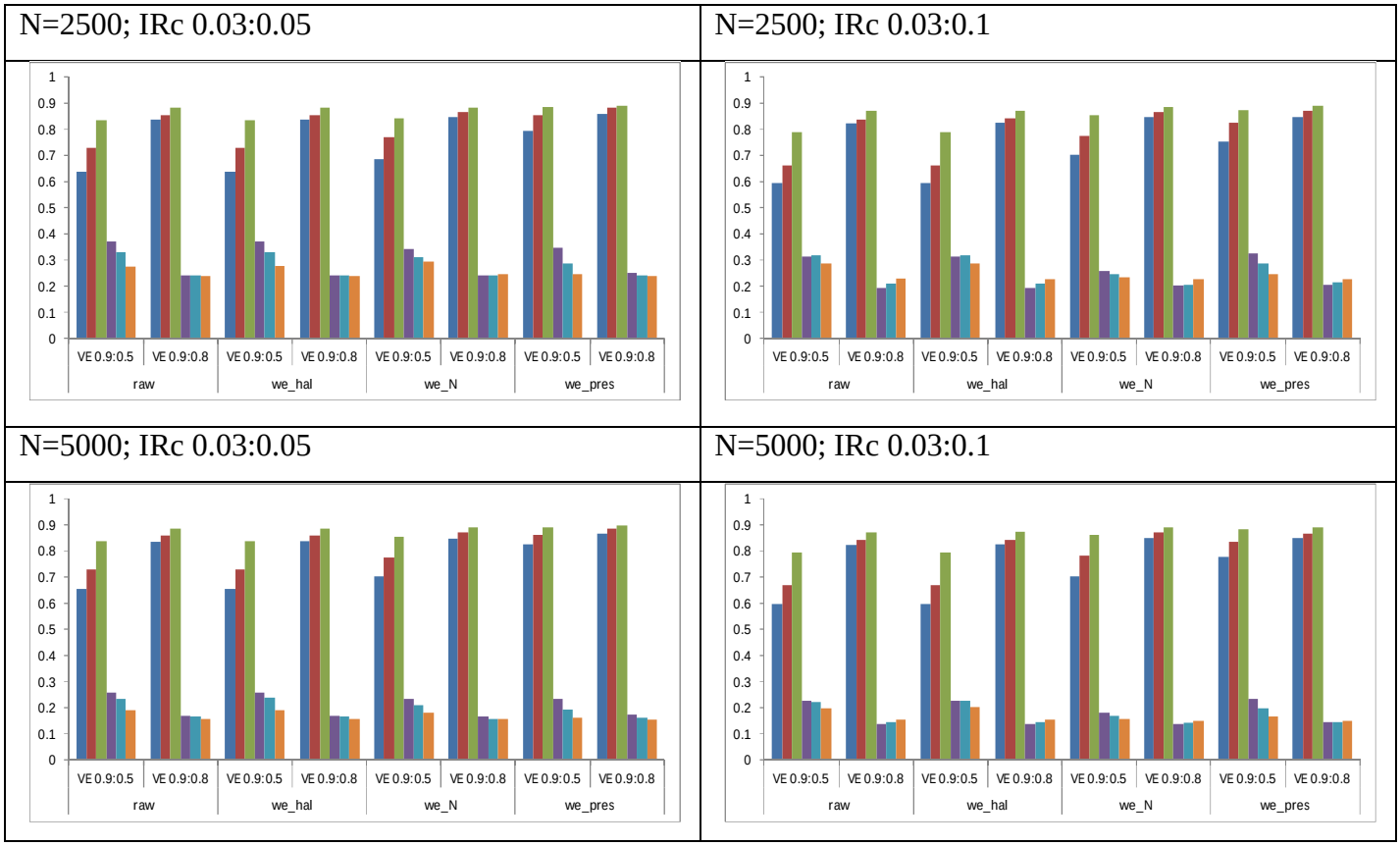
VE Approach	n	Ratio 50:50				Ratio 70:30				Ratio 90:10			
		VE 0.9:0.5		VE 0.9:0.8		VE 0.9:0.5		VE 0.9:0.8		VE 0.9:0.5		VE 0.9:0.8	
		IRc 0.03:0.05	IRc 0.03:0.10	IRc 0.03:0.05	IRc 0.03:0.10	IRc 0.03:0.05	IRc 0.03:0.10	IRc 0.03:0.05	IRc 0.03:0.10	IRc 0.03:0.05	IRc 0.03:0.10	IRc 0.03:0.05	IRc 0.03:0.10
raw	2500	96.600	79.500	96.100	92.700	95.000	71.400	95.800	92.400	96.500	84.400	97.065	94.800
	5000	92.300	57.400	94.700	89.000	90.000	46.800	94.700	89.200	94.200	72.900	96.400	92.600
	10000	87.400	26.300	94.100	84.500	82.400	15.400	94.800	82.100	89.300	45.500	94.700	90.400
we_hal	2500	97.100	94.800	96.200	95.700	95.900	96.200	97.000	95.100	97.800	96.500	97.773	97.100
	5000	95.600	95.800	95.800	94.300	95.900	95.900	95.100	95.700	96.800	96.100	96.700	95.400
	10000	96.200	94.400	96.100	95.000	93.700	95.600	96.800	93.800	94.300	94.700	95.500	95.500
we_N	2500	96.900	93.800	96.100	93.600	96.700	95.300	96.400	93.900	98.100	95.600	96.954	96.000
	5000	95.200	94.800	95.800	94.800	96.100	94.300	95.800	94.400	96.500	94.600	96.900	94.600
	10000	95.900	94.100	96.500	94.200	94.000	94.900	96.700	93.500	94.600	95.600	95.300	95.000
we_pres	2500	79.800	81.700	83.800	84.200	86.600	91.900	90.900	89.800	92.600	94.900	98.678	88.400
	5000	55.700	78.100	96.000	91.900	68.600	81.100	95.900	94.500	90.300	91.700	90.000	94.400
	10000	30.000	59.600	93.500	93.100	46.000	68.200	94.100	91.900	83.700	88.600	94.100	94.300

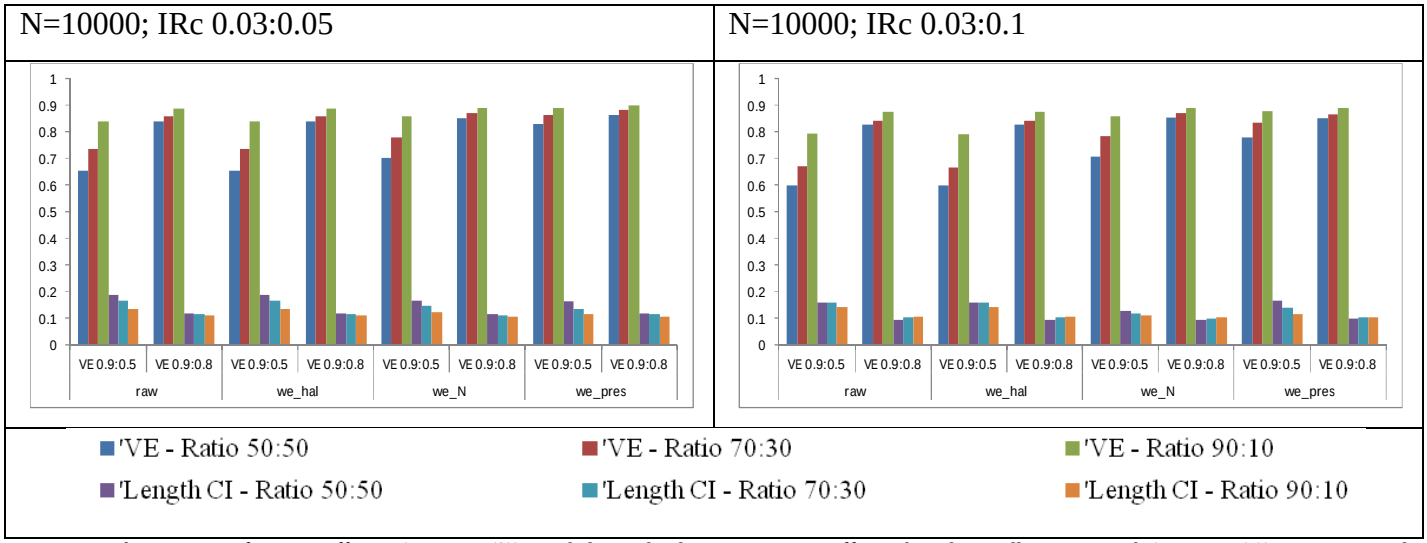
VE approach: raw: Raw of vaccine efficacy (Equation (3)), we_hal: weighted average vaccine efficacy based on Halloran approach (Equation (5)), we_N: general weighted average vaccine efficacy based on sample size for each sub-population (Equation (7)), we_pres : weighted average vaccine efficacy based on precision approach (Equation (9)).

Table 4 shows the percent coverage for each criterion in the simulation process. It shows that the coverage of raw VE is smaller compared to weighted average VE. The raw VE provided very low actual coverage for the combination of IRc 0.03:0.10 with VE 0.9:0.5. In overall, the actual coverage for raw VE is far from 95% that would have been expected.

Moreover, among the weighted average VE, the weighted average VE under precision approach has the lowest coverage and the farthest from 95% especially in the combination of VE 0,9:0.5 with others criterion. It shows that the coverage of VE 0,9:0.5 is smaller than those on VE 0.9:0.8. The weighted average VE based on the precision-based approach show permissive coverage with many of the nominally 95% CI capturing less than 95%. The weighted average VE based on Halloran et. al approach as well as general weighted average VE based on sample size provide better actual coverage than other approaches, the actual coverage probability is around the nominal 95% CI.

Figure 1. Estimate of VE and Length CI for RR as Attack Rate Ratio





VE approach: raw: Raw of vaccine efficacy (Equation (3)), we_hal: weighted average vaccine efficacy based on Halloran approach (Equation (5)), we_N: general weighted average vaccine efficacy based on sample size for each sub-population (Equation (7)), we_pres : weighted average vaccine efficacy based on precision approach (Equation (9)).

From figure 1 above show that the estimate of VE and length of CI based on the relative risk (RR) as Attack Rate (AR) ratio approach. In all settings, the estimator of VE and length of CI based on sample size (n) set to have specifications on 2,500; 5,000 and 10,000 performed similarly. The estimate of VE slightly increases along with the decreasing the distance between the estimate and true value when sample size increase. However, the length slightly decreases with increasing sample size.

For the incidence rate of the control group (IRc) in naïve and non-naïve sub-population, the estimate of VE is slightly higher in IRc 0.03:0.05 compared to IRc 0.03:0.1 with the same pattern for all VE approach (raw VE and weighted average VE). Meanwhile, the estimate of VE based on defined VE setting on naïve and non-naïve sub-population for specification VE 0.9:0.8 is higher for all VE approach (raw VE and weighted average VE) compared to VE 0.9:0.5. Moreover, the estimate of VE based on proportion number of patients in each sub-population naïve and non-naïve (Ratio naïve:non-naïve) is increasing with the increasing of proportion number of patients in naïve sub-population as well as the decreasing of proportion number of patients in non-naïve sub-population.

3.2.2 The Person Time Approach

The second sets of simulation presented in Appendix B2 (Table B2.1-B2.3), the estimate of vaccine efficacy based on person time approach with different set criteria as described in the previous section. The tables describe raw vaccine efficacy (Equation (12)); the weighted average vaccine efficacy using Halloran approach (Equation (17)); general weighted average vaccine efficacy based on sample size for each sub-population (Equation (7)); general

weighted average vaccine efficacy based on total follow up time for each sub-population (Equation (19)); and weighted average vaccine efficacy based on precision approach (Equation (9)). It shows that different criteria produce different estimates of vaccine efficacy. The true value was defined in equation (32) and (34) (Appendix A5).

The estimate of raw VE is less close to the true value compared to weighted average vaccine efficacy. Among three approaches of weighted average vaccine efficacy, the estimate of weighted average VE under precision-based approach is less close to the true value, meanwhile, estimate VE under Halloran approach is the closer to the true value. On average, between two general weighted average VE approaches, the estimate of VE weighted by the sample size is closer to the true value than those weighted by total follow up time approach.

In overall, the standard deviation for the sample of estimate VE (std VE) is very close to the mean standard error VE (se VE). The estimate of weighted average VE based on precision approach produce standard deviation of the sample mean farther to the mean standard error compared to those other approaches, while RAW VE produce the closest one. However, the MSE value in average close to zero, raw VE has the highest MSE value compared to weighted average VE approach. The weighted average VE based on precision give the highest value of MSE among the weighted average VE approaches, while general weighted average VE based on total follow up time is the lowest one.

Table 5. Percent Coverage for Person Time Approach

VE Approach	n	Ratio 50:50				Ratio 70:30				Ratio 90:10			
		VE 0.9:0.5		VE 0.9:0.8		VE 0.9:0.5		VE 0.9:0.8		VE 0.9:0.5		VE 0.9:0.8	
		IRc 0.03:0.0	IRc 0.03:0.1	IRc 0.03:0.0	IRc 0.03:0.1	IRc 0.03:0.0	IRc 0.03:0.1	IRc 0.03:0.0	IRc 0.03:0.1	IRc 0.03:0.0	IRc 0.03:0.1	IRc 0.03:0.0	IRc 0.03:0.1
raw	2500	92.400	75.500	96.200	93.600	93.700	69.300	96.200	93.100	96.700	85.900	97.874	95.700
	5000	90.100	54.300	95.500	89.900	89.300	45.900	95.300	90.100	95.100	75.800	96.900	94.100
	10000	86.300	24.700	94.900	85.500	82.500	15.300	95.400	83.800	90.600	48.200	95.800	91.700
we_hal	2500	82.900	80.000	82.700	83.700	88.900	90.400	90.000	88.100	91.100	93.500	97.654	88.000
	5000	93.400	94.500	94.700	92.100	95.600	95.600	94.800	95.100	96.600	96.300	90.100	94.900
	10000	96.200	94.000	96.100	94.100	93.800	95.000	96.800	94.500	94.300	94.900	95.500	95.900
we_N	2500	84.400	83.000	85.000	86.700	91.300	93.500	92.300	91.800	93.200	95.900	99.707	89.500
	5000	97.000	98.100	98.000	96.600	99.700	99.800	99.200	98.600	99.400	99.600	92.400	98.200
	10000	99.800	99.400	99.800	99.700	99.000	99.800	99.600	99.700	99.500	99.600	99.400	99.400
we_pres	2500	81.800	81.600	85.000	86.000	86.100	91.300	92.300	91.700	92.600	95.900	99.707	90.000
	5000	69.500	76.900	97.800	96.000	77.900	83.700	98.200	97.400	93.500	93.900	91.500	97.400
	10000	46.200	65.200	96.500	96.500	59.900	75.300	97.900	98.600	89.600	92.900	97.500	98.700
we_t	2500	84.400	83.100	85.000	86.700	91.300	93.500	92.300	91.800	93.200	95.900	99.707	89.500
	5000	97.100	98.100	98.100	96.500	99.700	99.800	99.200	98.600	99.400	99.500	92.400	98.200
	10000	99.800	99.200	99.800	99.700	99.000	99.800	99.600	99.700	99.400	99.500	99.400	99.300

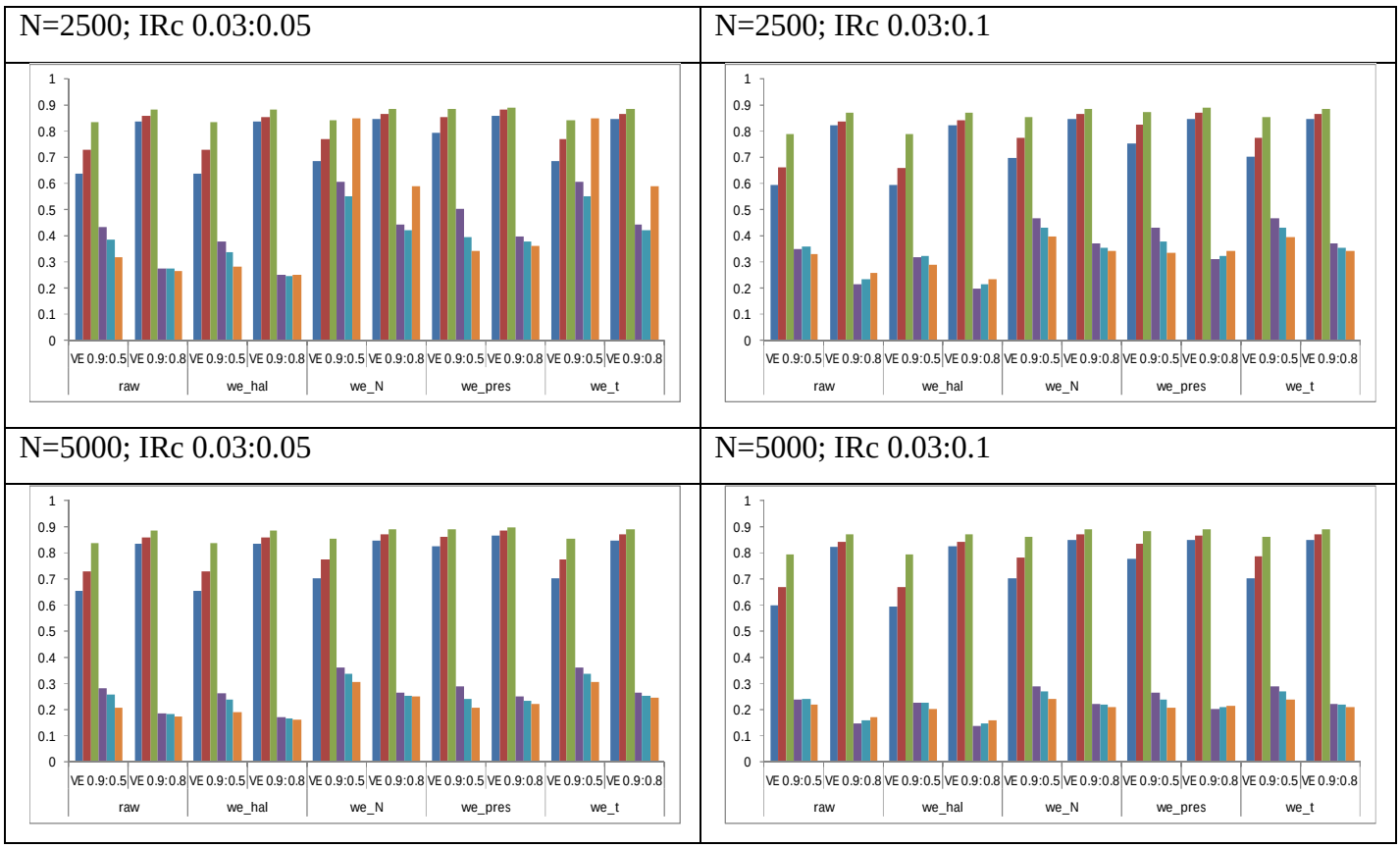
VE approach: raw: Raw of vaccine efficacy (Equation (12)), we_hal: weighted average vaccine efficacy based on Halloran approach (Equation (17)), we_N: general weighted average vaccine efficacy based on sample size for each sub-population (Equation (7)), we_pres : weighted average vaccine efficacy based on precision approach (Equation (9)), we_t: general weighted average vaccine efficacy based on total follow up time for each sub-population (Equation (19)).

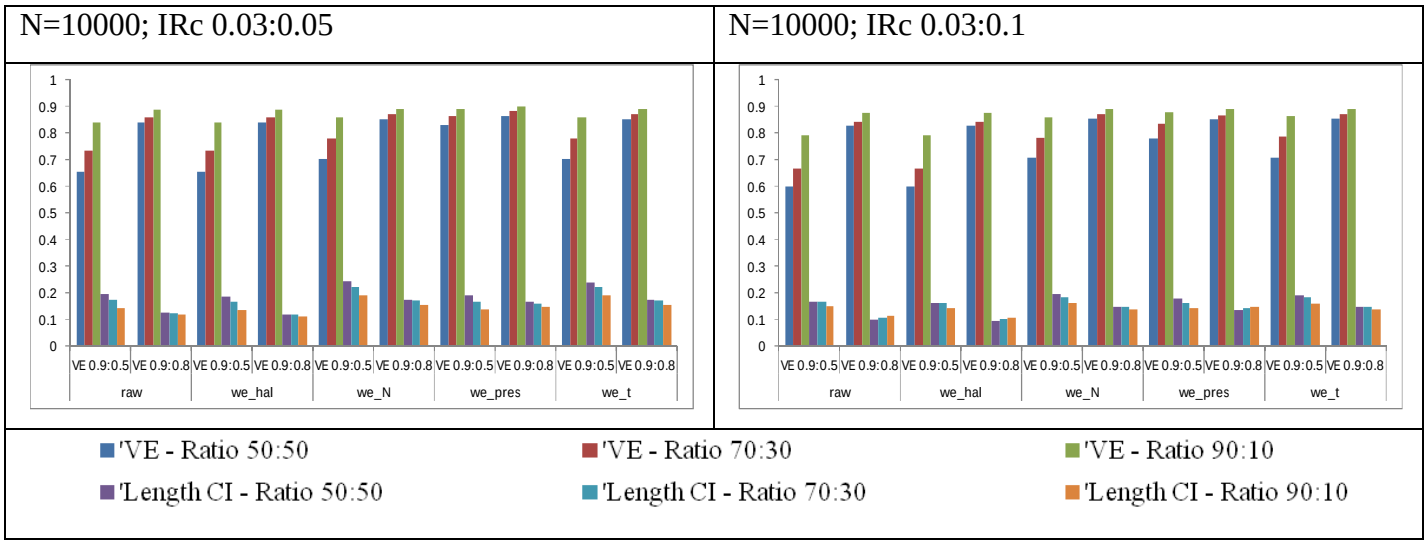
The percent coverage for person time approach provide in Table 5. The coverage of raw VE is smaller than the weighted average VE. While among the weighted average VE, the weighted average VE under precision approach has the lowest coverage.

The raw VE approach and the weighted average VE based on the precision approach show permissive coverage with many of the nominally 95% CI capturing less than 95%. For raw VE, smaller actual coverage was gained especially in the combination of IRc 0.03:0.10 with VE 0.9:0.5, while for weighted average VE based on precision approach in the combination of VE0.9:0.5 with others.

Moreover, the general weighted average VE based on sample size for each sub-population as well as total follow up time show conservative coverage, since many of the nominally 95% CI capture over 95% coverage, however, in average both of methods gave actual coverage around 95%. Moreover, the weighted average VE based on Halloran et. al approach captures the actual coverage probability close to the nominal 95%.

Figure 2. Estimate of VE and Length CI for RR as Person Time Approach





VE approach: raw: Raw of vaccine efficacy (Equation (12)), we_hal: weighted average vaccine efficacy based on Halloran approach (Equation (17)), we_N: general weighted average vaccine efficacy based on sample size for each sub-population (Equation (7)), we_pres : weighted average vaccine efficacy based on precision approach (Equation (9)), we_t: general weighted average vaccine efficacy based on total follow up time for each sub-population (Equation (19)).

The estimate of VE based on the relative risk (RR) as person time approach show in figure 2 seem to have the same pattern for each setting of sample size. However, the distance between the estimate of VE and true value decrease with sample size increase for all VE approach, raw VE and weighted average VE. The lengths of CI also decrease as sample size increase.

Increasing the incidence rate of the control group (IRc) in non-naïve sub-population produced a slightly lower estimate of VE for all VE approaches. The estimate of VE based on defined VE setting on naïve and non-naïve sub-population for VE 0.9:0.8 is higher than VE 0.9:0.5 for all VE approach. Moreover, the estimate of VE is increasing as the ratio of population in naïve increase.

3.2.3 Hazard Ratio Approach

The third sets of simulation presented in Appendix B3 (Table B3.1-B3.3) using Cox Proportional Hazard model in order to estimate vaccine efficacy with different set criteria as described in the previous section.

The tables contain the estimate of VE using different approaches of VE: the raw vaccine efficacy (Equation (21)); general weighted average vaccine efficacy based on sample size for each sub-population (Equation (7)); general weighted average vaccine efficacy based on total follow up time for each sub-population (Equation (19)); and weighted average vaccine efficacy based on precision approach (Equation (9)).

The estimate of raw VE and weighted average VE precision based approach are less close to the true value compared to general weighted average vaccine efficacy based on sample size and total follow up time. The true values of vaccine efficacy define as equations in equations (32) (Appendix A5). Meanwhile, the estimate of VE for two approaches of general weighted average VE, based on sample size for each sub-population and total follow up time, is almost the same. However, raw VE and weighted average VE based on precision approach have the highest MSE value compared to both general weighted average VE approaches. The MSE for both general weighted average approaches produce almost the same value.

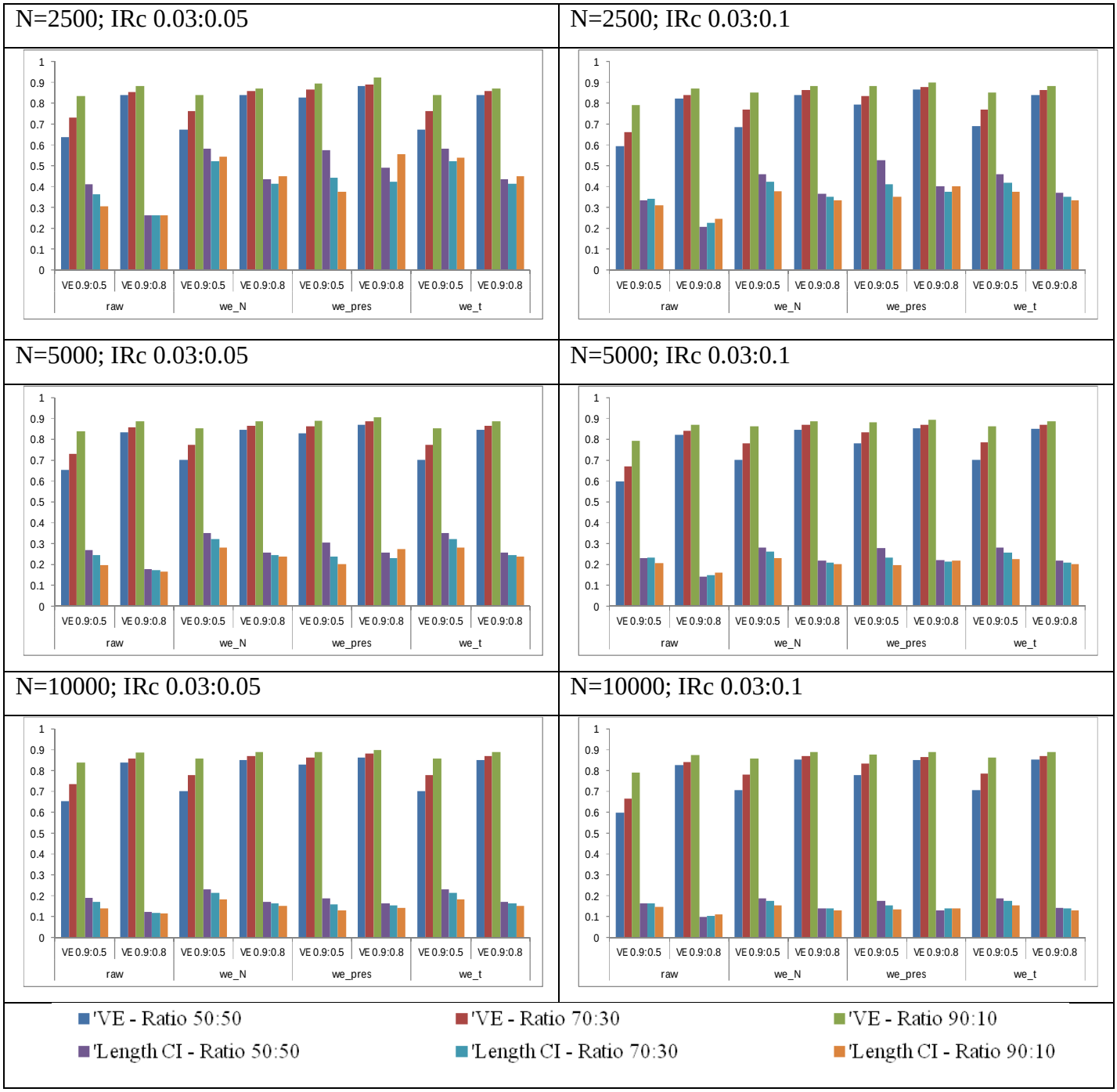
Table 6. Percent Coverage CI for Attack Proportional Hazard Approach

VE Approach	n	Ratio 50:50				Ratio 70:30				Ratio 90:10			
		VE 0.9:0.5		VE 0.9:0.8		VE 0.9:0.5		VE 0.9:0.8		VE 0.9:0.5		VE 0.9:0.8	
		IRc 0.03:0.0	IRc 0.03:0.1	IRc 0.03:0.0	IRc 0.03:0.1	IRc 0.03:0.0	IRc 0.03:0.1	IRc 0.03:0.0	IRc 0.03:0.1	IRc 0.03:0.0	IRc 0.03:0.1	IRc 0.03:0.0	IRc 0.03:0.1
raw	2500	90.200	71.300	94.595	90.900	91.000	62.800	95.281	90.000	94.200	79.200	96.360	93.065
	5000	88.100	49.800	94.400	87.100	85.700	41.400	94.200	87.300	92.800	70.400	95.800	90.800
	10000	84.300	22.800	93.300	82.400	78.800	13.100	94.000	80.200	87.600	41.900	94.500	89.100
we_N	2500	98.822	99.400	99.647	99.194	99.235	99.148	98.919	99.024	99.678	98.857	98.680	98.113
	5000	99.486	99.694	99.490	99.586	99.599	99.599	99.396	99.494	99.298	99.499	98.921	98.384
	10000	99.600	99.299	99.900	99.800	99.099	99.800	99.500	99.700	98.900	99.500	99.499	98.900
we_pres	2500	99.500	99.200	99.800	98.800	98.600	99.300	99.598	98.900	99.900	99.400	99.292	99.497
	5000	74.200	81.200	99.700	98.800	80.700	85.200	99.300	98.600	95.000	94.800	98.800	98.600
	10000	46.000	65.600	97.000	96.500	59.800	75.700	98.000	97.600	89.400	93.200	97.800	98.200
we_t	2500	98.940	99.520	99.647	99.194	99.235	99.148	98.919	99.024	99.678	98.857	98.680	98.113
	5000	99.486	99.796	99.490	99.586	99.599	99.699	99.396	99.494	99.298	99.499	98.921	98.283
	10000	99.600	99.299	99.900	99.800	99.099	99.700	99.400	99.700	98.900	99.400	99.499	98.900

VE approach: raw: Raw of vaccine efficacy (Equation (21)), we_N: general weighted average vaccine efficacy based on sample size for each sub-population (Equation (7)), we_pres : weighted average vaccine efficacy based on precision approach (Equation (9)), we_t: general weighted average vaccine efficacy based on total follow up time for each sub-population (Equation (19)).

Table 6 shows that the coverage of raw VE is smaller and farther away from 95% that would have been expected compared to weighted average VE especially in the combination of IRc 0.03:0.1 and VE 0.9:0.5. The weighted average VE based on precision approach show permissive coverage with many of the nominally 95% CI capturing less than 95% in the combination of VE 0.9:0.5 and ratio 50:50 as well as ratio 70:30. Meanwhile, the general weighted average VE based on sample size for each sub-population as well as total follow up time show conservative coverage, since many of the actual coverage capture over 95%.

Figure 3. Estimate of VE and Length CI for RR as Hazard Rate Ratio



VE approach: raw: Raw of vaccine efficacy (Equation (21)), we_N: general weighted average vaccine efficacy based on sample size for each sub-population (Equation (7)), we_pres : weighted average vaccine efficacy based on precision approach (Equation (9)), we_t: general weighted average vaccine efficacy based on total follow up time for each sub-population (Equation (19)).

Figure 3 present the estimate of VE and length of CI based on the relative risk (RR) as hazard ratio approach. The distance between the estimate of VE and true value is decreasing with sample size increase. Moreover, the length also decreases with increasing of sample size.

For the incidence rate of the control group (IRc) in naïve and non-naïve sub-population, the estimate of VE is slightly higher in IRc 0.03:0.05 compared to IRc 0.03:0.1. The estimate of VE based on defined VE setting on naïve and non-naïve sub-population for VE 0.9:0.8 is higher than VE 0.9:0.5 for all VE approach. The estimate of VE in Ratio 90:10 is higher than those on Ratio 70:30 and Ratio 50:50. The estimate of VE is increasing as the proportion number of population in naïve sub-population increase. However, the estimate of raw VE is lower than the estimate of the weighted average VE approaches. Moreover, the estimate of the weighted average of VE under precision approach give the highest value, while the estimate of both general weighted average VE approaches has almost the same estimate VE.

3.2.4 Poisson Regression Approach

The last sets of simulation presented in Appendix B4 (Table B4.1-B4.3) using Poisson Regression model in order to estimate relative risks (RR) as a way to estimate vaccine efficacy with different set criteria as explained in the previous section.

The estimate of VE using different approaches of VE was considered in the analysis: the raw vaccine efficacy (Equation (23)); general weighted average vaccine efficacy based on sample size for each sub-population (Equation (7));general weighted average vaccine efficacy based on total follow up time for each sub-population (Equation (19)); and weighted average vaccine efficacy based on precision approach (Equation (9)).

Under raw VE and weighted average VE based on precision, the estimate of vaccine efficacy is farther to the true value compared to general weighted average vaccine efficacy. Both general weighted average VE approaches based on sample size and total follow up time produce almost the same estimate of VE and have smallest distance to the true value. The true value of vaccine efficacy was defined in equations (32) (Appendix A5).

The estimate of raw VE produced a standard deviation for sample estimate VE (std VE) slightly farther to the mean standard error (se VE) compared to those on both general weighted average approaches. However, the MSE value in average close to zero. The precision based approach has the highest MSE followed by raw VE. Meanwhile, both general weighted averages VE have almost the same MSE, yet general weighted average VE based on follow up time approach has a slightly higher MSE than those on general weighted average VE based on sample size approach.

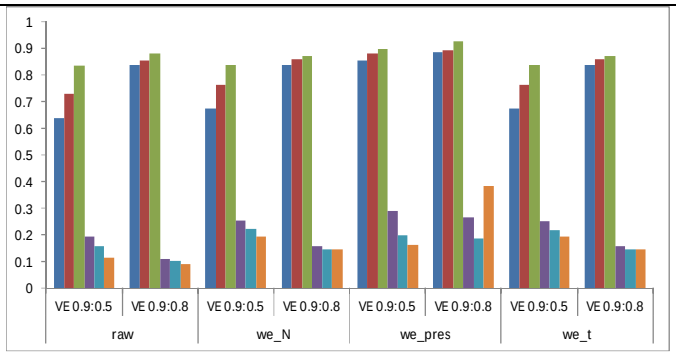
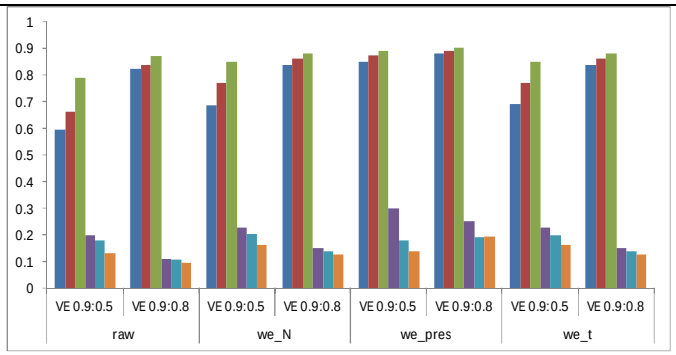
Table 77. Percent Coverage CI for Poisson Regression Approach

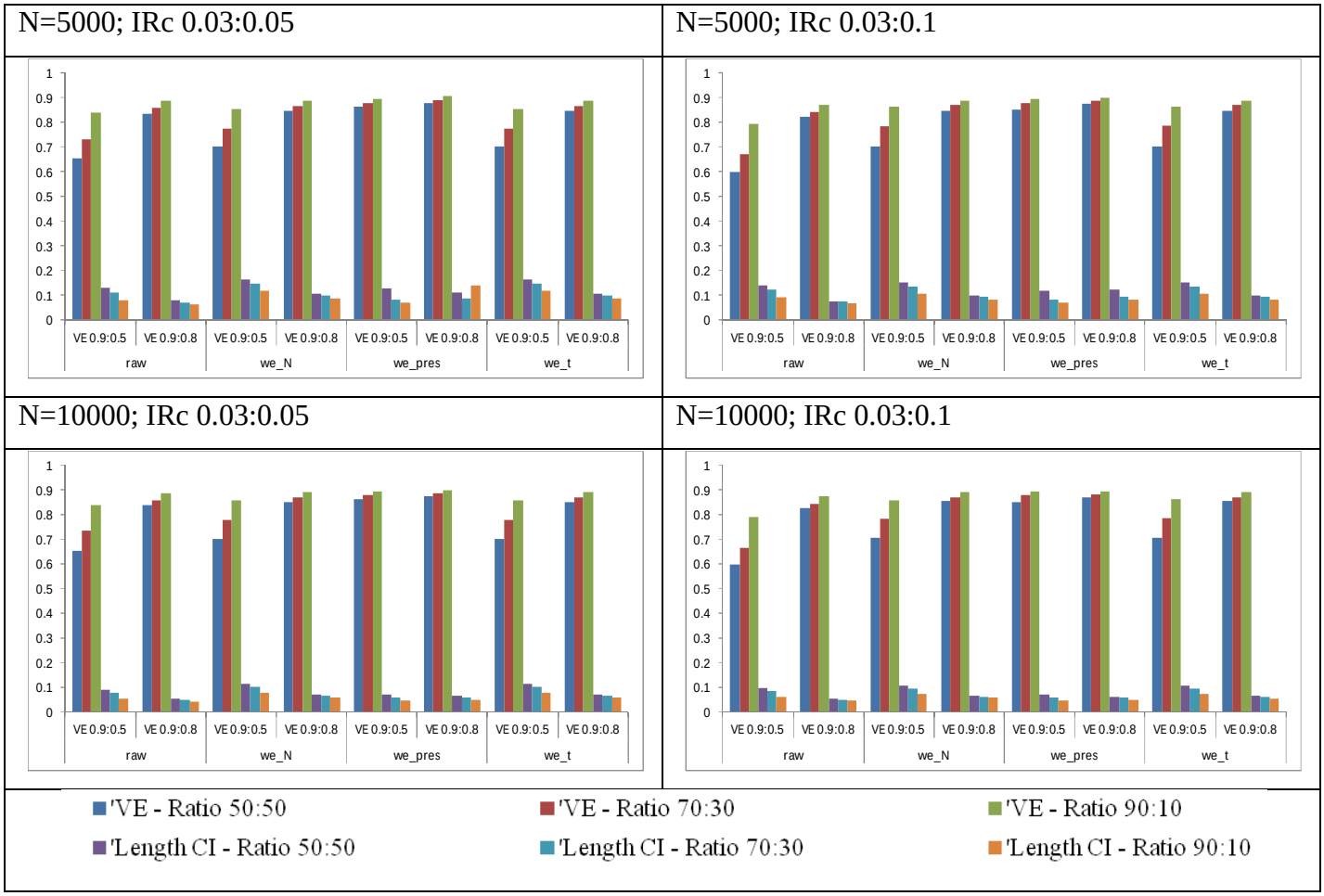
VE Approach	n	Ratio 50:50				Ratio 70:30				Ratio 90:10			
		VE 0.9:0.5		VE 0.9:0.8		VE 0.9:0.5		VE 0.9:0.8		VE 0.9:0.5		VE 0.9:0.8	
		IRc 0.03:0.0	IRc 0.03:0.1	IRc 0.03:0.0	IRc 0.03:0.1	IRc 0.03:0.0	IRc 0.03:0.1	IRc 0.03:0.0	IRc 0.03:0.1	IRc 0.03:0.0	IRc 0.03:0.1	IRc 0.03:0.0	IRc 0.03:0.1
raw	2500	57.800	43.700	59.459	65.200	55.400	29.500	62.651	57.800	55.800	39.800	55.308	58.593
	5000	56.500	21.600	62.100	61.300	48.800	13.900	58.800	54.700	56.400	30.000	56.800	55.200
	10000	48.800	7.000	60.200	53.800	38.300	2.100	59.700	43.600	48.300	12.100	56.500	53.900
we_N	2500	82.450	90.048	81.647	87.327	82.842	89.137	80.973	80.586	80.344	83.576	75.220	76.804
	5000	80.987	90.418	80.224	85.936	82.565	87.174	76.536	79.757	80.441	80.881	71.629	70.909
	10000	83.300	89.089	78.679	82.500	80.981	86.400	76.900	77.300	78.600	82.000	70.341	71.400
we_pres	2500	46.600	47.700	76.076	79.900	42.800	43.700	72.088	71.400	59.500	57.100	73.711	71.357
	5000	16.400	17.800	62.800	69.400	20.700	21.400	62.100	66.100	48.100	45.800	65.300	62.100
	10000	2.300	4.700	60.100	59.600	6.300	9.200	60.600	62.500	36.100	36.300	60.300	63.100
we_t	2500	82.568	90.408	81.765	87.327	83.169	89.031	80.973	80.260	80.559	83.680	75.220	76.471
	5000	80.987	90.622	80.428	85.936	82.565	86.373	76.435	79.453	80.542	80.380	71.521	71.111
	10000	83.600	88.388	78.879	82.500	81.481	85.200	76.900	77.200	79.100	81.800	70.140	71.100

VE approach: raw: Raw of vaccine efficacy (Equation (23)), we_N: general weighted average vaccine efficacy based on sample size for each sub-population (Equation (7)), we_pres : weighted average vaccine efficacy based on precision approach (Equation (9)), we_t: general weighted average vaccine efficacy based on total follow up time for each sub-population (Equation (19)).

The above table shows that the actual coverage of Poisson Regression based approach for all VE approaches seem to have permissive coverage, since most of them have actual coverage less than 95%. However, the coverage of raw VE is smaller compared to weighted average VE. While among weighted average VE approaches, the actual coverage for both general weighted average VE approach, weighted by sample size and follow up time have higher coverage and closer to the nominal 95% compared to weighted average VE based on precision approach. The raw VE and weighted average VE based on precision approach provide the worst coverage and far from 95% that would have been expected especially in the combination of IRc 0.3:0.1 and VE 0.9:0.5 for raw VE. Meanwhile for weighted average VE based on precision give the worst coverage in the combination of VE 0.9:0.5 and Ratio 50:50 and 70:30.

Figure 4. Estimate of VE and Length CI for RR as Poisson Regression

N=2500; IRc 0.03:0.05	N=2500; IRc 0.03:0.1
	



VE approach: raw: Raw of vaccine efficacy (Equation (23)), we_N: general weighted average vaccine efficacy based on sample size for each sub-population (Equation (7)), we_pres : weighted average vaccine efficacy based on precision approach (Equation (9)), we_t: general weighted average vaccine efficacy based on total follow up time for each sub-population (Equation (19)).

From figure 4 above shows that the estimate of VE and length of CI based on the relative risk (RR) as Poisson regression approach. In average, the estimate of VE slightly increases along with the decreasing the distance between the estimate and true value when sample size increase. However, the length decreases with increasing sample size.

The estimate of VE is slightly higher in IRc 0.03:0.05 compared to IRc 0.03:0.1 with the same pattern for all VE approach (raw VE and weighted average VE). Meanwhile, the estimate of VE based on defined VE setting on naïve and non-naïve sub-population for VE 0.9:0.8 is higher than VE 0.9:0.5. The estimate of VE in Ratio 90:10 is the highest compared to those on Ratio 70:30 and Ratio 50:50. The estimate of VE is increasing as the ratio of population in naïve increase. However, the estimate of raw VE is lower than those weighted average VE approaches. Both of general weighted average approach, weighted on sample size and total follow up time, have almost the same estimate VE. Meanwhile, the weighted average VE based on precision approach give the highest VE among all VE approaches.

4 Discussion and Conclusion

Vaccine efficacy has known as the risk reduction of the events in the vaccine group compared to the unvaccinated/control group. The estimation of vaccine efficacy depends on the study design, the structure of the population, the mixing patterns, the routes of transmission of the infectious agent, the way the vaccine confers protection, the characteristics of the vaccination program and the sampling procedure (Haber, et al., 1991). However, this requires information on exposure to infection, which is often difficult or impossible to obtain. The incidence rate, hazard rate, and the cumulative incidence (attack rates) are measures of disease frequency that generally do not require knowledge of contact with infective.

The heterogeneity of the vaccine efficacy of the two sub-populations related to baseline HPV status, naïve and non-naïve cohorts, need to be taken into account when estimating global vaccine efficacy of the vaccine group compared to the unvaccinated/control group. A weighted average vaccine efficacy, weighting on the baseline HPV status, was considered in the analysis. The adoption from one of three general approaches was considered when choosing the weighted factor: (1) the weighting factor may be chosen in order to achieve the optimal precision in the estimate, (2) to reflect the relative importance of various sub-population/strata or (3) to achieve the standardization of the results relative to standard population (Kleinbaum, et al., 1982).

The estimate of vaccine efficacy based on different relative risk (RR) approach give the same information on VE estimate. Choosing RR as the way to estimate vaccine efficacy related to the objective of the study. Primary vaccine efficacy studies often report VE based on relative events per person time as the cumulative of incidence rates, namely attack rate, when assumed that all patients in the study have the same follow up time. In reality, this assumption hardly fulfilled. Thus, the person time approach was considered for estimating vaccine efficacy. When covariates added in the analysis, the analyses modeled by the covariates or Poisson regression could be used. Under the assumption that the effect of the vaccine is multiplicative, constant, and homogeneous, the Cox proportional hazards model was used to estimate VE and requires only the ordering of the event time.

The estimate of raw VE estimated by the consideration on the homogenous population, meaning that do not take into account the heterogeneity of the population. This condition makes the raw VE produced the farthest estimate VE to the true value compared to weighted

average VE. Thus, yield the smallest deviation and produce the actual coverage farther apart from nominal 95% along with increasing sample size. The estimate raw VE was smaller than those on weighted average VE for all RR approaches. Moreover, the condition yields the same pattern for MSE.

Meanwhile, among weighted average VE, the precision-based weighted average vaccine efficacy has a higher estimate VE than others weighted average VE approaches. However, weighted average VE based on precision approach yield the largest difference between the standard deviation for sample estimate VE and the mean standard error VE. It's implied that the weighted average VE based on precision approach was less close to the true value and has higher MSE compared to others weighted average VE, since the precision based approach produce the smallest weighted for estimator when the population is not homogenous. Moreover, the actual coverage is farther apart from nominal 95% compared to others weighted average VE approaches.

The general weighted average VE, weighted on sample size for each sub-population and total follow up time gave almost the same result for all RR approaches. The heterogeneity in the population using different characteristic for each sub-population was taking into account when estimating vaccine efficacy. The number of patients for each sub-population was considered as the weighted factors that differentiate naïve and non-naïve sub-population. This condition yields the estimate of general weighted average VE produce higher VE and closer to the real value compare that those on other weighted average VE approaches. Moreover, the MSE produced by this method is the lowest one and the actual coverage is very close from nominal 95% compared to others weighted average VE approaches.

The weighted average VE based on the Halloran approach for attack rate and person time approach produce almost the same result as the general weighted average VE approaches. The reason is that weighted average VE based on Halloran approach using the number of patients for each sub-population and group as the weighted factor when estimating vaccine efficacy. However, weighted average VE based on Halloran approach gave the smallest distance between estimate VE and true value compared to those on general weighted average VE.

The estimate of weighted vaccine efficacy is closer to the real value as the number of sample increase. However, the length of CI slightly decreases with increasing sample size. Since

larger sample sizes generally lead to increased precision when estimating unknown parameters.

The estimate of weighted average VE is increasing as the ratio of population in the naïve increase along with the increasing of vaccine efficacy in non-naïve sub-population. Moreover, the estimate of weighted average VE is decreasing as the increasing of the incidence rate of the control group for non-naïve sub-population.

The increasing of incidence rate in the control group (IRc) for non-naïve sub-population yields the decreasing of estimate VE. The raw estimate of vaccine efficacy approach, give a lower vaccine efficacy estimate compared to weighted average vaccine efficacy approach in all RR approach. However, the estimate of VE increases as the ratio of population in naïve and non-naïve (R naïve:non-naïve) increase, while for the length CI in most settings seem to decrease as ratio naïve and non-naïve increase.

The discrepancy between actual and nominal coverage probability occur because of the assumptions used in deriving a confidence interval are not met. In this case, approximating a discrete distributional with continuous one. The approximation of the binomial distribution with the normal distribution yields the farthest distance between actual and nominal coverage probability. Another reason is there were excessive zero counts (no event) in the study that may cause the deviation from the Poisson distribution, this condition yields the low actual coverage for Poisson Regression approach in all VE methods.

The excessive of zero counts also yields the low coverage in the raw VE for all RR measurement approaches when there were large different between incidence rate in control group for naïve and non-naïve sub-population for low vaccine efficacy in non-naïve sub-population compared to naive. While for weighted average precision-based approach, low actual coverage appear when there were large discrepancy between vaccine efficacy in naïve and non-naïve sub-population.

The interpretation of estimate vaccine efficacy depends on which measurement risk is operating. Although the estimates of vaccine efficacy are desirable, these are not always estimable. Thus, the understanding summary vaccine efficacy estimates in light of the underlying biology are important.

The weighted average VE approach, the approach that takes into account the heterogeneity of the population, can provide the best estimate of VE and coverage probability compared to the

approach that do not take into account the heterogeneity in the estimation. Among weighted average VE approaches, both general weighted average VE approaches, general weighted average VE based on sample size for each sub-population and based on total follow up time give a better estimate of VE. However, the method that handles few events even zero events in the study should be considered in the future in order to get better estimation of weighted average vaccine efficacy.

5 References

1. Dragalin, V., Fedorov, V. & Cheuvart, B., 2002. Statistical Approaches to establishing Vaccine Efficacy. *Statistics in Medicine,* Volume 21, pp. 877-893.
2. FDA, 2006. *US Food and Drug Administration.* [Online] Available at: http://www.fda.gov/downloads/BiologicsBloodVaccines/Vaccines/ApprovedProducts/ucm111285.pdf
   [Accessed 26 July 2012].
3. Haber, M., Longini, J. I. M. & Halloran, M. E., 1991. Estimation of vaccine efficacy in outbreaks of acute infectious diseases. *Statistics in Medicine Vol 10,* pp. 1573-1584.
4. Halloran, E. M., Haber, M. & Longini, J. I. M., 1992. Intepretation and estimation of vaccine efficacy under heterogeneity. *American Journal of Epidemiology Vol. 136 No. 3,* pp. 328-343.
5. Halloran, E. M., Longini, J. I. M. & Struchiner, C. J., 2010. *Design and Analysis of Vaccine Studies.* New York: Springer.
6. HPV Patricia Study Group, 2007. Efficacy of prophylactic adjuvanted bivalent L1 virus-like-particle vaccine against infection with human papilomavirus type 16 and 18 in young women; an interim analysis of a phase III double blind, randomised controlled trial. *Lancet,* Volume 369, pp. 2161-2170.
7. HPV Patricia Study Group, 2009. Efficacy of human papilomavirus (HPV)-16/18 AS04-adjuvanted vaccine against cervical infection and precancer caused by oncogenic HPV types (PATRICIA) : Final analysis of double blind, randomized study in young women. *Lancet,* pp. 1-14.
8. IARC, 2007. *Monographs on the evaluation of carcinogenic risks to humans. Volume 90: human papillomaviruses.* Lyon, France: World Health Organization, International Agency for Research on Cancer (IARC).
9. Kleinbaum, D. G., Kupper, L. L. & Morgenstern, H., 1982. *Epidemiologic Research-Principles and Quantitative Methods.* New York: Lifetime Learning.
10. Monto, A. S., Hornbuckle, K. & Ohmit, S. E., 2001. Influenza Vaccine Effectiveness among elderly nursing home resident: A Cohort Study. *American Journal of Epidemiology,* 154(2), pp. 155-160.

11. Muñoz, N. et al., 2003. Epidemiologic classification of human papillomavirus types associated with cervical cancer. *N Engl J Med*, Volume 348, p. 518–527.
12. National Institute of Allergy and Infectious Disease, 2010. *Human Papillomavirus (HPV) and Genital Warts.* [Online] Available at: http://www.niaid.nih.gov/topics/genitalwarts/pages/default.aspx [Accessed 3 August 2012].
13. Schiffman, M. & Castle, P. E., 2003. Human Papillomavirus: Epidemiology and Public Health. *Archives of Pathology & Laboratory Medicine*, 127(8), pp. 930-934.
14. Small, D. S., Cheng, J. & Ten Have, T. R., 2010. Evaluating the Efficacy of a Malaria Vaccine. *The International Journal of Biostatistics*, pp. Vol. 6 , Iss. 2, Art. 4.

6 Appendix

A. Description of Equations

A1. Natural Logarithmic Transformation on Wald Confidence Interval

The estimate of $(1-\alpha)100\%$ confidence interval for VE is simply derived from Wald large-sample confidence interval as $\widehat{VE} \pm z_{1-\alpha/2}\sqrt{\widehat{Var}\left(\widehat{VE}\right)}$, where $z_{1-\alpha/2}$ is the value of standard normal distribution. The problem occurs when using this particular interval for VE is that there is asymmetric in the distribution of possible values for VE, which is lead to distortion in nominal confidence coefficient and cause uncover the true parameter value.

The natural logarithmic transformation is often used to transform highly skewed distribution into one that approximately normal in appearance (Kleinbaum, et al., 1982). In this study, the natural logarithmic transformation was applied on the logit scale of parameter estimate VE, namely:

$$\hat{\eta} = \text{logit}\left(\widehat{VE}\right) = \log\left(\frac{\widehat{VE}}{1-\widehat{VE}}\right). \tag{24}$$

$(1-\alpha)100\%$ Wald large-sample confidence interval of $\hat{\eta}$ is obtained as follows:

$$\hat{\eta} \pm z_{1-\alpha/2}\sqrt{\widehat{Var}\left(\hat{\eta}\right)}. \tag{25}$$

The variance of $\hat{\eta}$ can be derived using Delta methods as follows:

$$\widehat{Var}\left(\hat{\eta}\right) \approx \left(\frac{1}{\widehat{VE}\left(1-\widehat{VE}\right)}\right)^2 \widehat{Var}\left(\widehat{VE}\right). \tag{26}$$

The desired upper and lower endpoints of the $(1-\alpha)100\%$ Wald large-sample confidence interval for $\widehat{VE}$ are found by taking the back-transformed parameter of logit scale, namely expit scale: $\widehat{VE}_{LL} = \frac{e^{\hat{\eta}_{LL}}}{1+e^{\hat{\eta}_{LL}}}$ and $\widehat{VE}_{UL} = \frac{e^{\hat{\eta}_{UL}}}{1+e^{\hat{\eta}_{UL}}}$.

A2. The Variance Derivation for Weighted Average Vaccine Efficacy Halloran et. al for Attack Rate

The weighted average of vaccine efficacy based on Halloran et. al described as following expression:

$$\widehat{VE}\bullet = 1 - \frac{\sum_{k=1}^{K} \alpha_{vk} \ln\left(1 - AR_{vk}\right)}{\sum_{k=1}^{K} \gamma_{ck} \ln\left(1 - AR_{ck}\right)},$$

where k=1, 2, …,K. In this study, K=2 which corresponding to naïve and non-naïve sub populations. The weight defines as the proportion of patient enrolled in k^{th} sub-population for each group vaccine and control, i.e:

$\alpha_{vk} = N_{vk} \big/ N_v$ and $\gamma_{ck} = N_{ck} \big/ N_c$, with the restrictions, $\sum_{k=1}^{K} \alpha_{vk} = 1$ and $\sum_{k=1}^{K} \gamma_{ck} = 1$.

And simply $\widehat{VE}\bullet = 1 - \widehat{RR}\bullet$, where $\widehat{RR}\bullet = \dfrac{\sum_{k=1}^{K} \alpha_{vk} \ln\left(1 - AR_{vk}\right)}{\sum_{k=1}^{K} \gamma_{ck} \ln\left(1 - AR_{ck}\right)}$.

The variance of weighted average of vaccine efficacy can be derived analytically from the expression of variance of attack rate for k^{th} sub-population in each group that follow well-known properties of the binomial distribution as follow:

$\widehat{Var}\left(AR_{vk}\right) = \dfrac{AR_{vk}\left(1 - AR_{vk}\right)}{N_{vk}}$ and $\widehat{Var}\left(AR_{ck}\right) = \dfrac{AR_{ck}\left(1 - AR_{ck}\right)}{N_{ck}}$ for vaccine and control group respectively at k^{th} sub-population. Thus, the variance of $\ln\left(1 - AR_{vk}\right)$ in the vaccine group can be derived using Delta Methods as follows:

$$\widehat{Var}\left(\ln\left(1 - AR_{vk}\right)\right) = \left(\frac{1}{\left(1 - AR_{vk}\right)}\right)^2 Var\left(AR_{vk}\right). \tag{27}$$

The same expression applies to control group. Assuming that there is independency between naïve and non-naïve sub-population, Furthermore, the variance of numerator expression of $\widehat{RR}\bullet$ is: $\widehat{Var}\left(\sum_{k=1}^{K} \alpha_{vk} \ln\left(1 - AR_{vk}\right)\right) = \sum_{k=1}^{K} \alpha_{vk}^2 \left(\frac{1}{\left(1 - AR_{vk}\right)}\right)^2 Var\left(AR_{vk}\right).$

While the variance of denominator of $\widehat{RR}\bullet$ defines as:

$$\widehat{Var}\left(\sum_{k=1}^{K} \alpha_{ck} \ln\left(1 - AR_{ck}\right)\right) = \sum_{k=1}^{K} \alpha_{ck}^2 \left(\frac{1}{\left(1 - AR_{ck}\right)}\right)^2 Var\left(AR_{ck}\right).$$

To define the variance of $\widehat{RR}\bullet$ can simply derive from natural logarithmic transformation between the ratio of the numerator and denominator and assume that there is independency between vaccine and control group yield the result:

$$\widehat{Var}\left(\widehat{RR}\bullet\right)=\left(\widehat{RR}\bullet\right)^2\left[\frac{Var\left(\sum_{k=1}^{K}\alpha_{vk}\ln\left(1-AR_{vk}\right)\right)}{\left(\sum_{k=1}^{K}\alpha_{vk}\ln\left(1-AR_{vk}\right)\right)^2}+\frac{Var\left(\sum_{k=1}^{K}\alpha_{ck}\ln\left(1-AR_{ck}\right)\right)}{\left(\sum_{k=1}^{K}\alpha_{ck}\ln\left(1-AR_{ck}\right)\right)^2}\right],$$

$$\widehat{Var}\left(\widehat{RR}\bullet\right)=\left(\frac{\sum_{k=1}^{K}\alpha_{vk}\ln\left(1-AR_{vk}\right)}{\sum_{k=1}^{K}\gamma_{ck}\ln\left(1-AR_{ck}\right)}\right)^2\times$$
$$\left[\frac{\sum_{k=1}^{K}\alpha_{vk}^2\left(\frac{1}{\left(1-AR_{vk}\right)}\right)^2\left(\frac{AR_{vk}\left(1-AR_{vk}\right)}{N_{vk}}\right)}{\left(\sum_{k=1}^{K}\alpha_{vk}\ln\left(1-AR_{vk}\right)\right)^2}+\frac{\sum_{k=1}^{K}\alpha_{ck}^2\left(\frac{1}{\left(1-AR_{ck}\right)}\right)^2\left(\frac{AR_{ck}\left(1-AR_{ck}\right)}{N_{ck}}\right)}{\left(\sum_{k=1}^{K}\alpha_{ck}\ln\left(1-AR_{ck}\right)\right)^2}\right]. \quad (28)$$

The variance of weighted vaccine efficacy can be estimated using Delta Method as follow:

$$\widehat{Var}\left(\widehat{VE}\bullet\right)=Var\left(1-\widehat{RR}\bullet\right),$$
$$=\widehat{Var}\left(\widehat{RR}\bullet\right).$$

A3. The Variance Derivation for Weighted Average Vaccine Efficacy based on precision for Person Time Approach

The estimates of weighted average vaccine efficacy define as: $\widehat{VE}=1-\widehat{RR}.$ Where $\widehat{RR}=\frac{X_v/T_v}{X_c/T_c}$, which is equivalent to $\widehat{RR}=\frac{p}{r\left(1-p\right)}$, and p denote as the proportion of event in the vaccinated group, $p=X_v/\left(X_v+X_c\right)$. Thus, the variance of vaccine efficacy can be derived based on the log of the ratio of two binomial random variables as follow:

To simplify the estimation if variance, transformation of natural logarithmic was needed, yield: $\ln\left(\widehat{RR}\right)=\ln\left(\frac{X_v/T_v}{X_c/T_c}\right)$. The variance of estimate $\ln\left(\widehat{RR}\right)$ as described in (Halloran, et al., 2010) as follows:

$$\widehat{Var}\left(\ln\left(\widehat{RR}\right)\right)=\frac{T_v-X_v}{T_vX_v}+\frac{T_c-X_c}{T_cX_c},$$
$$=\frac{1}{X_v}+\frac{1}{T_v}+\frac{1}{X_c}+\frac{1}{T_c}. \quad (29)$$

The variance of $\widehat{RR}$ simply derived using Delta Methods as follows:

$$\begin{aligned}\widehat{Var}\left(\widehat{RR}\right) &= \widehat{RR}^2\,\widehat{Var}\left(\ln\left(\widehat{RR}\right)\right),\\ &= \left(\frac{X_v/T_v}{X_c/T_c}\right)^2\left(\frac{1}{X_v}+\frac{1}{T_v}+\frac{1}{X_c}+\frac{1}{T_c}\right).\end{aligned} \tag{30}$$

This equivalent to equation: $\widehat{Var}\left(\widehat{RR}\right)=\left(\frac{p}{r(1-p)}\right)^2\left(\frac{1}{X_v}+\frac{1}{T_v}+\frac{1}{X_c}+\frac{1}{T_c}\right).$

Thus, the variance of VE as follows:

$$\begin{aligned}\widehat{Var}\left(\widehat{VE}\right) &= Var\left(1-\widehat{RR}\right),\\ &= \widehat{Var}\left(\widehat{RR}\right).\end{aligned}$$

A4. The Variance Derivation for Weighted Average Vaccine Efficacy Halloran et. al for Person Time Approach

The weighted average of vaccine efficacy for person time approach using Halloran et al is defined as follows:

$$\widehat{VE}_{\bullet} = 1-\frac{\sum_{k=1}^{K}\alpha_{vk}\left(X_{vk}/T_{vk}\right)}{\sum_{k=1}^{K}\gamma_{ck}\left(X_{ck}/T_{ck}\right)},$$

where T_{vk} and T_{ck} denote as total follow up time in years for k[th] sub-population at vaccinated and control group respectively. While X_{ck} and X_{vk} denote the observed numbers of event at the end of the study for k[th] sub-population among control and vaccinated group. The weight is defined as the proportion of patients enrolled in k[th] sub-population for each group vaccine and control, i.e:

$\alpha_{vk} = N_{vk}/N_v$ and $\gamma_{ck} = N_{ck}/N_c$, with the restrictions, $\sum_{k=1}^{K}\alpha_{vk}=1$ and $\sum_{k=1}^{K}\gamma_{ck}=1$.

The weighted average vaccine efficacy defined as: $\widehat{VE}_{\bullet}=1-\widehat{RR}_{\bullet}$, where $\widehat{RR}$ is the relative risk for vaccine group to control group, $\widehat{RR}_{\bullet}=\frac{\sum_{k=1}^{K}\alpha_{vk}\left(X_{vk}/T_{vk}\right)}{\sum_{k=1}^{K}\gamma_{ck}\left(X_{ck}/T_{ck}\right)}$.

The variance of risk for k[th] sub-population between control and vaccinated group are defined as follow:

$$Var(X_{vk}/T_{vk}) = \left(X_{vk}/T_{vk}\right)^2 \times \left(1/X_{vk} + 1/T_{vk}\right) \text{ and,}$$
$$Var(X_{ck}/T_{ck}) = \left(X_{ck}/T_{ck}\right)^2 \times \left(1/X_{ck} + 1/T_{ck}\right).$$

The variance of the numerator and denominator of $\widehat{RR}_{\bullet}$ are defined as:

$$Var\left(\sum_{k=1}^{K}\alpha_{vk}\left(X_{vk}/T_{vk}\right)\right) = \sum_{k=1}^{K}\alpha_{vk}{}^2 Var\left(X_{vk}/T_{vk}\right),$$
$$Var\left(\sum_{k=1}^{K}\gamma_{ck}\left(X_{ck}/T_{ck}\right)\right) = \sum_{k=1}^{K}\gamma_{ck}{}^2 Var\left(X_{ck}/T_{ck}\right).$$

Thus the variance of $\widehat{RR}_{\bullet}$ is defined using natural logarithmic transformation between the ratio of the numerator and denominator and assume that there is independency between vaccine and control group yield the result:

$$\widehat{Var}\left(\widehat{RR}_{\bullet}\right) = \left(\widehat{RR}_{\bullet}\right)^2 \left[\frac{Var\left(\sum_{k=1}^{K}\alpha_{vk}\left(X_{vk}/T_{vk}\right)\right)}{\left(\sum_{k=1}^{K}\alpha_{vk}\left(X_{vk}/T_{vk}\right)\right)^2} + \frac{Var\left(\sum_{k=1}^{K}\gamma_{ck}\left(X_{ck}/T_{ck}\right)\right)}{\left(\sum_{k=1}^{K}\gamma_{ck}\left(X_{ck}/T_{ck}\right)\right)^2}\right], \textit{yield}:$$

$$\widehat{Var}\left(\widehat{RR}_{\bullet}\right) = \left(\frac{\sum_{k=1}^{K}\alpha_{vk}\left(X_{vk}/T_{vk}\right)}{\sum_{k=1}^{K}\gamma_{ck}\left(X_{ck}/T_{ck}\right)}\right)^2 \times \left[\frac{\sum_{k=1}^{K}\alpha_{vk}{}^2\left(\left(X_{vk}/T_{vk}\right)^2 \times \left(1/X_{vk} + 1/T_{vk}\right)\right)}{\left(\sum_{k=1}^{K}\alpha_{vk}\left(X_{vk}/T_{vk}\right)\right)^2} + \frac{\sum_{k=1}^{K}\gamma_{ck}{}^2\left(\left(X_{ck}/T_{ck}\right)^2 \times \left(1/X_{ck} + 1/T_{ck}\right)\right)}{\left(\sum_{k=1}^{K}\gamma_{ck}\left(X_{ck}/T_{ck}\right)\right)^2}\right] \quad (31)$$

The variance of weighted vaccine efficacy can be estimated using Delta Method as follows:

$$\widehat{Var}\left(\widehat{VE}_{\bullet}\right) = Var\left(1 - \widehat{RR}_{\bullet}\right),$$
$$= \widehat{Var}\left(\widehat{RR}_{\bullet}\right).$$

A5. The True value derivation methods

The derivation of true value vaccine efficacy is based on different approaches of VE and simulation setting. The simulation setting is presented with different criteria: (1) the proportion of the number of patient in naïve and non-naïve sub-population (Ratio naïve: non-naïve), (2) the incidence rate in the control group for naïve and non-naïve (IRc naïve: non-naïve), (3) vaccine efficacy (VE) for naïve and non-naïve (VE naïve: non-naïve) and (4) the sample mean (n).

The true value of vaccine efficacy was calculated based on the different VE approach. The true value of raw VE, general weighted of VE based on sample size and follow up time as well as weighted average VE based on precision approach defined as equation (7):

$$VE_{\bullet} = \left(Ratio_{naive} \times VE_{naive}\right) + \left(Ratio_{non-naive} \times VE_{non-naive}\right). \quad (32)$$

The true value the true value of weighted average VE based on Halloran et all defined as equation (5) for the attack rate (AR) ratio approach as follows:

$$VE_{\bullet} = 1 - \frac{\left(Ratio_{naive} \times \ln(1-((\mathbf{1}-VE_{naive}) \times IRc_{naive}))\right) + \left(Ratio_{non-naive} \times \ln(1-((\mathbf{1}-VE_{non-naive}) \times IRc_{non-naive}))\right)}{\left(Ratio_{naive} \times \ln(1-IRc_{naive})\right) + \left(Ratio_{non-naive} \times \ln(1-IRc_{non-naive})\right)}. \quad (33)$$

Meanwhile, the true value the true value of weighted average VE based on Halloran et all defined as equation (17) for person time approach is:

$$VE_{\bullet} = 1 - \frac{\left(Ratio_{naive} \times ((\mathbf{1}-VE_{naive}) \times IRc_{naive})\right) + \left(Ratio_{non-naive} \times ((\mathbf{1}-VE_{non-naive}) \times IRc_{non-naive})\right)}{\left(Ratio_{naive} \times IRc_{naive}\right) + \left(Ratio_{non-naive} \times IRc_{non-naive}\right)}. \quad (34)$$

A6. The Variance Derivation for Weighted Average Vaccine Efficacy Precision based Approach for Attack Rate

The estimate of weighted average vaccine efficacy based on precision approach is defined as follows: $\widehat{VE}_{\bullet} = \sum_{k=1}^{K} \hat{w}_k \widehat{VE}_k \Big/ \sum_{k-1}^{K} \hat{w}_k$, with $\hat{w}_k = 1/\widehat{Var}\left(\widehat{VE}_k\right)$.

Where $\widehat{VE}_k$ is the estimate of vaccine efficacy for k^{th} sub-population that can be derived using the equation (3) and $\hat{w}_k$ is a weight applied to k^{th} sub-population. From equation (4), the estimate of weighted factor $\hat{w}_k$ can be described as:

$$\hat{w}_k \simeq \left[\ln\left(\frac{N_{ck}}{N_{ck}-X_{ck}}\right)\right]^2 \Big/ \left[\frac{X_{vk}}{N_{vk}\left(N_{vk}-X_{vk}\right)} + \left(1-\widehat{VE}_k\right)^2 \frac{X_{ck}}{N_{ck}\left(N_{ck}-X_{ck}\right)}\right]. \quad (35)$$

The numerator part of $\hat{w}_k$ can be described as:

$$\begin{aligned}
\left[\ln\left(\frac{N_{ck}}{N_{ck}-X_{ck}}\right)\right]^2 &= \left[\ln\left(\frac{N_{ck}}{N_{ck}-AR_{ck}N_{ck}}\right)\right]^2, \\
&= \left[\ln\left(\frac{N_{ck}}{N_{ck}\left(1-AR_{ck}\right)}\right)\right]^2, \\
&= \left[-\ln\left(1-AR_{ck}\right)\right]^2, \\
&= \left[\ln\left(1-AR_{ck}\right)\right]^2.
\end{aligned}$$

The variance of the numerator part from Equation (35) described as follow with information of variance $\ln\left(1-AR_{ck}\right)$ from equation(27) in Appendix 2 :

$$\widehat{Var}\left(\ln\left(1-AR_{ck}\right)\right)=\left(\frac{1}{\left(1-AR_{ck}\right)}\right)^2 Var\left(AR_{ck}\right), \text{ and } \widehat{Var}\left(AR_{ck}\right)=\frac{AR_{ck}\left(1-AR_{ck}\right)}{N_{ck}}.$$

Using transformation of natural logarithm of numerator part of equation (35), the estimate variance of numerator part $\hat{w}_k$ yields as follows:

$$\begin{aligned}\widehat{Var}\left[\left(\ln\left(1-AR_{ck}\right)\right)^2\right] &= 4\left(\ln\left(1-AR_{ck}\right)\right)^2 \widehat{Var}\left(\ln\left(1-AR_{ck}\right)\right),\\ &= 4\left(\ln\left(1-AR_{ck}\right)\right)^2\left(\frac{1}{\left(1-AR_{ck}\right)}\right)^2 Var\left(AR_{ck}\right),\\ &= 4\left(\ln\left(1-AR_{ck}\right)\right)^2\left(\frac{1}{\left(1-AR_{ck}\right)}\right)^2 \frac{AR_{ck}\left(1-AR_{ck}\right)}{N_{ck}},\\ &= 4\left(\ln\left(1-AR_{ck}\right)\right)^2 \frac{AR_{ck}}{N_{ck}\left(1-AR_{ck}\right)}.\end{aligned} \tag{36}$$

The denominator part of $\hat{w}_k$ in equation (35) can be described as follows:

$$\begin{aligned}\left[\frac{X_{vk}}{N_{vk}\left(N_{vk}-X_{vk}\right)}+\left(1-\widehat{VE}_k\right)^2\frac{X_{ck}}{N_{ck}\left(N_{ck}-X_{ck}\right)}\right] &= \frac{AR_{vk}}{\left(N_{vk}-AR_{vk}N_{vk}\right)}+\left(1-\widehat{VE}_k\right)^2\frac{AR_{ck}}{\left(N_{ck}-AR_{ck}N_{ck}\right)},\\ &= \frac{AR_{vk}}{\left(N_{vk}\left(1-AR_{vk}\right)\right)}+\left(1-\widehat{VE}_k\right)^2\frac{AR_{ck}}{\left(N_{ck}\left(1-AR_{ck}\right)\right)}.\end{aligned}$$

The variance of $\frac{AR_{vk}}{\left(N_{vk}\left(1-AR_{vk}\right)\right)}$ defined using natural logarithmic transformation as follow:

$$\begin{aligned}\widehat{Var}\left(\ln\frac{AR_{vk}}{\left(N_{vk}\left(1-AR_{vk}\right)\right)}\right) &= \left(\frac{d\left(\ln\frac{AR_{vk}}{\left(N_{vk}\left(1-AR_{vk}\right)\right)}\right)}{dAR_{vk}}\right)^2 Var\left(AR_{vk}\right),\\ &= \left(\frac{d\left(\ln AR_{vk}-\ln\left(N_{vk}\left(1-AR_{vk}\right)\right)\right)}{dAR_{vk}}\right)^2 Var\left(AR_{vk}\right),\\ &= \left(\frac{1}{AR_{vk}}+\frac{1}{\left(1-AR_{vk}\right)}\right)^2\frac{AR_{vk}\left(1-AR_{vk}\right)}{N_{vk}},\\ &= \left(\frac{1}{AR_{vk}\left(1-AR_{vk}\right)}\right)^2\frac{AR_{vk}\left(1-AR_{vk}\right)}{N_{vk}},\\ &= \frac{1}{X_{vk}\left(1-AR_{vk}\right)}.\end{aligned}$$

Thus using Delta Methods yields:

$$\widehat{Var}\left(\frac{AR_{vk}}{\left(N_{vk}\left(1-AR_{vk}\right)\right)}\right)=\left(\frac{AR_{vk}}{\left(N_{vk}\left(1-AR_{vk}\right)\right)}\right)^{2}\frac{1}{X_{vk}\left(1-AR_{vk}\right)}.$$

The variance of $\left(1-\widehat{VE}_k\right)^2\frac{AR_{ck}}{\left(N_{ck}\left(1-AR_{ck}\right)\right)}$ defined as follows using logarithmic transformation approach:

$$\ln\left(\left(1-\widehat{VE}_k\right)^2\frac{AR_{ck}}{\left(N_{ck}\left(1-AR_{ck}\right)\right)}\right)=2\ln\left(1-\widehat{VE}_k\right)+\ln\left(\frac{AR_{ck}}{\left(N_{ck}\left(1-AR_{ck}\right)\right)}\right). \tag{37}$$

The variance of $\ln\left(\frac{AR_{ck}}{\left(N_{ck}\left(1-AR_{ck}\right)\right)}\right)$ defined as the same approach as before, yields:

$$\widehat{Var}\left(\ln\frac{AR_{ck}}{\left(N_{ck}\left(1-AR_{ck}\right)\right)}\right)=\left(\frac{1}{AR_{ck}\left(1-AR_{ck}\right)}\right)^{2}Var\left(AR_{ck}\right),$$

$$=\frac{1}{X_{ck}\left(1-AR_{ck}\right)}.$$

The variance of $2\ln\left(1-\widehat{VE}_k\right)$ yield as follows:

$$\widehat{Var}\left(2\ln\left(1-\widehat{VE}_k\right)\right)=4Var\left(\ln\left(1-\widehat{VE}_k\right)\right),$$

$$=4\left(\frac{1}{\left(1-\widehat{VE}_k\right)}\right)^{2}\widehat{Var}\left(\widehat{VE}_k\right).$$

The estimate of variance $\widehat{VE}_k$ defined in the equation (4) and using the property of the Cauchy–Schwarz inequality for estimating the covariance, the variance of equation (37) yields as follow:

$$\widehat{Var}\ln\left(\left(1-\widehat{VE}_k\right)^2\frac{AR_{ck}}{\left(N_{ck}\left(1-AR_{ck}\right)\right)}\right)\approx Var\left[2\ln\left(1-\widehat{VE}_k\right)+\ln\left(\frac{AR_{ck}}{\left(N_{ck}\left(1-AR_{ck}\right)\right)}\right)\right],$$

$$\approx\widehat{Var}\left(2\ln\left(1-\widehat{VE}_k\right)\right)+\widehat{Var}\left(\ln\left(\frac{AR_{ck}}{\left(N_{ck}\left(1-AR_{ck}\right)\right)}\right)\right)+2Cov\left[2\ln\left(1-\widehat{VE}_k\right),\ln\left(\frac{AR_{ck}}{\left(N_{ck}\left(1-AR_{ck}\right)\right)}\right)\right], \tag{38}$$

$$\approx 4\left(\frac{1}{\left(1-\widehat{VE}_k\right)}\right)^{2}\widehat{Var}\left(\widehat{VE}_k\right)+\frac{1}{X_{ck}\left(1-AR_{ck}\right)}+2\sqrt{4\left(\frac{1}{\left(1-\widehat{VE}_k\right)}\right)^{2}\widehat{Var}\left(\widehat{VE}_k\right)\times\frac{1}{X_{ck}\left(1-AR_{ck}\right)}}.$$

Thus, the estimate of variance $\left(1-\widehat{VE}_k\right)^2 \frac{AR_{ck}}{\left(N_{ck}\left(1-AR_{ck}\right)\right)}$:

$$\widehat{Var}\left[\left(1-\widehat{VE}_k\right)^2 \frac{AR_{ck}}{\left(N_{ck}\left(1-AR_{ck}\right)\right)}\right]=\left[\left(1-\widehat{VE}_k\right)^2 \frac{AR_{ck}}{\left(N_{ck}\left(1-AR_{ck}\right)\right)}\right]^2 Var\ln\left[\left(1-\widehat{VE}_k\right)^2 \frac{AR_{ck}}{\left(N_{ck}\left(1-AR_{ck}\right)\right)}\right]$$

The variance of denominator part of $\hat{w}_k$ can be described as following:

The denominator part of $\hat{w}_k$: $\frac{AR_{vk}}{\left(N_{vk}\left(1-AR_{vk}\right)\right)}+\left(1-\widehat{VE}_k\right)^2 \frac{AR_{ck}}{\left(N_{ck}\left(1-AR_{ck}\right)\right)}$.

$$Var\left(\frac{AR_{vk}}{\left(N_{vk}\left(1-AR_{vk}\right)\right)}+\left(1-\widehat{VE}_k\right)^2 \frac{AR_{ck}}{\left(N_{ck}\left(1-AR_{ck}\right)\right)}\right)=Var\left(\frac{AR_{vk}}{\left(N_{vk}\left(1-AR_{vk}\right)\right)}\right)+$$
$$Var\left(\left(1-\widehat{VE}_k\right)^2 \frac{AR_{ck}}{\left(N_{ck}\left(1-AR_{ck}\right)\right)}\right)+$$
$$2Cov\left(\frac{AR_{vk}}{\left(N_{vk}\left(1-AR_{vk}\right)\right)},\left(1-\widehat{VE}_k\right)^2 \frac{AR_{ck}}{\left(N_{ck}\left(1-AR_{ck}\right)\right)}\right).$$

The $Cov\left(\frac{AR_{vk}}{\left(N_{vk}\left(1-AR_{vk}\right)\right)},\left(1-\widehat{VE}_k\right)^2 \frac{AR_{ck}}{\left(N_{ck}\left(1-AR_{ck}\right)\right)}\right)$ can be approximated using Cauchy–Schwarz inequality as follows:

$$Cov\left(\frac{AR_{vk}}{\left(N_{vk}\left(1-AR_{vk}\right)\right)},\left(1-\widehat{VE}_k\right)^2 \frac{AR_{ck}}{\left(N_{ck}\left(1-AR_{ck}\right)\right)}\right)\le\sqrt{Var\left(\frac{AR_{vk}}{\left(N_{vk}\left(1-AR_{vk}\right)\right)}\right)\times Var\left(\left(1-\widehat{VE}_k\right)^2 \frac{AR_{ck}}{\left(N_{ck}\left(1-AR_{ck}\right)\right)}\right)}$$

Thus the variance of denominator of $\hat{w}_k$ yield as follow:

$$\widehat{Var}\left(\frac{AR_{vk}}{\left(N_{vk}\left(1-AR_{vk}\right)\right)}+\left(1-\widehat{VE}_k\right)^2 \frac{AR_{ck}}{\left(N_{ck}\left(1-AR_{ck}\right)\right)}\right)\approx\left(\frac{AR_{vk}}{\left(N_{vk}\left(1-AR_{vk}\right)\right)}\right)^2 \frac{1}{X_{vk}\left(1-AR_{vk}\right)}+$$
$$\left[\left(1-\widehat{VE}_k\right)^2 \frac{AR_{ck}}{\left(N_{ck}\left(1-AR_{ck}\right)\right)}\right]^2 Var\ln\left[\left(1-\widehat{VE}_k\right)^2 \frac{AR_{ck}}{\left(N_{ck}\left(1-AR_{ck}\right)\right)}\right]+ \quad (39)$$
$$2\sqrt{\left(\frac{AR_{vk}}{\left(N_{vk}\left(1-AR_{vk}\right)\right)}\right)^2 \frac{1}{X_{vk}\left(1-AR_{vk}\right)}\times\left(\left[\left(1-\widehat{VE}_k\right)^2 \frac{AR_{ck}}{\left(N_{ck}\left(1-AR_{ck}\right)\right)}\right]^2 Var\ln\left[\left(1-\widehat{VE}_k\right)^2 \frac{AR_{ck}}{\left(N_{ck}\left(1-AR_{ck}\right)\right)}\right]\right)}$$

The variance of weighted factor $\hat{w}_k$ describes using transformation of natural logarithm as follow:

Suppose that the numerator of $\hat{w}_k$ (equation (35)), $\left(\ln\left(1-AR_{ck}\right)\right)^2 = a$ and denominator of $\hat{w}_k$, $\left(\frac{AR_{vk}}{\left(N_{vk}\left(1-AR_{vk}\right)\right)}+\left(1-\widehat{VE}_k\right)^2\frac{AR_{ck}}{\left(N_{ck}\left(1-AR_{ck}\right)\right)}\right)=b$. Thus, $\ln\hat{w}_k = \ln a - \ln b$. The variance of a and b describes from equation (36) and (39) respectively, yields:

$$\begin{aligned}\widehat{Var}\left(\ln\hat{w}_k\right) &= Var\left(\ln a\right)+Var\left(\ln b\right)-2Cov\left(\ln a,\ln b\right),\\ &\approx\left(\frac{1}{a}\right)^2 Var\left(a\right)+\left(\frac{1}{b}\right)^2 Var\left(b\right)-2\sqrt{\left(\frac{1}{ab}\right)^2 Var\left(a\right)Var\left(b\right)}.\end{aligned} \tag{40}$$

Thus yields:

$$\widehat{Var}\left(\hat{w}_k\right) = \hat{w}_k^2\widehat{Var}\left(\ln\hat{w}_k\right). \tag{41}$$

The weighted average vaccine efficacy based on precision approach defines as equation (9): $\widehat{VE}_{\bullet} = \sum_{k=1}^{K}\hat{w}_k\widehat{VE}_k \Big/ \sum_{k-1}^{K}\hat{w}_k$. The variance of $\hat{w}_k\widehat{VE}_k$ for each k[th] sub-population is defines using natural logarithmic transformation as follows:

$$\begin{aligned}\widehat{Var}\left(\ln\left(\hat{w}_k\widehat{VE}_k\right)\right) &= Var\left(\ln\hat{w}_k\right)+Var\left(\ln\widehat{VE}_k\right)+2Cov\left(\hat{w}_k,\widehat{VE}_k\right),\\ &\approx\widehat{Var}\left(\ln\hat{w}_k\right)+\left(\frac{1}{\widehat{VE}_k}\right)^2\widehat{Var}\left(\widehat{VE}_k\right)+2\sqrt{\widehat{Var}\left(\ln\hat{w}_k\right)\times\left(\frac{1}{\widehat{VE}_k}\right)^2\widehat{Var}\left(\widehat{VE}_k\right)}.\end{aligned}$$

Thus: $\widehat{Var}\left(\hat{w}_k\widehat{VE}_k\right) = \left(\hat{w}_k\widehat{VE}_k\right)^2\widehat{Var}\left(\ln\left(\hat{w}_k\widehat{VE}_k\right)\right)$.

Assumed that there is independency between naïve and non-naïve sub-population, the variance of numerator part of weighted average of vaccine efficacy, $\sum_{k=1}^{K}\hat{w}_k\widehat{VE}_k$ yield as follows:

$$\widehat{Var}\left(\sum_{k=1}^{K}\hat{w}_k\widehat{VE}_k\right)=\sum_{k=1}^{K}Var\left(\hat{w}_k\widehat{VE}_k\right)=\sum_{k=1}^{K}\left(\hat{w}_k\widehat{VE}_k\right)^2\widehat{Var}\left(\ln\left(\hat{w}_k\widehat{VE}_k\right)\right). \tag{42}$$

Meanwhile, the variance of denominator part of , weighted average vaccine efficacy, $\widehat{VE}_{\bullet} = \sum_{k=1}^{K}\hat{w}_k\widehat{VE}_k \Big/ \sum_{k-1}^{K}\hat{w}_k$ as follows:

$$\widehat{Var}\left(\sum_{k-1}^{K}\hat{w}_k\right)=\sum_{k-1}^{K}\widehat{Var}\left(\hat{w}_k\right). \tag{43}$$

The variance of weighted average vaccine efficacy $\widehat{VE}_{\bullet} = \sum_{k=1}^{K} \hat{w}_k \widehat{VE}_k \Big/ \sum_{k-1}^{K} \hat{w}_k$, described using natural logarithmic transformation as follow:

$$\widehat{Var}\left(\ln \widehat{VE}_{\bullet}\right) = \widehat{Var}\left(\ln\left(\sum_{k=1}^{K} \hat{w}_k \widehat{VE}_k\right)\right) + \widehat{Var}\left(\ln\left(\sum_{k-1}^{K} \hat{w}_k\right)\right) - 2Cov\left(\ln\left(\sum_{k=1}^{K} \hat{w}_k \widehat{VE}_k\right), \ln\left(\sum_{k-1}^{K} \hat{w}_k\right)\right),$$

$$\approx \left(\frac{1}{\sum_{k=1}^{K} \hat{w}_k \widehat{VE}_k}\right)^2 \widehat{Var}\left(\sum_{k=1}^{K} \hat{w}_k \widehat{VE}_k\right) + \left(\frac{1}{\sum_{k-1}^{K} \hat{w}_k}\right)^2 \widehat{Var}\left(\sum_{k-1}^{K} \hat{w}_k\right) -$$

$$2\sqrt{\left(\left(\frac{1}{\sum_{k=1}^{K} \hat{w}_k \widehat{VE}_k}\right)^2 \widehat{Var}\left(\sum_{k=1}^{K} \hat{w}_k \widehat{VE}_k\right)\right) \times \left(\left(\frac{1}{\sum_{k-1}^{K} \hat{w}_k}\right)^2 \widehat{Var}\left(\sum_{k-1}^{K} \hat{w}_k\right)\right)}.$$

Thus, the estimate variance of weighted average VE yields:

$$\widehat{Var}\left(\widehat{VE}_{\bullet}\right) = \left(\widehat{VE}_{\bullet}\right)^2 \widehat{Var}\left(\ln \widehat{VE}_{\bullet}\right),$$

$$\approx \left(\widehat{VE}_{\bullet}\right)^2 \left[\left(\frac{1}{\sum_{k=1}^{K} \hat{w}_k \widehat{VE}_k}\right)^2 \widehat{Var}\left(\sum_{k=1}^{K} \hat{w}_k \widehat{VE}_k\right) + \left(\frac{1}{\sum_{k-1}^{K} \hat{w}_k}\right)^2 \widehat{Var}\left(\sum_{k-1}^{K} \hat{w}_k\right) - 2\sqrt{\left(\left(\frac{1}{\sum_{k=1}^{K} \hat{w}_k \widehat{VE}_k}\right)^2 \widehat{Var}\left(\sum_{k=1}^{K} \hat{w}_k \widehat{VE}_k\right)\right) \times \left(\left(\frac{1}{\sum_{k-1}^{K} \hat{w}_k}\right)^2 \widehat{Var}\left(\sum_{k-1}^{K} \hat{w}_k\right)\right)}\right]. \quad (44)$$

A7. The Variance Derivation for Weighted Average Vaccine Efficacy based on Precision based approach for Person Time Approach

The estimate of weighted average vaccine efficacy for person time approach is defined as: $\widehat{VE}_{\bullet} = \sum_{k=1}^{K} \hat{w}_k \widehat{VE}_k \Big/ \sum_{k-1}^{K} \hat{w}_k$, with $\hat{w}_k = 1/\widehat{Var}\left(\widehat{VE}_k\right)$.

Where $\widehat{VE}_k$ is the estimate of vaccine efficacy for k^{th} sub-population that can be derive using the equation (12) and $\hat{w}_k$ is a weight applied to k^{th} sub-population. From equation describe in Appendix A3 and equation (16), the estimate of variance $\widehat{VE}_k$ can be describe as:

$$\widehat{Var}\left(\widehat{VE}_k\right) = \widehat{RR_k}^2 \widehat{Var}\left(\ln\left(\widehat{RR_k}\right)\right).$$
$$= \left(\frac{X_{vk}/T_{vk}}{X_{ck}/T_{ck}}\right)^2 \left(\frac{T_{vk}-X_{vk}}{T_{vk}X_{vk}} + \frac{T_{ck}-X_{ck}}{T_{ck}X_{ck}}\right), \quad (45)$$
$$= \left(\frac{X_{vk}/T_{vk}}{X_{ck}/T_{ck}}\right)^2 \left(\frac{1}{X_{vk}} + \frac{1}{T_{vk}} + \frac{1}{X_{ck}} + \frac{1}{T_{ck}}\right),$$

where $\widehat{VE}_k = 1 - \widehat{RR_k}$ and $\widehat{RR_k} = \left(\frac{X_{vk}/T_{vk}}{X_{ck}/T_{ck}}\right)^2$.

Thus, the weighted factor $\hat{w}_k$ can be defined as: $\hat{w}_k = 1/\widehat{Var}\left(\widehat{VE}_k\right) = \frac{1}{\widehat{RR_k}^2 \widehat{Var}\left(\ln\left(\widehat{RR_k}\right)\right)}$.

The variance of the weighted factor can be derived using natural logarithmic transformation as follows:

$$\ln\left(\hat{w}_k\right) = -2\ln\left(\widehat{RR_k}\right) - \ln\left(\widehat{Var}\left(\ln\left(\widehat{RR_k}\right)\right)\right).$$

The variance defined as:

$$Var\left(\ln\left(\hat{w}_k\right)\right) = 4Var\left(\ln\left(\widehat{RR_k}\right)\right) + Var\left(\ln\left(\widehat{Var}\left(\ln\left(\widehat{RR_k}\right)\right)\right)\right) + 4Cov\left(\ln\left(\widehat{RR_k}\right), \ln\left(\widehat{Var}\left(\ln\left(\widehat{RR_k}\right)\right)\right)\right),$$
$$\approx 4Var\left(\ln\left(\widehat{RR_k}\right)\right) + Var\left(\ln\left(\widehat{Var}\left(\ln\left(\widehat{RR_k}\right)\right)\right)\right) + 4\sqrt{Var\left(\ln\left(\widehat{RR_k}\right)\right) \times Var\left(\ln\left(\widehat{Var}\left(\ln\left(\widehat{RR_k}\right)\right)\right)\right)}. \quad (46)$$

From equation (45), the variance of $\left(\ln\left(\widehat{RR_k}\right)\right)$ can be describes as:

$$Var\left(\ln\left(\widehat{RR_k}\right)\right) = \left(\frac{1}{X_{vk}} + \frac{1}{T_{vk}} + \frac{1}{X_{ck}} + \frac{1}{T_{ck}}\right).$$

Since in vaccine study the total follow up time in both groups are usually large, so that the estimate of variance is approximated by the function of the number of cases in vaccinated and control group:

$$Var\left(\ln\left(\widehat{RR_k}\right)\right) \approx \left(\frac{1}{X_{vk}} + \frac{1}{X_{ck}}\right). \quad (47)$$

Assumed independency between vaccine and control group, the variance of $\widehat{Var}\left(\ln\left(\widehat{RR_k}\right)\right)$ defines as: $\widehat{Var}\left(\widehat{Var}\left(\ln\left(\widehat{RR_k}\right)\right)\right) = Var\left(\frac{1}{X_{vk}}\right) + Var\left(\frac{1}{X_{ck}}\right)$.

Where $IR_{vk} = X_{vk}/T_{vk}$ and $IR_{ck} = X_{ck}/T_{ck}$ follow Binomial distribution with parameter $Bin\left(T_{vk}, X_{vk}/T_{vk}\right)$ and $Bin\left(T_{ck}, X_{ck}/T_{ck}\right)$ respectively. Thus, yield:

$$
\begin{aligned}
\widehat{Var}\left(\widehat{Var}\left(\ln\left(\widehat{RR_k}\right)\right)\right) &= Var\left(\frac{1}{X_{vk}}\right)+Var\left(\frac{1}{X_{ck}}\right),\\
&\approx \left(d\frac{\left(X_{vk}\right)^{-1}}{dX_{vk}}\right)^2 Var\left(X_{vk}\right)+\left(d\frac{\left(X_{ck}\right)^{-1}}{dX_{ck}}\right)^2 Var\left(X_{ck}\right),\\
&\approx \left(-\left(X_{vk}\right)^{-2}\right)^2 Var\left(T_{vk}IR_{vk}\right)+\left(-\left(X_{ck}\right)^{-2}\right)^2 Var\left(T_{ck}IR_{ck}\right),\\
&\approx \left(X_{vk}\right)^{-4}\left(T_{vk}\right)^2\frac{\left(X_{vk}/T_{vk}\right)\left(1-X_{vk}/T_{vk}\right)}{T_{vk}}+\left(X_{ck}\right)^{-4}\left(T_{ck}\right)^2\frac{\left(X_{ck}/T_{ck}\right)\left(1-X_{ck}/T_{ck}\right)}{T_{ck}},\\
&\approx \left(\left(X_{vk}\right)^{-4}\left(1-X_{vk}/T_{vk}\right)X_{vk}\right)+\left(\left(X_{ck}\right)^{-4}\left(1-X_{ck}/T_{ck}\right)X_{ck}\right).
\end{aligned}
\tag{48}
$$

Thus, using Delta Methods the variance of $\ln\left(\widehat{Var}\left(\ln\left(\widehat{RR_k}\right)\right)\right)$ can be derives as:

$$
\begin{aligned}
\widehat{Var}\left(\ln\left(\widehat{Var}\left(\ln\left(\widehat{RR_k}\right)\right)\right)\right) &\approx \left(\frac{d\left(\ln\left(\widehat{Var}\left(\ln\left(\widehat{RR_k}\right)\right)\right)\right)}{d\widehat{Var}\left(\ln\left(\widehat{RR_k}\right)\right)}\right)^2 \widehat{Var}\left(\widehat{Var}\left(\ln\left(\widehat{RR_k}\right)\right)\right),\\
&\approx \left(\frac{1}{\widehat{Var}\left(\ln\left(\widehat{RR_k}\right)\right)}\right)^2 \widehat{Var}\left(\widehat{Var}\left(\ln\left(\widehat{RR_k}\right)\right)\right),\\
&\approx \frac{1}{\left(\frac{1}{\left(X_{vk}\right)}+\frac{1}{\left(X_{ck}\right)}\right)^2}\left(\frac{\left(1-X_{vk}/T_{vk}\right)X_{vk}}{\left(X_{vk}\right)^4}+\frac{\left(1-X_{ck}/T_{ck}\right)X_{ck}}{\left(X_{ck}\right)^4}\right).
\end{aligned}
\tag{49}
$$

The variance of $\hat{w}_k$ can be derives using Delta Methods as following with knowledge from equation (46) (47) and (49):

$$
\begin{aligned}
\widehat{Var}\left(\hat{w}_k\right) &= \left(\hat{w}_k\right)^2\widehat{Var}\left(\ln\hat{w}_k\right),\\
&\approx \left(\frac{1}{\widehat{RR_k}^2\widehat{Var}\left(\ln\left(\widehat{RR_k}\right)\right)}\right)^2 \widehat{Var}\left(\ln\hat{w}_k\right).
\end{aligned}
\tag{50}
$$

The weighted average vaccine efficacy based on precision approach for Person Time Approach defines as: $\widehat{VE}_{\bullet}=\sum_{k=1}^{K}\hat{w}_k\widehat{VE}_k\Big/\sum_{k-1}^{K}\hat{w}_k$. The variance of $\hat{w}_k\widehat{VE}_k$ for each k[th] sub-population is defines using natural logarithmic transformation as follow:

$$
\begin{aligned}
\widehat{Var}\left(\ln\left(\hat{w}_k\widehat{VE}_k\right)\right) &= Var\left(\ln\hat{w}_k\right)+Var\left(\ln\widehat{VE}_k\right)+2Cov\left(\hat{w}_k,\widehat{VE}_k\right),\\
&\approx \widehat{Var}\left(\ln\hat{w}_k\right)+\left(\frac{1}{\widehat{VE}_k}\right)^2\widehat{Var}\left(\widehat{VE}_k\right)+2\sqrt{\widehat{Var}\left(\ln\hat{w}_k\right)\times\left(\frac{1}{\widehat{VE}_k}\right)^2\widehat{Var}\left(\widehat{VE}_k\right)},
\end{aligned}
$$

where $\widehat{Var}\left(\ln \hat{w}_k\right)$ describes in equation (46) and $\widehat{Var}\left(\widehat{VE}_k\right)$ describes in equation (45). Thus, yield: $\widehat{Var}\left(\hat{w}_k \widehat{VE}_k\right) = \left(\hat{w}_k \widehat{VE}_k\right)^2 \widehat{Var}\left(\ln\left(\hat{w}_k \widehat{VE}_k\right)\right)$.

Assumed that there is independency between naïve and non-naïve sub-population, the variance of numerator part of weighted average of vaccine efficacy, $\sum_{k=1}^{K} \hat{w}_k \widehat{VE}_k$ yield as follow:

$$\widehat{Var}\left(\sum_{k=1}^{K} \hat{w}_k \widehat{VE}_k\right) = \sum_{k=1}^{K} Var\left(\hat{w}_k \widehat{VE}_k\right) = \sum_{k=1}^{K} \left(\hat{w}_k \widehat{VE}_k\right)^2 \widehat{Var}\left(\ln\left(\hat{w}_k \widehat{VE}_k\right)\right). \tag{51}$$

Meanwhile, the variance of denominator part of weighted average vaccine efficacy, $\sum_{k-1}^{K} \hat{w}_k$, as follow:

$$\widehat{Var}\left(\sum_{k-1}^{K} \hat{w}_k\right) = \sum_{k-1}^{K} \widehat{Var}\left(\hat{w}_k\right). \tag{52}$$

The variance of weighted average vaccine efficacy $\widehat{VE}\bullet = \sum_{k=1}^{K} \hat{w}_k \widehat{VE}_k \Big/ \sum_{k-1}^{K} \hat{w}_k$, describe using natural logarithmic transformation as follow:

$$\widehat{Var}\left(\ln \widehat{VE}\bullet\right) = \widehat{Var}\left(\ln\left(\sum_{k=1}^{K} \hat{w}_k \widehat{VE}_k\right)\right) + \widehat{Var}\left(\ln\left(\sum_{k-1}^{K} \hat{w}_k\right)\right) - 2Cov\left(\ln\left(\sum_{k=1}^{K} \hat{w}_k \widehat{VE}_k\right), \ln\left(\sum_{k-1}^{K} \hat{w}_k\right)\right),$$

$$\approx \left(\frac{1}{\sum_{k=1}^{K} \hat{w}_k \widehat{VE}_k}\right)^2 \widehat{Var}\left(\sum_{k=1}^{K} \hat{w}_k \widehat{VE}_k\right) + \left(\frac{1}{\sum_{k-1}^{K} \hat{w}_k}\right)^2 \widehat{Var}\left(\sum_{k-1}^{K} \hat{w}_k\right) -$$

$$2\sqrt{\left(\left(\frac{1}{\sum_{k=1}^{K} \hat{w}_k \widehat{VE}_k}\right)^2 \widehat{Var}\left(\sum_{k=1}^{K} \hat{w}_k \widehat{VE}_k\right)\right) \times \left(\left(\frac{1}{\sum_{k-1}^{K} \hat{w}_k}\right)^2 \widehat{Var}\left(\sum_{k-1}^{K} \hat{w}_k\right)\right)}.$$

Thus, the estimate variance of weighted average VE yields:

$$\widehat{Var}\left(\widehat{VE}\bullet\right) = \left(\widehat{VE}\bullet\right)^2 \widehat{Var}\left(\ln \widehat{VE}\bullet\right),$$

$$\approx \left(\widehat{VE}\bullet\right)^2 \left[\left(\frac{1}{\sum_{k=1}^{K} \hat{w}_k \widehat{VE}_k}\right)^2 \widehat{Var}\left(\sum_{k=1}^{K} \hat{w}_k \widehat{VE}_k\right) + \left(\frac{1}{\sum_{k-1}^{K} \hat{w}_k}\right)^2 \widehat{Var}\left(\sum_{k-1}^{K} \hat{w}_k\right) - \right.$$

$$\left. 2\sqrt{\left(\left(\frac{1}{\sum_{k=1}^{K} \hat{w}_k \widehat{VE}_k}\right)^2 \widehat{Var}\left(\sum_{k=1}^{K} \hat{w}_k \widehat{VE}_k\right)\right) \times \left(\left(\frac{1}{\sum_{k-1}^{K} \hat{w}_k}\right)^2 \widehat{Var}\left(\sum_{k-1}^{K} \hat{w}_k\right)\right)}\right]. \tag{53}$$

B. Tables and Figures

B1.1. Parameter Estimate of Vaccine efficacy for Attack Rate (AR) approach with n=2500

n	VE Aproach	Ratio naïve:non	VE naïve:non	IRc naïve:non	True Value	VE	std VE	se VE	bias squared	MSE	length CI	coverage CI
2500	raw	50:50	0.9:0.5	0.03:0.05	0.700	0.641	0.101	0.099	0.004	0.014	0.371	96.600
				0.03:0.1	0.700	0.597	0.088	0.083	0.011	0.018	0.314	79.500
			0.9:0.8	0.03:0.05	0.850	0.838	0.063	0.061	0.000	0.004	0.244	96.100
				0.03:0.1	0.850	0.824	0.051	0.050	0.001	0.003	0.197	92.700
		70:30	0.9:0.5	0.03:0.05	0.780	0.732	0.089	0.087	0.002	0.010	0.331	95.000
				0.03:0.1	0.780	0.663	0.089	0.084	0.014	0.022	0.318	71.400
			0.9:0.8	0.03:0.05	0.870	0.857	0.060	0.060	0.000	0.004	0.244	95.800
				0.03:0.1	0.870	0.841	0.056	0.053	0.001	0.004	0.213	92.400
		90:10	0.9:0.5	0.03:0.05	0.860	0.834	0.070	0.069	0.001	0.006	0.277	96.500
				0.03:0.1	0.860	0.792	0.072	0.073	0.005	0.010	0.287	84.400
			0.9:0.8	0.03:0.05	0.890	0.882	0.058	0.056	0.000	0.003	0.239	97.065
				0.03:0.1	0.890	0.870	0.056	0.056	0.000	0.004	0.230	94.800
	we_hal	50:50	0.9:0.5	0.03:0.05	0.654	0.640	0.102	0.100	0.000	0.011	0.372	97.100
				0.03:0.1	0.600	0.597	0.089	0.083	0.000	0.008	0.314	94.800
			0.9:0.8	0.03:0.05	0.840	0.838	0.063	0.061	0.000	0.004	0.244	96.200
				0.03:0.1	0.829	0.825	0.051	0.050	0.000	0.003	0.196	95.700
		70:30	0.9:0.5	0.03:0.05	0.736	0.730	0.090	0.087	0.000	0.008	0.332	95.900
				0.03:0.1	0.670	0.662	0.090	0.084	0.000	0.008	0.319	96.200
			0.9:0.8	0.03:0.05	0.861	0.857	0.060	0.060	0.000	0.004	0.244	97.000
				0.03:0.1	0.846	0.842	0.056	0.053	0.000	0.003	0.211	95.100
		90:10	0.9:0.5	0.03:0.05	0.839	0.834	0.070	0.069	0.000	0.005	0.278	97.800
				0.03:0.1	0.794	0.790	0.073	0.074	0.000	0.005	0.289	96.500
			0.9:0.8	0.03:0.05	0.886	0.882	0.058	0.057	0.000	0.003	0.239	97.773
				0.03:0.1	0.876	0.870	0.056	0.056	0.000	0.003	0.229	97.100
	we_N	50:50	0.9:0.5	0.03:0.05	0.700	0.687	0.093	0.092	0.000	0.009	0.345	96.900
				0.03:0.1	0.700	0.701	0.072	0.068	0.000	0.005	0.260	93.800
			0.9:0.8	0.03:0.05	0.850	0.848	0.061	0.060	0.000	0.004	0.242	96.100
				0.03:0.1	0.850	0.848	0.054	0.051	0.000	0.003	0.202	93.600
		70:30	0.9:0.5	0.03:0.05	0.780	0.771	0.084	0.080	0.000	0.007	0.310	96.700
				0.03:0.1	0.780	0.775	0.065	0.064	0.000	0.004	0.249	95.300
			0.9:0.8	0.03:0.05	0.870	0.866	0.059	0.059	0.000	0.004	0.242	96.400
				0.03:0.1	0.870	0.868	0.055	0.052	0.000	0.003	0.209	93.900
		90:10	0.9:0.5	0.03:0.05	0.860	0.845	0.076	0.075	0.000	0.006	0.294	98.100
				0.03:0.1	0.860	0.856	0.060	0.059	0.000	0.004	0.237	95.600
			0.9:0.8	0.03:0.05	0.890	0.884	0.060	0.058	0.000	0.004	0.247	96.954
				0.03:0.1	0.890	0.886	0.055	0.053	0.000	0.003	0.226	96.000
	we_pres	50:50	0.9:0.5	0.03:0.05	0.700	0.795	0.100	0.087	0.009	0.019	0.349	79.800
				0.03:0.1	0.700	0.755	0.098	0.083	0.003	0.013	0.327	81.700
			0.9:0.8	0.03:0.05	0.850	0.861	0.058	0.061	0.000	0.003	0.253	83.800
				0.03:0.1	0.850	0.847	0.048	0.052	0.000	0.002	0.209	84.200
		70:30	0.9:0.5	0.03:0.05	0.780	0.856	0.077	0.068	0.006	0.012	0.287	86.600
				0.03:0.1	0.780	0.827	0.083	0.070	0.002	0.009	0.286	91.900
			0.9:0.8	0.03:0.05	0.870	0.882	0.056	0.057	0.000	0.003	0.245	90.900
				0.03:0.1	0.870	0.870	0.055	0.052	0.000	0.003	0.217	89.800
		90:10	0.9:0.5	0.03:0.05	0.860	0.887	0.057	0.058	0.001	0.004	0.248	92.600
				0.03:0.1	0.860	0.877	0.062	0.058	0.000	0.004	0.248	94.900
			0.9:0.8	0.03:0.05	0.890	0.891	0.055	0.056	0.000	0.003	0.241	98.678
				0.03:0.1	0.890	0.891	0.051	0.053	0.000	0.003	0.226	88.400

VE approach: raw: Raw of vaccine efficacy (Equation (3)), we_hal: weighted average vaccine efficacy based on Halloran approach (Equation (5)), we_N: general weighted average vaccine efficacy based on sample size for each sub-population (Equation (7)), we_pres : weighted average vaccine efficacy based on precision approach (Equation (9)).

B1.2. Parameter Estimate of Vaccine efficacy for Attack Rate (AR) approach with n=5000

n	VE Aproach	Ratio naïve:non	VE naïve:non	IRc naïve:non	True Value	VE	std VE	se VE	bias squared	MSE	length CI	coverage CI
5000	raw	50:50	0.9:0.5	0.03:0.05	0.700	0.655	0.069	0.068	0.002	0.007	0.261	92.300
				0.03:0.1	0.700	0.598	0.056	0.059	0.010	0.014	0.226	57.400
			0.9:0.8	0.03:0.05	0.850	0.837	0.043	0.043	0.000	0.002	0.171	94.700
				0.03:0.1	0.850	0.825	0.035	0.035	0.001	0.002	0.139	89.000
		70:30	0.9:0.5	0.03:0.05	0.780	0.733	0.061	0.061	0.002	0.006	0.236	90.000
				0.03:0.1	0.780	0.671	0.058	0.058	0.012	0.015	0.225	46.800
			0.9:0.8	0.03:0.05	0.870	0.859	0.043	0.042	0.000	0.002	0.167	94.700
				0.03:0.1	0.870	0.843	0.038	0.037	0.001	0.002	0.148	89.200
		90:10	0.9:0.5	0.03:0.05	0.860	0.838	0.046	0.048	0.000	0.003	0.191	94.200
				0.03:0.1	0.860	0.797	0.051	0.051	0.004	0.007	0.201	72.900
			0.9:0.8	0.03:0.05	0.890	0.886	0.039	0.039	0.000	0.002	0.160	96.400
				0.03:0.1	0.890	0.873	0.040	0.039	0.000	0.002	0.157	92.600
	we_hal	50:50	0.9:0.5	0.03:0.05	0.654	0.654	0.070	0.068	0.000	0.005	0.261	95.600
				0.03:0.1	0.600	0.598	0.057	0.059	0.000	0.003	0.226	95.800
			0.9:0.8	0.03:0.05	0.840	0.837	0.043	0.043	0.000	0.002	0.171	95.800
				0.03:0.1	0.829	0.826	0.035	0.035	0.000	0.001	0.138	94.300
		70:30	0.9:0.5	0.03:0.05	0.736	0.732	0.062	0.061	0.000	0.004	0.237	95.900
				0.03:0.1	0.670	0.670	0.059	0.059	0.000	0.003	0.226	95.900
			0.9:0.8	0.03:0.05	0.861	0.859	0.043	0.042	0.000	0.002	0.167	95.100
				0.03:0.1	0.846	0.844	0.037	0.037	0.000	0.001	0.147	95.700
		90:10	0.9:0.5	0.03:0.05	0.839	0.838	0.047	0.048	0.000	0.002	0.192	96.800
				0.03:0.1	0.794	0.795	0.052	0.052	0.000	0.003	0.202	96.100
			0.9:0.8	0.03:0.05	0.886	0.886	0.039	0.039	0.000	0.002	0.160	96.700
				0.03:0.1	0.876	0.873	0.040	0.039	0.000	0.002	0.157	95.400
	we_N	50:50	0.9:0.5	0.03:0.05	0.700	0.703	0.063	0.061	0.000	0.004	0.236	95.200
				0.03:0.1	0.700	0.703	0.047	0.047	0.000	0.002	0.184	94.800
			0.9:0.8	0.03:0.05	0.850	0.848	0.041	0.042	0.000	0.002	0.167	95.800
				0.03:0.1	0.850	0.851	0.035	0.035	0.000	0.001	0.138	94.800
		70:30	0.9:0.5	0.03:0.05	0.780	0.776	0.055	0.055	0.000	0.003	0.213	96.100
				0.03:0.1	0.780	0.784	0.044	0.044	0.000	0.002	0.171	94.300
			0.9:0.8	0.03:0.05	0.870	0.870	0.041	0.040	0.000	0.002	0.161	95.800
				0.03:0.1	0.870	0.871	0.037	0.036	0.000	0.001	0.145	94.400
		90:10	0.9:0.5	0.03:0.05	0.860	0.857	0.046	0.046	0.000	0.002	0.185	96.500
				0.03:0.1	0.860	0.862	0.041	0.040	0.000	0.002	0.160	94.600
			0.9:0.8	0.03:0.05	0.890	0.890	0.039	0.039	0.000	0.002	0.160	96.900
				0.03:0.1	0.890	0.890	0.040	0.037	0.000	0.002	0.152	94.600
	we_pres	50:50	0.9:0.5	0.03:0.05	0.700	0.826	0.078	0.058	0.016	0.022	0.237	55.700
				0.03:0.1	0.700	0.778	0.086	0.059	0.006	0.014	0.236	78.100
			0.9:0.8	0.03:0.05	0.850	0.867	0.047	0.043	0.000	0.003	0.175	96.000
				0.03:0.1	0.850	0.851	0.040	0.036	0.000	0.002	0.146	91.900
		70:30	0.9:0.5	0.03:0.05	0.780	0.864	0.058	0.047	0.007	0.010	0.194	68.600
				0.03:0.1	0.780	0.836	0.065	0.049	0.003	0.007	0.198	81.100
			0.9:0.8	0.03:0.05	0.870	0.885	0.043	0.040	0.000	0.002	0.164	95.900
				0.03:0.1	0.870	0.869	0.042	0.037	0.000	0.002	0.148	94.500
		90:10	0.9:0.5	0.03:0.05	0.860	0.892	0.043	0.040	0.001	0.003	0.165	90.300
				0.03:0.1	0.860	0.883	0.047	0.041	0.001	0.003	0.166	91.700
			0.9:0.8	0.03:0.05	0.890	0.899	0.038	0.038	0.000	0.002	0.157	90.000
				0.03:0.1	0.890	0.892	0.041	0.037	0.000	0.002	0.152	94.400

VE approach: raw: Raw of vaccine efficacy (Equation (3)), we_hal: weighted average vaccine efficacy based on Halloran approach (Equation (5)), we_N: general weighted average vaccine efficacy based on sample size for each sub-population (Equation (7)), we_pres : weighted average vaccine efficacy based on precision approach (Equation (9)).

B1.3. Parameter Estimate of Vaccine efficacy for Attack Rate (AR) approach with n=10000

n	VE Aproach	Ratio naïve:non	VE naïve:non	IRc naïve:non	True Value	VE	std VE	se VE	bias squared	MSE	length CI	coverage CI
10000	raw	50:50	0.9:0.5	0.03:0.05	0.700	0.655	0.046	0.048	0.002	0.004	0.186	87.400
				0.03:0.1	0.700	0.600	0.042	0.041	0.010	0.012	0.161	26.300
			0.9:0.8	0.03:0.05	0.850	0.839	0.030	0.030	0.000	0.001	0.119	94.100
				0.03:0.1	0.850	0.828	0.025	0.025	0.000	0.001	0.097	84.500
		70:30	0.9:0.5	0.03:0.05	0.780	0.736	0.046	0.043	0.002	0.004	0.167	82.400
				0.03:0.1	0.780	0.670	0.041	0.041	0.012	0.014	0.160	15.400
			0.9:0.8	0.03:0.05	0.870	0.860	0.029	0.030	0.000	0.001	0.117	94.800
				0.03:0.1	0.870	0.844	0.027	0.026	0.001	0.001	0.104	82.100
		90:10	0.9:0.5	0.03:0.05	0.860	0.838	0.036	0.034	0.000	0.002	0.135	89.300
				0.03:0.1	0.860	0.793	0.037	0.037	0.004	0.006	0.143	45.500
			0.9:0.8	0.03:0.05	0.890	0.886	0.029	0.028	0.000	0.001	0.111	94.700
				0.03:0.1	0.890	0.874	0.028	0.027	0.000	0.001	0.109	90.400
	we_hal	50:50	0.9:0.5	0.03:0.05	0.654	0.654	0.047	0.048	0.000	0.002	0.186	96.200
				0.03:0.1	0.600	0.600	0.042	0.041	0.000	0.002	0.161	94.400
			0.9:0.8	0.03:0.05	0.840	0.839	0.030	0.030	0.000	0.001	0.119	96.100
				0.03:0.1	0.829	0.829	0.025	0.024	0.000	0.001	0.096	95.000
		70:30	0.9:0.5	0.03:0.05	0.736	0.734	0.046	0.043	0.000	0.002	0.168	93.700
				0.03:0.1	0.670	0.669	0.042	0.041	0.000	0.002	0.161	95.600
			0.9:0.8	0.03:0.05	0.861	0.860	0.029	0.030	0.000	0.001	0.117	96.800
				0.03:0.1	0.846	0.845	0.027	0.026	0.000	0.001	0.103	93.800
		90:10	0.9:0.5	0.03:0.05	0.839	0.838	0.036	0.034	0.000	0.001	0.135	94.300
				0.03:0.1	0.794	0.792	0.038	0.037	0.000	0.001	0.144	94.700
			0.9:0.8	0.03:0.05	0.886	0.886	0.029	0.028	0.000	0.001	0.111	95.500
				0.03:0.1	0.876	0.875	0.028	0.027	0.000	0.001	0.109	95.500
	we_N	50:50	0.9:0.5	0.03:0.05	0.700	0.704	0.042	0.043	0.000	0.002	0.167	95.900
				0.03:0.1	0.700	0.706	0.033	0.033	0.000	0.001	0.129	94.100
			0.9:0.8	0.03:0.05	0.850	0.851	0.028	0.029	0.000	0.001	0.115	96.500
				0.03:0.1	0.850	0.854	0.025	0.024	0.000	0.001	0.096	94.200
		70:30	0.9:0.5	0.03:0.05	0.780	0.781	0.040	0.038	0.000	0.002	0.148	94.000
				0.03:0.1	0.780	0.784	0.031	0.031	0.000	0.001	0.120	94.900
			0.9:0.8	0.03:0.05	0.870	0.871	0.028	0.028	0.000	0.001	0.112	96.700
				0.03:0.1	0.870	0.872	0.026	0.025	0.000	0.001	0.101	93.500
		90:10	0.9:0.5	0.03:0.05	0.860	0.859	0.033	0.032	0.000	0.001	0.125	94.600
				0.03:0.1	0.860	0.861	0.028	0.028	0.000	0.001	0.112	95.600
			0.9:0.8	0.03:0.05	0.890	0.891	0.028	0.027	0.000	0.001	0.109	95.300
				0.03:0.1	0.890	0.891	0.027	0.026	0.000	0.001	0.105	95.000
	we_pres	50:50	0.9:0.5	0.03:0.05	0.700	0.830	0.058	0.041	0.017	0.020	0.165	30.000
				0.03:0.1	0.700	0.778	0.063	0.042	0.006	0.010	0.166	59.600
			0.9:0.8	0.03:0.05	0.850	0.864	0.035	0.030	0.000	0.001	0.120	93.500
				0.03:0.1	0.850	0.851	0.031	0.025	0.000	0.001	0.100	93.100
		70:30	0.9:0.5	0.03:0.05	0.780	0.863	0.041	0.034	0.007	0.009	0.136	46.000
				0.03:0.1	0.780	0.834	0.049	0.035	0.003	0.005	0.139	68.200
			0.9:0.8	0.03:0.05	0.870	0.882	0.031	0.029	0.000	0.001	0.114	94.100
				0.03:0.1	0.870	0.867	0.031	0.026	0.000	0.001	0.102	91.900
		90:10	0.9:0.5	0.03:0.05	0.860	0.890	0.032	0.029	0.001	0.002	0.115	83.700
				0.03:0.1	0.860	0.881	0.033	0.029	0.000	0.002	0.117	88.600
			0.9:0.8	0.03:0.05	0.890	0.898	0.029	0.027	0.000	0.001	0.109	94.100
				0.03:0.1	0.890	0.890	0.028	0.026	0.000	0.001	0.105	94.300

VE approach: raw: Raw of vaccine efficacy (Equation (3)), we_hal: weighted average vaccine efficacy based on Halloran approach (Equation (5)), we_N: general weighted average vaccine efficacy based on sample size for each sub-population (Equation (7)), we_pres : weighted average vaccine efficacy based on precision approach (Equation (9)).

B2.1. Parameter Estimate of Vaccine efficacy for Person Time Approach with n=2500

n	VE Aproach	Ratio naïve:non	VE naïve:non	IRc naïve:non	True Value	VE	std VE	se VE	bias squared	MSE	length CI	coverage CI
2500	raw	50:50	0.9:0.5	0.03:0.05	0.700	0.640	0.101	0.100	0.004	0.014	0.434	92.400
				0.03:0.1	0.700	0.596	0.088	0.084	0.011	0.019	0.350	75.500
			0.9:0.8	0.03:0.05	0.850	0.838	0.063	0.061	0.000	0.004	0.275	96.200
				0.03:0.1	0.850	0.824	0.051	0.050	0.001	0.003	0.216	93.600
		70:30	0.9:0.5	0.03:0.05	0.780	0.731	0.090	0.087	0.002	0.010	0.386	93.700
				0.03:0.1	0.780	0.662	0.089	0.084	0.014	0.022	0.361	69.300
			0.9:0.8	0.03:0.05	0.870	0.857	0.061	0.060	0.000	0.004	0.276	96.200
				0.03:0.1	0.870	0.840	0.056	0.053	0.001	0.004	0.236	93.100
		90:10	0.9:0.5	0.03:0.05	0.860	0.834	0.070	0.069	0.001	0.006	0.320	96.700
				0.03:0.1	0.860	0.792	0.072	0.074	0.005	0.010	0.330	85.900
			0.9:0.8	0.03:0.05	0.890	0.883	0.059	0.057	0.000	0.004	0.269	97.874
				0.03:0.1	0.890	0.870	0.056	0.056	0.000	0.004	0.259	95.700
	we_hal	50:50	0.9:0.5	0.03:0.05	0.650	0.639	0.102	0.102	0.000	0.011	0.379	82.900
				0.03:0.1	0.592	0.596	0.089	0.085	0.000	0.008	0.320	80.000
			0.9:0.8	0.03:0.05	0.837	0.838	0.063	0.063	0.000	0.004	0.250	82.700
				0.03:0.1	0.823	0.825	0.051	0.051	0.000	0.003	0.200	83.700
		70:30	0.9:0.5	0.03:0.05	0.733	0.730	0.090	0.089	0.000	0.008	0.338	88.900
				0.03:0.1	0.665	0.661	0.090	0.085	0.000	0.008	0.323	90.400
			0.9:0.8	0.03:0.05	0.858	0.857	0.060	0.061	0.000	0.004	0.248	90.000
				0.03:0.1	0.841	0.841	0.056	0.054	0.000	0.003	0.215	88.100
		90:10	0.9:0.5	0.03:0.05	0.837	0.833	0.070	0.071	0.000	0.005	0.283	91.100
				0.03:0.1	0.792	0.790	0.073	0.075	0.000	0.005	0.293	93.500
			0.9:0.8	0.03:0.05	0.884	0.882	0.059	0.061	0.000	0.003	0.251	97.654
				0.03:0.1	0.873	0.870	0.056	0.057	0.000	0.003	0.234	88.000
	we_N	50:50	0.9:0.5	0.03:0.05	0.700	0.686	0.093	0.094	0.000	0.009	0.606	84.400
				0.03:0.1	0.700	0.701	0.072	0.071	0.000	0.005	0.467	83.000
			0.9:0.8	0.03:0.05	0.850	0.847	0.061	0.063	0.000	0.004	0.445	85.000
				0.03:0.1	0.850	0.848	0.054	0.054	0.000	0.003	0.371	86.700
		70:30	0.9:0.5	0.03:0.05	0.780	0.771	0.084	0.083	0.000	0.007	0.553	91.300
				0.03:0.1	0.780	0.775	0.065	0.066	0.000	0.004	0.433	93.500
			0.9:0.8	0.03:0.05	0.870	0.866	0.060	0.061	0.000	0.004	0.423	92.300
				0.03:0.1	0.870	0.868	0.055	0.054	0.000	0.003	0.354	91.800
		90:10	0.9:0.5	0.03:0.05	0.860	0.845	0.076	0.077	0.000	0.006	0.853	93.200
				0.03:0.1	0.860	0.855	0.060	0.060	0.000	0.004	0.399	95.900
			0.9:0.8	0.03:0.05	0.890	0.885	0.061	0.063	0.000	0.004	0.591	99.707
				0.03:0.1	0.890	0.887	0.055	0.055	0.000	0.003	0.344	89.500
	we_pres	50:50	0.9:0.5	0.03:0.05	0.700	0.795	0.100	0.176	0.009	0.019	0.504	81.800
				0.03:0.1	0.700	0.756	0.098	0.206	0.003	0.013	0.432	81.600
			0.9:0.8	0.03:0.05	0.850	0.861	0.058	0.094	0.000	0.003	0.401	85.000
				0.03:0.1	0.850	0.847	0.048	0.088	0.000	0.002	0.313	86.000
		70:30	0.9:0.5	0.03:0.05	0.780	0.856	0.077	0.109	0.006	0.012	0.396	86.100
				0.03:0.1	0.780	0.827	0.083	0.140	0.002	0.009	0.378	91.300
			0.9:0.8	0.03:0.05	0.870	0.882	0.056	0.079	0.000	0.003	0.380	92.300
				0.03:0.1	0.870	0.870	0.055	0.080	0.000	0.003	0.325	91.700
		90:10	0.9:0.5	0.03:0.05	0.860	0.887	0.057	0.068	0.001	0.004	0.342	92.600
				0.03:0.1	0.860	0.877	0.062	0.079	0.000	0.004	0.335	95.900
			0.9:0.8	0.03:0.05	0.890	0.891	0.055	0.064	0.000	0.003	0.362	99.707
				0.03:0.1	0.890	0.890	0.052	0.067	0.000	0.003	0.343	90.000
	we_t	50:50	0.9:0.5	0.03:0.05	0.700	0.687	0.092	0.094	0.000	0.009	0.605	84.400
				0.03:0.1	0.700	0.703	0.072	0.071	0.000	0.005	0.468	83.100
			0.9:0.8	0.03:0.05	0.850	0.848	0.061	0.063	0.000	0.004	0.445	85.000
				0.03:0.1	0.850	0.848	0.055	0.054	0.000	0.003	0.373	86.700
		70:30	0.9:0.5	0.03:0.05	0.780	0.771	0.084	0.082	0.000	0.007	0.551	91.300
				0.03:0.1	0.780	0.777	0.064	0.065	0.000	0.004	0.431	93.500
			0.9:0.8	0.03:0.05	0.870	0.866	0.060	0.061	0.000	0.004	0.423	92.300
				0.03:0.1	0.870	0.868	0.056	0.054	0.000	0.003	0.354	91.800
		90:10	0.9:0.5	0.03:0.05	0.860	0.845	0.076	0.077	0.000	0.006	0.851	93.200
				0.03:0.1	0.860	0.856	0.060	0.060	0.000	0.004	0.396	95.900
			0.9:0.8	0.03:0.05	0.890	0.886	0.061	0.063	0.000	0.004	0.590	99.707
				0.03:0.1	0.890	0.887	0.055	0.055	0.000	0.003	0.343	89.500

VE approach: raw: Raw of vaccine efficacy (Equation (12)), we_hal: weighted average vaccine efficacy based on Halloran approach (Equation (17)), we_N: general weighted average vaccine efficacy based on sample size for each sub-population (Equation (7)), we_pres : weighted average vaccine efficacy based on precision approach (Equation (9)), we_t: general weighted average vaccine efficacy based on total follow up time for each sub-population (Equation (19)).

B2.2. Parameter Estimate of Vaccine efficacy for Person Time Approach with n=5000

n	VE Aproach	Ratio naïve:non	VE naïve:non	IRc naïve:non	True Value	VE	std VE	se VE	bias squared	MSE	length CI	coverage CI
5000	raw	50:50	0.9:0.5	0.03:0.05	0.700	0.655	0.069	0.068	0.002	0.007	0.285	90.100
				0.03:0.1	0.700	0.598	0.056	0.059	0.010	0.014	0.241	54.300
			0.9:0.8	0.03:0.05	0.850	0.837	0.043	0.043	0.000	0.002	0.186	95.500
				0.03:0.1	0.850	0.825	0.035	0.035	0.001	0.002	0.147	89.900
		70:30	0.9:0.5	0.03:0.05	0.780	0.733	0.061	0.061	0.002	0.006	0.259	89.300
				0.03:0.1	0.780	0.670	0.058	0.059	0.012	0.015	0.243	45.900
			0.9:0.8	0.03:0.05	0.870	0.859	0.043	0.042	0.000	0.002	0.182	95.300
				0.03:0.1	0.870	0.843	0.038	0.038	0.001	0.002	0.159	90.100
		90:10	0.9:0.5	0.03:0.05	0.860	0.838	0.046	0.048	0.000	0.003	0.209	95.100
				0.03:0.1	0.860	0.796	0.051	0.051	0.004	0.007	0.219	75.800
			0.9:0.8	0.03:0.05	0.890	0.886	0.039	0.039	0.000	0.002	0.174	96.900
				0.03:0.1	0.890	0.873	0.040	0.039	0.000	0.002	0.170	94.100
	we_hal	50:50	0.9:0.5	0.03:0.05	0.650	0.654	0.070	0.068	0.000	0.005	0.263	93.400
				0.03:0.1	0.592	0.597	0.057	0.059	0.000	0.003	0.228	94.500
			0.9:0.8	0.03:0.05	0.837	0.837	0.043	0.044	0.000	0.002	0.172	94.700
				0.03:0.1	0.823	0.826	0.035	0.035	0.000	0.001	0.139	92.100
		70:30	0.9:0.5	0.03:0.05	0.733	0.731	0.062	0.062	0.000	0.004	0.238	95.600
				0.03:0.1	0.665	0.669	0.058	0.059	0.000	0.003	0.228	95.600
			0.9:0.8	0.03:0.05	0.858	0.859	0.043	0.042	0.000	0.002	0.168	94.800
				0.03:0.1	0.841	0.843	0.038	0.038	0.000	0.001	0.149	95.100
		90:10	0.9:0.5	0.03:0.05	0.837	0.838	0.047	0.049	0.000	0.002	0.193	96.600
				0.03:0.1	0.792	0.795	0.052	0.052	0.000	0.003	0.203	96.300
			0.9:0.8	0.03:0.05	0.884	0.886	0.039	0.040	0.000	0.002	0.162	90.100
				0.03:0.1	0.873	0.873	0.040	0.039	0.000	0.002	0.158	94.900
	we_N	50:50	0.9:0.5	0.03:0.05	0.700	0.703	0.063	0.062	0.000	0.004	0.363	97.000
				0.03:0.1	0.700	0.703	0.047	0.048	0.000	0.002	0.290	98.100
			0.9:0.8	0.03:0.05	0.850	0.848	0.041	0.042	0.000	0.002	0.269	98.000
				0.03:0.1	0.850	0.850	0.035	0.035	0.000	0.001	0.223	96.600
		70:30	0.9:0.5	0.03:0.05	0.780	0.776	0.055	0.055	0.000	0.003	0.339	99.700
				0.03:0.1	0.780	0.784	0.044	0.044	0.000	0.002	0.271	99.800
			0.9:0.8	0.03:0.05	0.870	0.869	0.041	0.040	0.000	0.002	0.257	99.200
				0.03:0.1	0.870	0.870	0.037	0.036	0.000	0.001	0.221	98.600
		90:10	0.9:0.5	0.03:0.05	0.860	0.856	0.046	0.046	0.000	0.002	0.308	99.400
				0.03:0.1	0.860	0.861	0.041	0.040	0.000	0.002	0.242	99.600
			0.9:0.8	0.03:0.05	0.890	0.890	0.039	0.039	0.000	0.002	0.250	92.400
				0.03:0.1	0.890	0.889	0.040	0.038	0.000	0.002	0.211	98.200
	we_pres	50:50	0.9:0.5	0.03:0.05	0.700	0.826	0.078	0.128	0.016	0.022	0.292	69.500
				0.03:0.1	0.700	0.779	0.086	0.155	0.006	0.014	0.268	76.900
			0.9:0.8	0.03:0.05	0.850	0.867	0.047	0.068	0.000	0.003	0.251	97.800
				0.03:0.1	0.850	0.851	0.040	0.063	0.000	0.002	0.203	96.000
		70:30	0.9:0.5	0.03:0.05	0.780	0.865	0.058	0.079	0.007	0.010	0.245	77.900
				0.03:0.1	0.780	0.836	0.065	0.103	0.003	0.007	0.240	83.700
			0.9:0.8	0.03:0.05	0.870	0.885	0.043	0.055	0.000	0.002	0.235	98.200
				0.03:0.1	0.870	0.869	0.042	0.055	0.000	0.002	0.213	97.400
		90:10	0.9:0.5	0.03:0.05	0.860	0.892	0.043	0.048	0.001	0.003	0.208	93.500
				0.03:0.1	0.860	0.883	0.047	0.056	0.001	0.003	0.209	93.900
			0.9:0.8	0.03:0.05	0.890	0.898	0.038	0.044	0.000	0.002	0.222	91.500
				0.03:0.1	0.890	0.892	0.041	0.046	0.000	0.002	0.217	97.400
	we_t	50:50	0.9:0.5	0.03:0.05	0.700	0.704	0.063	0.061	0.000	0.004	0.362	97.100
				0.03:0.1	0.700	0.705	0.046	0.048	0.000	0.002	0.290	98.100
			0.9:0.8	0.03:0.05	0.850	0.848	0.041	0.042	0.000	0.002	0.269	98.100
				0.03:0.1	0.850	0.851	0.035	0.036	0.000	0.001	0.224	96.500
		70:30	0.9:0.5	0.03:0.05	0.780	0.777	0.054	0.055	0.000	0.003	0.338	99.700
				0.03:0.1	0.780	0.786	0.044	0.044	0.000	0.002	0.270	99.800
			0.9:0.8	0.03:0.05	0.870	0.870	0.041	0.040	0.000	0.002	0.257	99.200
				0.03:0.1	0.870	0.871	0.037	0.037	0.000	0.001	0.221	98.600
		90:10	0.9:0.5	0.03:0.05	0.860	0.857	0.046	0.046	0.000	0.002	0.307	99.400
				0.03:0.1	0.860	0.862	0.041	0.040	0.000	0.002	0.240	99.500
			0.9:0.8	0.03:0.05	0.890	0.890	0.039	0.039	0.000	0.002	0.249	92.400
				0.03:0.1	0.890	0.890	0.040	0.038	0.000	0.002	0.211	98.200

VE approach: raw: Raw of vaccine efficacy (Equation (12)), we_hal: weighted average vaccine efficacy based on Halloran approach (Equation (17)), we_N: general weighted average vaccine efficacy based on sample size for each sub-population (Equation (7)), we_pres : weighted average vaccine efficacy based on precision approach (Equation (9)), we_t: general weighted average vaccine efficacy based on total follow up time for each sub-population (Equation (19)).

B2.3. Parameter Estimate of Vaccine efficacy for Person Time Approach with n=10000

n	VE Aproach	Ratio naïve:non	VE naïve:non	IRc naïve:non	True Value	VE	std VE	se VE	bias squared	MSE	length CI	coverage CI
10000	raw	50:50	0.9:0.5	0.03:0.05	0.700	0.655	0.046	0.048	0.002	0.004	0.196	86.300
				0.03:0.1	0.700	0.600	0.042	0.042	0.010	0.012	0.167	24.700
			0.9:0.8	0.03:0.05	0.850	0.839	0.030	0.030	0.000	0.001	0.126	94.900
				0.03:0.1	0.850	0.827	0.025	0.025	0.001	0.001	0.101	85.500
		70:30	0.9:0.5	0.03:0.05	0.780	0.735	0.046	0.043	0.002	0.004	0.177	82.500
				0.03:0.1	0.780	0.669	0.041	0.041	0.012	0.014	0.168	15.300
			0.9:0.8	0.03:0.05	0.870	0.860	0.029	0.030	0.000	0.001	0.124	95.400
				0.03:0.1	0.870	0.844	0.027	0.026	0.001	0.001	0.109	83.800
		90:10	0.9:0.5	0.03:0.05	0.860	0.838	0.036	0.034	0.000	0.002	0.143	90.600
				0.03:0.1	0.860	0.793	0.037	0.037	0.004	0.006	0.152	48.200
			0.9:0.8	0.03:0.05	0.890	0.886	0.029	0.028	0.000	0.001	0.118	95.800
				0.03:0.1	0.890	0.874	0.028	0.028	0.000	0.001	0.115	91.700
	we_hal	50:50	0.9:0.5	0.03:0.05	0.650	0.654	0.046	0.048	0.000	0.002	0.187	96.200
				0.03:0.1	0.592	0.599	0.043	0.042	0.000	0.002	0.162	94.000
			0.9:0.8	0.03:0.05	0.837	0.839	0.030	0.030	0.000	0.001	0.120	96.100
				0.03:0.1	0.823	0.829	0.025	0.025	0.000	0.001	0.097	94.100
		70:30	0.9:0.5	0.03:0.05	0.733	0.734	0.046	0.043	0.000	0.002	0.169	93.800
				0.03:0.1	0.665	0.668	0.042	0.042	0.000	0.002	0.162	95.000
			0.9:0.8	0.03:0.05	0.858	0.860	0.029	0.030	0.000	0.001	0.118	96.800
				0.03:0.1	0.841	0.845	0.027	0.026	0.000	0.001	0.104	94.500
		90:10	0.9:0.5	0.03:0.05	0.837	0.838	0.036	0.034	0.000	0.001	0.135	94.300
				0.03:0.1	0.792	0.792	0.038	0.037	0.000	0.001	0.145	94.900
			0.9:0.8	0.03:0.05	0.884	0.886	0.029	0.028	0.000	0.001	0.112	95.500
				0.03:0.1	0.873	0.874	0.028	0.027	0.000	0.001	0.109	95.900
	we_N	50:50	0.9:0.5	0.03:0.05	0.700	0.704	0.042	0.043	0.000	0.002	0.242	99.800
				0.03:0.1	0.700	0.706	0.033	0.033	0.000	0.001	0.194	99.400
			0.9:0.8	0.03:0.05	0.850	0.851	0.029	0.029	0.000	0.001	0.176	99.800
				0.03:0.1	0.850	0.853	0.025	0.024	0.000	0.001	0.146	99.700
		70:30	0.9:0.5	0.03:0.05	0.780	0.780	0.040	0.038	0.000	0.002	0.225	99.000
				0.03:0.1	0.780	0.783	0.031	0.031	0.000	0.001	0.183	99.800
			0.9:0.8	0.03:0.05	0.870	0.871	0.028	0.028	0.000	0.001	0.171	99.600
				0.03:0.1	0.870	0.871	0.026	0.025	0.000	0.001	0.147	99.700
		90:10	0.9:0.5	0.03:0.05	0.860	0.858	0.033	0.032	0.000	0.001	0.193	99.500
				0.03:0.1	0.860	0.861	0.028	0.028	0.000	0.001	0.162	99.600
			0.9:0.8	0.03:0.05	0.890	0.891	0.028	0.027	0.000	0.001	0.157	99.400
				0.03:0.1	0.890	0.891	0.027	0.026	0.000	0.001	0.139	99.400
	we_pres	50:50	0.9:0.5	0.03:0.05	0.700	0.830	0.058	0.093	0.017	0.020	0.192	46.200
				0.03:0.1	0.700	0.779	0.063	0.114	0.006	0.010	0.181	65.200
			0.9:0.8	0.03:0.05	0.850	0.864	0.035	0.047	0.000	0.001	0.169	96.500
				0.03:0.1	0.850	0.851	0.031	0.043	0.000	0.001	0.137	96.500
		70:30	0.9:0.5	0.03:0.05	0.780	0.864	0.041	0.058	0.007	0.009	0.165	59.900
				0.03:0.1	0.780	0.835	0.049	0.075	0.003	0.005	0.162	75.300
			0.9:0.8	0.03:0.05	0.870	0.882	0.031	0.039	0.000	0.001	0.161	97.900
				0.03:0.1	0.870	0.867	0.031	0.039	0.000	0.001	0.145	98.600
		90:10	0.9:0.5	0.03:0.05	0.860	0.890	0.032	0.035	0.001	0.002	0.139	89.600
				0.03:0.1	0.860	0.881	0.033	0.041	0.000	0.002	0.142	92.900
			0.9:0.8	0.03:0.05	0.890	0.898	0.029	0.031	0.000	0.001	0.146	97.500
				0.03:0.1	0.890	0.890	0.028	0.032	0.000	0.001	0.146	98.700
	we_t	50:50	0.9:0.5	0.03:0.05	0.700	0.705	0.041	0.043	0.000	0.002	0.241	99.800
				0.03:0.1	0.700	0.708	0.033	0.033	0.000	0.001	0.193	99.200
			0.9:0.8	0.03:0.05	0.850	0.851	0.029	0.029	0.000	0.001	0.176	99.800
				0.03:0.1	0.850	0.854	0.025	0.024	0.000	0.001	0.147	99.700
		70:30	0.9:0.5	0.03:0.05	0.780	0.781	0.040	0.038	0.000	0.002	0.224	99.000
				0.03:0.1	0.780	0.786	0.031	0.031	0.000	0.001	0.182	99.800
			0.9:0.8	0.03:0.05	0.870	0.871	0.028	0.028	0.000	0.001	0.170	99.600
				0.03:0.1	0.870	0.872	0.026	0.026	0.000	0.001	0.146	99.700
		90:10	0.9:0.5	0.03:0.05	0.860	0.859	0.033	0.032	0.000	0.001	0.192	99.400
				0.03:0.1	0.860	0.862	0.028	0.028	0.000	0.001	0.161	99.500
			0.9:0.8	0.03:0.05	0.890	0.891	0.028	0.027	0.000	0.001	0.157	99.400
				0.03:0.1	0.890	0.891	0.027	0.026	0.000	0.001	0.139	99.300

VE approach: raw: Raw of vaccine efficacy (Equation (12)), we_hal: weighted average vaccine efficacy based on Halloran approach (Equation (17)), we_N: general weighted average vaccine efficacy based on sample size for each sub-population (Equation (7)), we_pres : weighted average vaccine efficacy based on precision approach (Equation (9)), we_t: general weighted average vaccine efficacy based on total follow up time for each sub-population (Equation (19)).

B3.1. Parameter Estimate of Vaccine efficacy for Hazard Ratio Approach with n=2500

n	VE Aproach	Ratio naïve:non	VE naïve:non	IRc naïve:non	True Value	VE	std VE	bias squared	MSE	length CI	coverage CI
2500	raw	50:50	0.9:0.5	0.03:0.05	0.700	0.640	0.101	0.004	0.014	0.410	90.200
				0.03:0.1	0.700	0.596	0.088	0.011	0.019	0.335	71.300
			0.9:0.8	0.03:0.05	0.850	0.838	0.063	0.000	0.004	0.263	94.595
				0.03:0.1	0.850	0.824	0.051	0.001	0.003	0.207	90.900
		70:30	0.9:0.5	0.03:0.05	0.780	0.731	0.090	0.002	0.010	0.365	91.000
				0.03:0.1	0.780	0.662	0.089	0.014	0.022	0.343	62.800
			0.9:0.8	0.03:0.05	0.870	0.857	0.060	0.000	0.004	0.265	95.281
				0.03:0.1	0.870	0.840	0.056	0.001	0.004	0.226	90.000
		90:10	0.9:0.5	0.03:0.05	0.860	0.834	0.070	0.001	0.006	0.306	94.200
				0.03:0.1	0.860	0.792	0.072	0.005	0.010	0.313	79.200
			0.9:0.8	0.03:0.05	0.890	0.882	0.058	0.000	0.003	0.262	96.360
				0.03:0.1	0.890	0.870	0.056	0.000	0.004	0.249	93.065
	we_N	50:50	0.9:0.5	0.03:0.05	0.700	0.675	0.090	0.001	0.009	0.582	98.822
				0.03:0.1	0.700	0.689	0.069	0.000	0.005	0.458	99.400
			0.9:0.8	0.03:0.05	0.850	0.838	0.058	0.000	0.004	0.437	99.647
				0.03:0.1	0.850	0.839	0.052	0.000	0.003	0.368	99.194
		70:30	0.9:0.5	0.03:0.05	0.780	0.763	0.082	0.000	0.007	0.524	99.235
				0.03:0.1	0.780	0.770	0.063	0.000	0.004	0.421	99.148
			0.9:0.8	0.03:0.05	0.870	0.860	0.058	0.000	0.003	0.416	98.919
				0.03:0.1	0.870	0.862	0.053	0.000	0.003	0.349	99.024
		90:10	0.9:0.5	0.03:0.05	0.860	0.840	0.075	0.000	0.006	0.543	99.678
				0.03:0.1	0.860	0.852	0.058	0.000	0.003	0.379	98.857
			0.9:0.8	0.03:0.05	0.890	0.873	0.060	0.000	0.004	0.453	98.680
				0.03:0.1	0.890	0.883	0.053	0.000	0.003	0.336	98.113
	we_pres	50:50	0.9:0.5	0.03:0.05	0.700	0.826	0.118	0.016	0.030	0.576	99.500
				0.03:0.1	0.700	0.795	0.128	0.009	0.026	0.527	99.200
			0.9:0.8	0.03:0.05	0.850	0.881	0.073	0.001	0.006	0.491	99.800
				0.03:0.1	0.850	0.867	0.069	0.000	0.005	0.404	98.800
		70:30	0.9:0.5	0.03:0.05	0.780	0.868	0.084	0.008	0.015	0.443	98.600
				0.03:0.1	0.780	0.837	0.090	0.003	0.011	0.413	99.300
			0.9:0.8	0.03:0.05	0.870	0.891	0.062	0.000	0.004	0.425	99.598
				0.03:0.1	0.870	0.880	0.063	0.000	0.004	0.376	98.900
		90:10	0.9:0.5	0.03:0.05	0.860	0.895	0.062	0.001	0.005	0.376	99.900
				0.03:0.1	0.860	0.881	0.065	0.000	0.005	0.352	99.400
			0.9:0.8	0.03:0.05	0.890	0.925	0.068	0.001	0.006	0.554	99.292
				0.03:0.1	0.890	0.901	0.059	0.000	0.004	0.404	99.497
	we_t	50:50	0.9:0.5	0.03:0.05	0.700	0.676	0.090	0.001	0.009	0.582	98.940
				0.03:0.1	0.700	0.692	0.068	0.000	0.005	0.458	99.520
			0.9:0.8	0.03:0.05	0.850	0.838	0.058	0.000	0.004	0.437	99.647
				0.03:0.1	0.850	0.839	0.052	0.000	0.003	0.369	99.194
		70:30	0.9:0.5	0.03:0.05	0.780	0.764	0.082	0.000	0.007	0.523	99.235
				0.03:0.1	0.780	0.772	0.063	0.000	0.004	0.420	99.148
			0.9:0.8	0.03:0.05	0.870	0.861	0.058	0.000	0.003	0.416	98.919
				0.03:0.1	0.870	0.862	0.053	0.000	0.003	0.350	99.024
		90:10	0.9:0.5	0.03:0.05	0.860	0.841	0.075	0.000	0.006	0.541	99.678
				0.03:0.1	0.860	0.853	0.058	0.000	0.003	0.377	98.857
			0.9:0.8	0.03:0.05	0.890	0.873	0.060	0.000	0.004	0.453	98.680
				0.03:0.1	0.890	0.883	0.053	0.000	0.003	0.335	98.113

VE approach: raw: Raw of vaccine efficacy (Equation (21)), we_N: general weighted average vaccine efficacy based on sample size for each sub-population (Equation (7)), we_pres : weighted average vaccine efficacy based on precision approach (Equation (9)), we_t: general weighted average vaccine efficacy based on total follow up time for each sub-population (Equation (19)).

B3.2. Parameter Estimate of Vaccine efficacy for Hazard Ratio Approach with n=5000

n	VE Aproach	Ratio naïve:non	VE naïve:non	IRc naïve:non	True Value	VE	std VE	bias squared	MSE	length CI	coverage CI
5000	raw	50:50	0.9:0.5	0.03:0.05	0.700	0.655	0.070	0.002	0.007	0.273	88.100
				0.03:0.1	0.700	0.598	0.056	0.010	0.014	0.233	49.800
			0.9:0.8	0.03:0.05	0.850	0.837	0.043	0.000	0.002	0.178	94.400
				0.03:0.1	0.850	0.825	0.035	0.001	0.002	0.142	87.100
		70:30	0.9:0.5	0.03:0.05	0.780	0.733	0.061	0.002	0.006	0.248	85.700
				0.03:0.1	0.780	0.670	0.058	0.012	0.015	0.234	41.400
			0.9:0.8	0.03:0.05	0.870	0.859	0.043	0.000	0.002	0.174	94.200
				0.03:0.1	0.870	0.843	0.038	0.001	0.002	0.153	87.300
		90:10	0.9:0.5	0.03:0.05	0.860	0.838	0.046	0.000	0.003	0.200	92.800
				0.03:0.1	0.860	0.796	0.051	0.004	0.007	0.209	70.400
			0.9:0.8	0.03:0.05	0.890	0.886	0.039	0.000	0.002	0.167	95.800
				0.03:0.1	0.890	0.873	0.040	0.000	0.002	0.163	90.800
	we_N	50:50	0.9:0.5	0.03:0.05	0.700	0.702	0.062	0.000	0.004	0.350	99.486
				0.03:0.1	0.700	0.702	0.046	0.000	0.002	0.283	99.694
			0.9:0.8	0.03:0.05	0.850	0.847	0.040	0.000	0.002	0.261	99.490
				0.03:0.1	0.850	0.849	0.034	0.000	0.001	0.218	99.586
		70:30	0.9:0.5	0.03:0.05	0.780	0.776	0.055	0.000	0.003	0.323	99.599
				0.03:0.1	0.780	0.783	0.044	0.000	0.002	0.262	99.599
			0.9:0.8	0.03:0.05	0.870	0.869	0.041	0.000	0.002	0.248	99.396
				0.03:0.1	0.870	0.869	0.037	0.000	0.001	0.213	99.494
		90:10	0.9:0.5	0.03:0.05	0.860	0.856	0.046	0.000	0.002	0.283	99.298
				0.03:0.1	0.860	0.861	0.041	0.000	0.002	0.230	99.499
			0.9:0.8	0.03:0.05	0.890	0.888	0.039	0.000	0.001	0.238	98.921
				0.03:0.1	0.890	0.889	0.039	0.000	0.002	0.203	98.384
	we_pres	50:50	0.9:0.5	0.03:0.05	0.700	0.830	0.082	0.017	0.024	0.306	74.200
				0.03:0.1	0.700	0.782	0.091	0.007	0.015	0.278	81.200
			0.9:0.8	0.03:0.05	0.850	0.870	0.050	0.000	0.003	0.260	99.700
				0.03:0.1	0.850	0.856	0.048	0.000	0.002	0.225	98.800
		70:30	0.9:0.5	0.03:0.05	0.780	0.865	0.058	0.007	0.011	0.238	80.700
				0.03:0.1	0.780	0.836	0.066	0.003	0.007	0.235	85.200
			0.9:0.8	0.03:0.05	0.870	0.886	0.044	0.000	0.002	0.233	99.300
				0.03:0.1	0.870	0.870	0.044	0.000	0.002	0.216	98.600
		90:10	0.9:0.5	0.03:0.05	0.860	0.892	0.043	0.001	0.003	0.202	95.000
				0.03:0.1	0.860	0.883	0.047	0.001	0.003	0.201	94.800
			0.9:0.8	0.03:0.05	0.890	0.906	0.045	0.000	0.002	0.275	98.800
				0.03:0.1	0.890	0.893	0.042	0.000	0.002	0.218	98.600
	we_t	50:50	0.9:0.5	0.03:0.05	0.700	0.703	0.062	0.000	0.004	0.349	99.486
				0.03:0.1	0.700	0.704	0.046	0.000	0.002	0.282	99.796
			0.9:0.8	0.03:0.05	0.850	0.847	0.040	0.000	0.002	0.261	99.490
				0.03:0.1	0.850	0.849	0.034	0.000	0.001	0.218	99.586
		70:30	0.9:0.5	0.03:0.05	0.780	0.776	0.054	0.000	0.003	0.322	99.599
				0.03:0.1	0.780	0.786	0.044	0.000	0.002	0.261	99.699
			0.9:0.8	0.03:0.05	0.870	0.869	0.041	0.000	0.002	0.247	99.396
				0.03:0.1	0.870	0.870	0.037	0.000	0.001	0.213	99.494
		90:10	0.9:0.5	0.03:0.05	0.860	0.856	0.046	0.000	0.002	0.282	99.298
				0.03:0.1	0.860	0.862	0.041	0.000	0.002	0.229	99.499
			0.9:0.8	0.03:0.05	0.890	0.888	0.039	0.000	0.001	0.238	98.921
				0.03:0.1	0.890	0.889	0.039	0.000	0.002	0.203	98.283

VE approach: raw: Raw of vaccine efficacy (Equation (21)), we_N: general weighted average vaccine efficacy based on sample size for each sub-population (Equation (7)), we_pres : weighted average vaccine efficacy based on precision approach (Equation (9)), we_t: general weighted average vaccine efficacy based on total follow up time for each sub-population (Equation (19)).

B3.3. Parameter Estimate of Vaccine efficacy for Hazard Ratio Approach with n=10000

n	VE Aproach	Ratio naïve:non	VE naïve:non	IRc naïve:non	True Value	VE	std VE	bias squared	MSE	length CI	coverage CI
10000	raw	50:50	0.9:0.5	0.03:0.05	0.700	0.655	0.046	0.002	0.004	0.190	84.300
				0.03:0.1	0.700	0.600	0.042	0.010	0.012	0.163	22.800
			0.9:0.8	0.03:0.05	0.850	0.839	0.030	0.000	0.001	0.122	93.300
				0.03:0.1	0.850	0.827	0.025	0.001	0.001	0.098	82.400
		70:30	0.9:0.5	0.03:0.05	0.780	0.735	0.046	0.002	0.004	0.171	78.800
				0.03:0.1	0.780	0.669	0.041	0.012	0.014	0.163	13.100
			0.9:0.8	0.03:0.05	0.870	0.860	0.029	0.000	0.001	0.120	94.000
				0.03:0.1	0.870	0.844	0.027	0.001	0.001	0.105	80.200
		90:10	0.9:0.5	0.03:0.05	0.860	0.838	0.036	0.000	0.002	0.138	87.600
				0.03:0.1	0.860	0.793	0.037	0.004	0.006	0.146	41.900
			0.9:0.8	0.03:0.05	0.890	0.886	0.029	0.000	0.001	0.114	94.500
				0.03:0.1	0.890	0.874	0.028	0.000	0.001	0.111	89.100
	we_N	50:50	0.9:0.5	0.03:0.05	0.700	0.704	0.042	0.000	0.002	0.233	99.600
				0.03:0.1	0.700	0.706	0.033	0.000	0.001	0.188	99.299
			0.9:0.8	0.03:0.05	0.850	0.851	0.028	0.000	0.001	0.170	99.900
				0.03:0.1	0.850	0.854	0.025	0.000	0.001	0.141	99.800
		70:30	0.9:0.5	0.03:0.05	0.780	0.780	0.040	0.000	0.002	0.215	99.099
				0.03:0.1	0.780	0.783	0.031	0.000	0.001	0.176	99.800
			0.9:0.8	0.03:0.05	0.870	0.871	0.028	0.000	0.001	0.163	99.500
				0.03:0.1	0.870	0.872	0.026	0.000	0.001	0.141	99.700
		90:10	0.9:0.5	0.03:0.05	0.860	0.858	0.033	0.000	0.001	0.182	98.900
				0.03:0.1	0.860	0.861	0.028	0.000	0.001	0.155	99.500
			0.9:0.8	0.03:0.05	0.890	0.891	0.028	0.000	0.001	0.150	99.499
				0.03:0.1	0.890	0.891	0.027	0.000	0.001	0.133	98.900
	we_pres	50:50	0.9:0.5	0.03:0.05	0.700	0.830	0.058	0.017	0.020	0.185	46.000
				0.03:0.1	0.700	0.778	0.063	0.006	0.010	0.176	65.600
			0.9:0.8	0.03:0.05	0.850	0.864	0.035	0.000	0.001	0.164	97.000
				0.03:0.1	0.850	0.851	0.031	0.000	0.001	0.132	96.500
		70:30	0.9:0.5	0.03:0.05	0.780	0.863	0.042	0.007	0.009	0.160	59.800
				0.03:0.1	0.780	0.834	0.049	0.003	0.005	0.157	75.700
			0.9:0.8	0.03:0.05	0.870	0.882	0.031	0.000	0.001	0.154	98.000
				0.03:0.1	0.870	0.867	0.031	0.000	0.001	0.139	97.600
		90:10	0.9:0.5	0.03:0.05	0.860	0.890	0.032	0.001	0.002	0.133	89.400
				0.03:0.1	0.860	0.881	0.033	0.000	0.002	0.136	93.200
			0.9:0.8	0.03:0.05	0.890	0.898	0.029	0.000	0.001	0.143	97.800
				0.03:0.1	0.890	0.890	0.028	0.000	0.001	0.140	98.200
	we_t	50:50	0.9:0.5	0.03:0.05	0.700	0.705	0.041	0.000	0.002	0.233	99.600
				0.03:0.1	0.700	0.709	0.033	0.000	0.001	0.187	99.299
			0.9:0.8	0.03:0.05	0.850	0.851	0.028	0.000	0.001	0.170	99.900
				0.03:0.1	0.850	0.854	0.025	0.000	0.001	0.142	99.800
		70:30	0.9:0.5	0.03:0.05	0.780	0.781	0.040	0.000	0.002	0.214	99.099
				0.03:0.1	0.780	0.786	0.031	0.000	0.001	0.175	99.700
			0.9:0.8	0.03:0.05	0.870	0.871	0.028	0.000	0.001	0.163	99.400
				0.03:0.1	0.870	0.872	0.026	0.000	0.001	0.141	99.700
		90:10	0.9:0.5	0.03:0.05	0.860	0.859	0.033	0.000	0.001	0.181	98.900
				0.03:0.1	0.860	0.862	0.028	0.000	0.001	0.154	99.400
			0.9:0.8	0.03:0.05	0.890	0.891	0.028	0.000	0.001	0.150	99.499
				0.03:0.1	0.890	0.891	0.027	0.000	0.001	0.133	98.900

VE approach: raw: Raw of vaccine efficacy (Equation (21)), we_N: general weighted average vaccine efficacy based on sample size for each sub-population (Equation (7)), we_pres : weighted average vaccine efficacy based on precision approach (Equation (9)), we_t: general weighted average vaccine efficacy based on total follow up time for each sub-population (Equation (19)).

B4.1. Parameter Estimate of Vaccine efficacy for Poisson Regression Approach with n=2500

n	VE Aproach	Ratio naïve:non	VE naïve:non	IRc naïve:non	True Value	VE	std VE	bias squared	MSE	length CI	coverage CI
2500	raw	50:50	0.9:0.5	0.03:0.05	0.700	0.640	0.101	0.004	0.014	0.194	57.800
				0.03:0.1	0.700	0.596	0.088	0.011	0.019	0.199	43.700
			0.9:0.8	0.03:0.05	0.850	0.838	0.063	0.000	0.004	0.111	59.459
				0.03:0.1	0.850	0.824	0.051	0.001	0.003	0.110	65.200
		70:30	0.9:0.5	0.03:0.05	0.780	0.731	0.090	0.002	0.010	0.158	55.400
				0.03:0.1	0.780	0.662	0.089	0.014	0.022	0.180	29.500
			0.9:0.8	0.03:0.05	0.870	0.857	0.060	0.000	0.004	0.104	62.651
				0.03:0.1	0.870	0.840	0.056	0.001	0.004	0.107	57.800
		90:10	0.9:0.5	0.03:0.05	0.860	0.834	0.070	0.001	0.006	0.115	55.800
				0.03:0.1	0.860	0.792	0.072	0.005	0.010	0.132	39.800
			0.9:0.8	0.03:0.05	0.890	0.882	0.058	0.000	0.003	0.093	55.308
				0.03:0.1	0.890	0.870	0.056	0.000	0.004	0.097	58.593
	we_N	50:50	0.9:0.5	0.03:0.05	0.700	0.675	0.090	0.001	0.009	0.254	82.450
				0.03:0.1	0.700	0.689	0.069	0.000	0.005	0.228	90.048
			0.9:0.8	0.03:0.05	0.850	0.838	0.058	0.000	0.004	0.159	81.647
				0.03:0.1	0.850	0.839	0.052	0.000	0.003	0.150	87.327
		70:30	0.9:0.5	0.03:0.05	0.780	0.763	0.082	0.000	0.007	0.222	82.842
				0.03:0.1	0.780	0.770	0.063	0.000	0.004	0.203	89.137
			0.9:0.8	0.03:0.05	0.870	0.860	0.058	0.000	0.003	0.148	80.973
				0.03:0.1	0.870	0.862	0.053	0.000	0.003	0.140	80.586
		90:10	0.9:0.5	0.03:0.05	0.860	0.840	0.075	0.000	0.006	0.197	80.344
				0.03:0.1	0.860	0.852	0.058	0.000	0.003	0.165	83.576
			0.9:0.8	0.03:0.05	0.890	0.873	0.060	0.000	0.004	0.149	75.220
				0.03:0.1	0.890	0.883	0.053	0.000	0.003	0.127	76.804
	we_pres	50:50	0.9:0.5	0.03:0.05	0.700	0.853	0.105	0.023	0.034	0.293	46.600
				0.03:0.1	0.700	0.853	0.106	0.023	0.035	0.301	47.700
			0.9:0.8	0.03:0.05	0.850	0.888	0.071	0.001	0.006	0.269	76.076
				0.03:0.1	0.850	0.882	0.068	0.001	0.006	0.250	79.900
		70:30	0.9:0.5	0.03:0.05	0.780	0.882	0.077	0.010	0.016	0.201	42.800
				0.03:0.1	0.780	0.875	0.075	0.009	0.015	0.180	43.700
			0.9:0.8	0.03:0.05	0.870	0.895	0.061	0.001	0.004	0.186	72.088
				0.03:0.1	0.870	0.891	0.063	0.000	0.004	0.192	71.400
		90:10	0.9:0.5	0.03:0.05	0.860	0.898	0.061	0.001	0.005	0.164	59.500
				0.03:0.1	0.860	0.893	0.061	0.001	0.005	0.139	57.100
			0.9:0.8	0.03:0.05	0.890	0.926	0.068	0.001	0.006	0.383	73.711
				0.03:0.1	0.890	0.905	0.059	0.000	0.004	0.194	71.357
	we_t	50:50	0.9:0.5	0.03:0.05	0.700	0.676	0.090	0.001	0.009	0.253	82.568
				0.03:0.1	0.700	0.692	0.068	0.000	0.005	0.227	90.408
			0.9:0.8	0.03:0.05	0.850	0.838	0.058	0.000	0.004	0.159	81.765
				0.03:0.1	0.850	0.839	0.052	0.000	0.003	0.150	87.327
		70:30	0.9:0.5	0.03:0.05	0.780	0.764	0.082	0.000	0.007	0.221	83.169
				0.03:0.1	0.780	0.772	0.063	0.000	0.004	0.201	89.031
			0.9:0.8	0.03:0.05	0.870	0.861	0.058	0.000	0.003	0.148	80.973
				0.03:0.1	0.870	0.862	0.053	0.000	0.003	0.139	80.260
		90:10	0.9:0.5	0.03:0.05	0.860	0.841	0.075	0.000	0.006	0.197	80.559
				0.03:0.1	0.860	0.853	0.058	0.000	0.003	0.163	83.680
			0.9:0.8	0.03:0.05	0.890	0.873	0.060	0.000	0.004	0.148	75.220
				0.03:0.1	0.890	0.883	0.053	0.000	0.003	0.126	76.471

VE approach: raw: Raw of vaccine efficacy (Equation (23)), we_N: general weighted average vaccine efficacy based on sample size for each sub-population (Equation (7)), we_pres : weighted average vaccine efficacy based on precision approach (Equation (9)), we_t: general weighted average vaccine efficacy based on total follow up time for each sub-population (Equation (19)).

B4.2. Parameter Estimate of Vaccine efficacy for Poisson Regression Approach with n=5000

n	VE Aproach	Ratio naïve:non	VE naïve:non	IRc naïve:non	True Value	VE	std VE	bias squared	MSE	length CI	coverage CI
5000	raw	50:50	0.9:0.5	0.03:0.05	0.700	0.655	0.069	0.002	0.007	0.132	56.500
				0.03:0.1	0.700	0.598	0.056	0.010	0.014	0.140	21.600
			0.9:0.8	0.03:0.05	0.850	0.837	0.043	0.000	0.002	0.078	62.100
				0.03:0.1	0.850	0.825	0.035	0.001	0.002	0.077	61.300
		70:30	0.9:0.5	0.03:0.05	0.780	0.733	0.061	0.002	0.006	0.111	48.800
				0.03:0.1	0.780	0.670	0.058	0.012	0.015	0.124	13.900
			0.9:0.8	0.03:0.05	0.870	0.859	0.043	0.000	0.002	0.072	58.800
				0.03:0.1	0.870	0.843	0.038	0.001	0.002	0.074	54.700
		90:10	0.9:0.5	0.03:0.05	0.860	0.838	0.046	0.000	0.003	0.080	56.400
				0.03:0.1	0.860	0.796	0.051	0.004	0.007	0.091	30.000
			0.9:0.8	0.03:0.05	0.890	0.886	0.039	0.000	0.002	0.064	56.800
				0.03:0.1	0.890	0.873	0.040	0.000	0.002	0.067	55.200
	we_N	50:50	0.9:0.5	0.03:0.05	0.700	0.702	0.062	0.000	0.004	0.164	80.987
				0.03:0.1	0.700	0.702	0.046	0.000	0.002	0.153	90.418
			0.9:0.8	0.03:0.05	0.850	0.847	0.040	0.000	0.002	0.106	80.224
				0.03:0.1	0.850	0.849	0.034	0.000	0.001	0.099	85.936
		70:30	0.9:0.5	0.03:0.05	0.780	0.776	0.055	0.000	0.003	0.148	82.565
				0.03:0.1	0.780	0.783	0.044	0.000	0.002	0.135	87.174
			0.9:0.8	0.03:0.05	0.870	0.869	0.041	0.000	0.002	0.099	76.536
				0.03:0.1	0.870	0.869	0.037	0.000	0.001	0.094	79.757
		90:10	0.9:0.5	0.03:0.05	0.860	0.856	0.046	0.000	0.002	0.119	80.441
				0.03:0.1	0.860	0.861	0.041	0.000	0.002	0.108	80.881
			0.9:0.8	0.03:0.05	0.890	0.888	0.039	0.000	0.001	0.088	71.629
				0.03:0.1	0.890	0.889	0.039	0.000	0.002	0.083	70.909
	we_pres	50:50	0.9:0.5	0.03:0.05	0.700	0.862	0.072	0.026	0.031	0.126	16.400
				0.03:0.1	0.700	0.851	0.076	0.023	0.029	0.119	17.800
			0.9:0.8	0.03:0.05	0.850	0.879	0.050	0.001	0.003	0.111	62.800
				0.03:0.1	0.850	0.875	0.050	0.001	0.003	0.122	69.400
		70:30	0.9:0.5	0.03:0.05	0.780	0.881	0.053	0.010	0.013	0.085	20.700
				0.03:0.1	0.780	0.878	0.054	0.010	0.013	0.085	21.400
			0.9:0.8	0.03:0.05	0.870	0.892	0.044	0.000	0.002	0.089	62.100
				0.03:0.1	0.870	0.886	0.045	0.000	0.002	0.096	66.100
		90:10	0.9:0.5	0.03:0.05	0.860	0.896	0.042	0.001	0.003	0.072	48.100
				0.03:0.1	0.860	0.896	0.043	0.001	0.003	0.070	45.800
			0.9:0.8	0.03:0.05	0.890	0.907	0.045	0.000	0.002	0.141	65.300
				0.03:0.1	0.890	0.898	0.042	0.000	0.002	0.084	62.100
	we_t	50:50	0.9:0.5	0.03:0.05	0.700	0.703	0.062	0.000	0.004	0.164	80.987
				0.03:0.1	0.700	0.704	0.046	0.000	0.002	0.152	90.622
			0.9:0.8	0.03:0.05	0.850	0.847	0.040	0.000	0.002	0.106	80.428
				0.03:0.1	0.850	0.849	0.034	0.000	0.001	0.099	85.936
		70:30	0.9:0.5	0.03:0.05	0.780	0.776	0.054	0.000	0.003	0.147	82.565
				0.03:0.1	0.780	0.786	0.044	0.000	0.002	0.134	86.373
			0.9:0.8	0.03:0.05	0.870	0.869	0.041	0.000	0.002	0.099	76.435
				0.03:0.1	0.870	0.870	0.037	0.000	0.001	0.094	79.453
		90:10	0.9:0.5	0.03:0.05	0.860	0.857	0.046	0.000	0.002	0.118	80.542
				0.03:0.1	0.860	0.862	0.041	0.000	0.002	0.107	80.380
			0.9:0.8	0.03:0.05	0.890	0.889	0.039	0.000	0.001	0.088	71.521
				0.03:0.1	0.890	0.889	0.039	0.000	0.002	0.083	71.111

VE approach: raw: Raw of vaccine efficacy (Equation (23)), we_N: general weighted average vaccine efficacy based on sample size for each sub-population (Equation (7)), we_pres : weighted average vaccine efficacy based on precision approach (Equation (9)), we_t: general weighted average vaccine efficacy based on total follow up time for each sub-population (Equation (19)).

B4.3. Parameter Estimate of Vaccine efficacy for Poisson Regression Approach with n=10000

n	VE Aproach	Ratio naïve:non	VE naïve:non	IRc naïve:non	True Value	VE	std VE	bias squared	MSE	length CI	coverage CI
10000	raw	50:50	0.9:0.5	0.03:0.05	0.700	0.655	0.046	0.002	0.004	0.093	48.800
				0.03:0.1	0.700	0.600	0.042	0.010	0.012	0.098	7.000
			0.9:0.8	0.03:0.05	0.850	0.839	0.030	0.000	0.001	0.055	60.200
				0.03:0.1	0.850	0.827	0.025	0.001	0.001	0.054	53.800
		70:30	0.9:0.5	0.03:0.05	0.780	0.735	0.046	0.002	0.004	0.078	38.300
				0.03:0.1	0.780	0.669	0.041	0.012	0.014	0.088	2.100
			0.9:0.8	0.03:0.05	0.870	0.860	0.029	0.000	0.001	0.051	59.700
				0.03:0.1	0.870	0.844	0.027	0.001	0.001	0.052	43.600
		90:10	0.9:0.5	0.03:0.05	0.860	0.838	0.036	0.000	0.002	0.056	48.300
				0.03:0.1	0.860	0.793	0.037	0.004	0.006	0.065	12.100
			0.9:0.8	0.03:0.05	0.890	0.886	0.029	0.000	0.001	0.045	56.500
				0.03:0.1	0.890	0.874	0.028	0.000	0.001	0.047	53.900
	we_N	50:50	0.9:0.5	0.03:0.05	0.700	0.704	0.042	0.000	0.002	0.115	83.300
				0.03:0.1	0.700	0.706	0.033	0.000	0.001	0.107	89.089
			0.9:0.8	0.03:0.05	0.850	0.851	0.029	0.000	0.001	0.073	78.679
				0.03:0.1	0.850	0.853	0.025	0.000	0.001	0.068	82.500
		70:30	0.9:0.5	0.03:0.05	0.780	0.780	0.040	0.000	0.002	0.102	80.981
				0.03:0.1	0.780	0.783	0.031	0.000	0.001	0.095	86.400
			0.9:0.8	0.03:0.05	0.870	0.871	0.028	0.000	0.001	0.069	76.900
				0.03:0.1	0.870	0.871	0.026	0.000	0.001	0.065	77.300
		90:10	0.9:0.5	0.03:0.05	0.860	0.858	0.033	0.000	0.001	0.081	78.600
				0.03:0.1	0.860	0.861	0.028	0.000	0.001	0.075	82.000
			0.9:0.8	0.03:0.05	0.890	0.891	0.028	0.000	0.001	0.060	70.341
				0.03:0.1	0.890	0.891	0.027	0.000	0.001	0.058	71.400
	we_pres	50:50	0.9:0.5	0.03:0.05	0.700	0.863	0.051	0.027	0.029	0.070	2.300
				0.03:0.1	0.700	0.852	0.052	0.023	0.026	0.072	4.700
			0.9:0.8	0.03:0.05	0.850	0.874	0.036	0.001	0.002	0.067	60.100
				0.03:0.1	0.850	0.872	0.036	0.000	0.002	0.064	59.600
		70:30	0.9:0.5	0.03:0.05	0.780	0.881	0.037	0.010	0.012	0.060	6.300
				0.03:0.1	0.780	0.878	0.040	0.010	0.011	0.059	9.200
			0.9:0.8	0.03:0.05	0.870	0.888	0.031	0.000	0.001	0.059	60.600
				0.03:0.1	0.870	0.884	0.032	0.000	0.001	0.060	62.500
		90:10	0.9:0.5	0.03:0.05	0.860	0.895	0.031	0.001	0.002	0.049	36.100
				0.03:0.1	0.860	0.895	0.030	0.001	0.002	0.049	36.300
			0.9:0.8	0.03:0.05	0.890	0.900	0.029	0.000	0.001	0.053	60.300
				0.03:0.1	0.890	0.897	0.028	0.000	0.001	0.052	63.100
	we_t	50:50	0.9:0.5	0.03:0.05	0.700	0.705	0.041	0.000	0.002	0.114	83.600
				0.03:0.1	0.700	0.708	0.033	0.000	0.001	0.106	88.388
			0.9:0.8	0.03:0.05	0.850	0.851	0.028	0.000	0.001	0.073	78.879
				0.03:0.1	0.850	0.854	0.025	0.000	0.001	0.068	82.500
		70:30	0.9:0.5	0.03:0.05	0.780	0.781	0.040	0.000	0.002	0.102	81.481
				0.03:0.1	0.780	0.786	0.031	0.000	0.001	0.094	85.200
			0.9:0.8	0.03:0.05	0.870	0.871	0.028	0.000	0.001	0.069	76.900
				0.03:0.1	0.870	0.872	0.026	0.000	0.001	0.065	77.200
		90:10	0.9:0.5	0.03:0.05	0.860	0.859	0.033	0.000	0.001	0.080	79.100
				0.03:0.1	0.860	0.862	0.028	0.000	0.001	0.075	81.800
			0.9:0.8	0.03:0.05	0.890	0.891	0.028	0.000	0.001	0.060	70.140
				0.03:0.1	0.890	0.891	0.027	0.000	0.001	0.057	71.100

VE approach: raw: Raw of vaccine efficacy (Equation (23)), we_N: general weighted average vaccine efficacy based on sample size for each sub-population (Equation (7)), we_pres : weighted average vaccine efficacy based on precision approach (Equation (9)), we_t: general weighted average vaccine efficacy based on total follow up time for each sub-population (Equation (19)).

C. SAS Codes

C1. Macro for running the simulation

```
*note :
n : total sample size in the study
p_naive : prop of being naive pop
p_c01: prop incidence rate (IR) for control group on naive
sub-pop
p_c00: prop incidence rate (IR) for control group on non-
naive sub-pop
ve11: vaccine efficacy for vaccine group on naive sub-pop
ve10: vaccine efficacy for vaccine group on non-naive sub-
pop
;
%macro generate;
%let n=5000;
%let p_naive=0.9;
%let p_c01=0.03;
%let p_c00=0.1;
%let ve11=0.9;
%let ve10=0.65;
%let m=100;

%do i=1 %to &m;
dm "out;clear;log;clear;";
%simulate (&n,&p_naive, &p_c01,&p_c00,&ve11,&ve10);
%sim_AR; %sim_FU; %sim_hazard; %sim_poisson;
%end;
%compile_AR; %compile_FU;
%compile_hazard;%compile_poisson;

data compilation;
set final_ar final_fu final_HR final_poisson;run;
proc print data=compilation;run;

%mend generate;
```

C2. Macro Generating the Data

```
%macro simulate (n, p_naive, p_c01,p_c00,ve11,ve10) ;
 data sim;
  do j = 1 to &n ;
  group=rand('bern',0.5);
 if group=0 then do ;naive=rand('bern',&p_naive);
              if naive =1 then case=rand('bern',&p_c01) ;
              else case=rand('bern',&p_c00) ;end;
else do;
              naive=rand('bern',&p_naive);
if naive =1 then case=rand('bern',(1-&ve11)*&p_c01) ;
else case=rand('bern',(1-&ve10)*&p_c00) ;end;

x=ranexp(12345)/0.34384;
if case=1 then do;censored = 0;time = min(4,x);end;
              else do;censored = 1; time = 4;end;
output; end;
run;
%mend simulate;
```

C3. Macro VE approach

C3.1 Attack Rate (AR)

```
%macro sim_AR;

*### RAW VE_AR ####;
proc sort data=sim; by group ;run;
proc means data=sim noprint;
output out=out2 sum(case)=sum_case sum(time)=sum_time;
by group;run;

data raw_ar;set out2;
AR=sum_case/_freq_;drop _type_;run;

data raw_group0 (where=(group=0) rename=(_freq_=n0
sum_case=case0 ar=ar0))
    raw_group1 (where=(group=1) rename=(_freq_=n1
sum_case=case1 ar=ar1));
set raw_ar;run;

data r;
merge raw_group0  raw_group1;
ve=1-( log(1-ar1)/log(1-ar0) );
var_VE=( (case1/(n1*(n1-case1))) + ((1-ve)*(1-
ve)*case0/(n0*(n0-case0))) )/((log(n0/(n0-case0)))**2);
se_ve=sqrt(var_ve);
**********calculating CI ***********;
eta=log(ve/(1-ve)); * using transformation of logit
FUNCTION OF ve;
var_eta=(1/(ve*(1-ve)))*(1/(ve*(1-ve)))*var_VE;*use delta
methods;
ll_eta=eta-(1.96*sqrt(var_eta));
ul_eta=eta+(1.96*sqrt(var_eta));
ve=exp(eta)/(1+exp(eta));* using transformation of expit
FUNCTION OF ve;
ll_ve=exp(ll_eta)/(1+exp(ll_eta));
ul_ve=exp(ul_eta)/(1+exp(ul_eta));
lenght=ul_ve-ll_ve;
*define true value (tp);
tp=(&p_naive*&ve11)+ ((1-&p_naive)*&ve10);*define true
parameter of VE general weighted;
if ll_ve<=tp and ul_ve >=tp then covtot=1; else covtot=0;
*define the coverage;
keep ve se_ve ll_ve ul_ve tp lenght covtot; run;

%if &i = 1 %then %do;
     data outall_ar_raw;
       set r;
     %end;
   %else %do;
     proc append base=outall_ar_raw data=r;quit;
     %end;

*######## ar halloran ##########;
proc sort data=sim; by naive group ;run;
proc means data=sim noprint;
output out=out1 sum(case)=sum_case sum(time)=sum_time;
by naive group ;run;

data hal_ar_group0 (where=(group=0))
    hal_ar_group1 (where=(group=1));
set out1;
AR=sum_case/_freq_;drop _type_;run;
```

```
data hal_group0na0_ar (where=(naive=0)
rename=(_freq_=n0 sum_case=case0 sum_time=t0 ar=ar0))
            hal_group0na1_ar (where=(naive=1)
rename=(_freq_=n1 sum_case=case1 sum_time=t1 ar=ar1));
set hal_ar_group0; run;

data hal_group1na0_ar (where=(naive=0)
rename=(_freq_=n0 sum_case=case0 sum_time=t0 ar=ar0))
            hal_group1na1_ar (where=(naive=1)
rename=(_freq_=n1 sum_case=case1 sum_time=t1 ar=ar1));
set hal_ar_group1; run;

data hal_gr0_ar;merge hal_group0na0_ar  hal_group0na1_ar
;
we_na0=n0/(n0+n1);
we_na1=n1/(n0+n1);
den=we_na0*log(1-ar0)+we_na1*log(1-ar1);
var_ar0=ar0*(1-ar0)/n0;
var_ar1=ar1*(1-ar1)/n1;
var_den=(we_na0*we_na0*(1/(1-ar0))*(1/(1-
ar0))*var_ar0)+(we_na1*we_na1*(1/(1-ar1))*(1/(1-
ar1))*var_ar1);
keep den var_den;run;

data hal_gr1_ar;merge hal_group1na0_ar  hal_group1na1_ar
;
we_na0=n0/(n0+n1);
we_na1=n1/(n0+n1);
num=we_na0*log(1-ar0)+we_na1*log(1-ar1);
var_ar0=ar0*(1-ar0)/n0;
var_ar1=ar1*(1-ar1)/n1;
var_num=((we_na0*we_na0)*(1/(1-ar0))*(1/(1-
ar0))*var_ar0)+((we_na1*we_na1)*(1/(1-ar1))*(1/(1-
ar1))*var_ar1);
keep num var_num;run;

data result_hal_ar;merge hal_gr0_ar hal_gr1_ar;
rr=num/den;
VE=1-rr;
var_rr=(((rr*rr)/(num*num))*var_num)+(((rr*rr)/(den*den))
*var_den);
var_ve=var_rr;
se_ve=sqrt(var_ve);
**********calculating CI ***********;
eta=log(ve/(1-ve)); * using transformation of logit
FUNCTION OF ve;
var_eta=(1/(ve*(1-ve)))*(1/(ve*(1-ve)))*var_VE;*use delta
methods;
ll_eta=eta-(1.96*sqrt(var_eta));
ul_eta=eta+(1.96*sqrt(var_eta));
ve=exp(eta)/(1+exp(eta));* using transformation of expit
FUNCTION OF ve;
ll_ve=exp(ll_eta)/(1+exp(ll_eta));
ul_ve=exp(ul_eta)/(1+exp(ul_eta));
*define true parameter of VE hal;
numtp=(&p_naive*log(1-((1-&ve11)*&p_c01)))+((1-
&p_naive)*log(1-((1-&ve10)*&p_c00)));
dentp=(&p_naive*log(1-&p_c01))+((1-&p_naive)*log(1-
&p_c00));
tp=1- numtp/dentp ;
lenght=ul_ve-ll_ve;
if ll_ve<=tp and ul_ve >=tp then covtot=1; else covtot=0;
*define the coverage;
keep ve se_ve ll_ve ul_ve tp lenght covtot; run;

%if &i = 1 %then %do;
    data outall_ar_hal;
    set result_hal_ar;
  %end;
 %else %do;
  proc append base=outall_ar_hal data=result_hal_ar;quit;
  %end;

*########## ar general weighted############;
data g_naive0_ar (where=(naive=0))
   g_naive1_ar (where=(naive=1));
set out1;
AR=sum_case/_freq_;drop _type_;run;

data g_group0na0 (where=(group=0) rename=(_freq_=n0
sum_case=case0 sum_time=t0 ar=ar0))
            g_group1na0 (where=(group=1)
rename=(_freq_=n1 sum_case=case1 sum_time=t1 ar=ar1));
set  g_naive0_ar; run;

data g_group0na1 (where=(group=0) rename=(_freq_=N0
sum_case=case0 sum_time=t0 ar=ar0))
            g_group1na1 (where=(group=1)
rename=(_freq_=N1 sum_case=case1 sum_time=t1 ar=ar1));
set g_naive1_ar; run;

data g_na0_ar;merge g_group0na0  g_group1na0 ;
VE_na0_ar=1-(log(1-ar1)/log(1-ar0));
var_VE_na0_ar=((case1/(n1*(n1-case1)))+((1-
VE_na0_ar)*(1-VE_na0_ar)*case0/(n0*(n0-
case0))))/((log(n0/(n0-case0)))**2);
n_na0_ar=n0+n1;
keep VE_na0_ar var_VE_na0_ar n_na0_ar ;run;

data g_na1_ar;merge g_group0na1  g_group1na1 ;
VE_na1_ar=1-(log(1-ar1)/log(1-ar0));
var_VE_na1_ar=((case1/(n1*(n1-case1)))+((1-
VE_na1_ar)*(1-VE_na1_ar)*case0/(n0*(n0-
case0))))/((log(n0/(n0-case0)))**2);
n_na1_ar=n0+n1;
keep VE_na1_ar var_VE_na1_ar n_na1_ar ;run;

***weighted by N;
data result_g_ar;merge g_na0_ar g_na1_ar;
we_na0=n_na0_ar/(n_na0_ar+n_na1_ar);
we_na1=n_na1_ar/(n_na0_ar+n_na1_ar);
ve=we_na0*VE_na0_ar+we_na1*VE_na1_ar;
var_ve=((we_na0*we_na0)*var_VE_na0_ar)
+((we_na1*we_na1)*var_VE_na1_ar);
se_ve=sqrt(var_ve);
**********calculating CI ***********;
eta=log(ve/(1-ve)); * using transformation of logit
FUNCTION OF ve;
var_eta=(1/(ve*(1-ve)))*(1/(ve*(1-ve)))*var_VE;*use delta
methods;
ll_eta=eta-(1.96*sqrt(var_eta));
ul_eta=eta+(1.96*sqrt(var_eta));
ve=exp(eta)/(1+exp(eta));* using transformation of expit
FUNCTION OF ve;
ll_ve=exp(ll_eta)/(1+exp(ll_eta));
ul_ve=exp(ul_eta)/(1+exp(ul_eta));
tp=(&p_naive*&ve11)+ ((1-&p_naive)*&ve10);*define true
parameter of VE general weighted;
lenght=ul_ve-ll_ve;
if ll_ve<=tp and ul_ve >=tp then covtot=1; else covtot=0;
*define the coverage;
keep ve se_ve ll_ve ul_ve lenght tp covtot; run;

%if &i = 1 %then %do;
```

```
    data outall_ar_g;
      set result_g_ar;
    %end;
  %else %do;
    proc append base=outall_ar_g data=result_g_ar;quit;
    %end;

*############ar by naive###############;
data g_na1;set g_na1_ar;name='naive';
ve=ve_na1_ar;
se_ve=sqrt(var_ve_na1_ar);
*********calculating CI ***********;
eta=log(ve/(1-ve)); * using transformation of logit
FUNCTION OF ve;
var_eta=(1/(ve*(1-ve)))*(1/(ve*(1-ve)))*se_ve*se_ve;*use
delta methods;
ll_eta=eta-(1.96*sqrt(var_eta));
ul_eta=eta+(1.96*sqrt(var_eta));
ve=exp(eta)/(1+exp(eta));* using transformation of expit
FUNCTION OF ve;
ll_ve=exp(ll_eta)/(1+exp(ll_eta));
ul_ve=exp(ul_eta)/(1+exp(ul_eta));
tp=&ve11;*define true parameter of VE naive;
lenght=ul_ve-ll_ve;
if ll_ve<=tp and ul_ve >=tp then covtot=1; else covtot=0;
*define the coverage;
keep name ve se_ve ll_ve ul_ve tp lenght covtot; run;

data g_na0;set g_na0_ar;name='non-n';
ve=ve_na0_ar;
se_ve=sqrt(var_ve_na0_ar);
*********calculating CI ***********;
eta=log(ve/(1-ve)); * using transformation of logit
FUNCTION OF ve;
var_eta=(1/(ve*(1-ve)))*(1/(ve*(1-ve)))*se_ve*se_ve;*use
delta methods;
ll_eta=eta-(1.96*sqrt(var_eta));
ul_eta=eta+(1.96*sqrt(var_eta));
ve=exp(eta)/(1+exp(eta));* using transformation of expit
FUNCTION OF ve;
ll_ve=exp(ll_eta)/(1+exp(ll_eta));
ul_ve=exp(ul_eta)/(1+exp(ul_eta));
tp=&ve10;*define true parameter of VE non-naive;
lenght=ul_ve-ll_ve;
if ll_ve<=tp and ul_ve >=tp then covtot=1; else covtot=0;
*define the coverage;
keep name ve se_ve ll_ve ul_ve tp lenght covtot; run;

data result_bynaive_ar;set g_na1 g_na0;run;

%if &i = 1 %then %do;
    data outall_ar_bynaive;
      set result_bynaive_ar;
    %end;
  %else %do;
    proc append base=outall_ar_bynaive
data=result_bynaive_ar;quit;
    %end;

*#########precision based approach#############;
*non-naive*;
data hbr_na0_ar;merge g_group0na0  g_group1na0 ;
VE_na0_ar=1-(log(1-ar1)/log(1-ar0));
var_VE_na0_ar=(   ((case1)/(n1*(n1-case1)))+ ((1-
VE_na0_ar)*(1-VE_na0_ar)*(case0)/(n0*(n0-case0)))
)/((log(n0/(n0-case0)))**2);
we_na0=1/var_VE_na0_ar;
*********calculating the variance of we_na0 ***********;
Num=(log(1-ar0))**2;
Var_num=4*num*ar0/(n0*(1-ar0));
Den_a=ar1/(n1*(1-ar1));
Var_den_a=(den_a**2)*(1/((case1)*(1-ar1)));
Den_b=((1- VE_na0_ar)**2)*ar0/(n0*(1-ar0));
var_log_den_b=( 4*((1/(1- VE_na0_ar))**2)*
var_VE_na0_ar )+ 1/((case0)*(1-ar0))+  2*sqrt( ( 4*((1/(1-
VE_na0_ar))**2)* var_VE_na0_ar )* (1/((case0)*(1-ar0)))
);
var_den_b= (Den_b**2)* var_log_den_b;
den= den_a+den_b;
var_den= Var_den_a+ var_den_b+2*sqrt(Var_den_a*
var_den_b);
var_log_we_na0=((1/num)**2)* Var_num + ((1/den)**2)*
Var_den-2*sqrt( ((1/(num*den))**2) *var_num*var_den );
var_we_na0=(we_na0**2)* var_log_we_na0;
wexVE_na0= we_na0*VE_na0_ar;
var_log_wexVE_na0= var_log_we_na0 + ((1/
VE_na0_ar)**2)* var_VE_na0_ar
+2*sqrt(var_log_we_na0*((1/ VE_na0_ar)**2)*
var_VE_na0_ar);
var_wexVE_na0= (wexVE_na0**2)* var_log_wexVE_na0;
keep wexVE_na0 var_wexVE_na0 var_we_na0 we_na0;
run;

*naive*;
data hbr_na1_ar;merge g_group0na1  g_group1na1 ;
VE_na1_ar=1-(log(1-ar1)/log(1-ar0));
var_VE_na1_ar=(((case1)/(n1*(n1-case1)))+((1-
VE_na1_ar)*(1-VE_na1_ar)*(case0)/(n0*(n0-
case0))))/((log(n0/(n0-case0)))**2);
we_na1=1/var_VE_na1_ar;
*********calculating the variance of we_na0 ***********;
Num=(log(1-ar0))**2;
Var_num=4*num*ar0/(n0*(1-ar0));
Den_a=ar1/(n1*(1-ar1));
Var_den_a=(den_a**2)*(1/((case1)*(1-ar1)));
Den_b=((1- VE_na1_ar)**2)*ar0/(n0*(1-ar0));
var_log_den_b=( 4*((1/(1- VE_na1_ar))**2)*
var_VE_na1_ar )+ 1/((case0)*(1-ar0))+ 2*sqrt( ( 4*((1/(1-
VE_na1_ar))**2)* var_VE_na1_ar )* (1/((case0)*(1-ar0)))
);
var_den_b= (Den_b**2)* var_log_den_b;
den= den_a+den_b;
var_den= Var_den_a+ var_den_b+2*sqrt(Var_den_a*
var_den_b);
var_log_we_na1=((1/num)**2)* Var_num + ((1/den)**2)*
Var_den-2*sqrt(  ((1/(num*den))**2)*var_num*var_den);
var_we_na1=(we_na1**2)* var_log_we_na1;
wexVE_na1= we_na1*VE_na1_ar;
var_log_wexVE_na1= var_log_we_na1 + ((1/
VE_na1_ar)**2)* var_VE_na1_ar
+2*sqrt(var_log_we_na1*((1/ VE_na1_ar)**2)*
var_VE_na1_ar);
var_wexVE_na1= (wexVE_na1**2)* var_log_wexVE_na1;
keep wexVE_na1 var_wexVE_na1 var_we_na1 we_na1
;run;

data result_hbr_ar;merge hbr_na0_ar hbr_na1_ar;
num= wexVE_na0 + wexVE_na1;
den= we_na0+we_na1;
Var_num= var_wexVE_na1+ var_wexVE_na0;
Var_den= var_we_na1+ var_we_na0;
VE=num/den;
Var_ve_log=((1/num)**2)*var_num+ ((1/den)**2)*var_den-
2*sqrt( ((1/num)**2)*var_num * ((1/den)**2)*var_den );
```

```
Var_ve=(ve**2)* Var_ve_log;
se_ve=sqrt(var_ve);
*********calculating CI **********;
eta=log(ve/(1-ve)); * using transformation of logit
FUNCTION OF ve;
var_eta=(1/(ve*(1-ve)))*(1/(ve*(1-ve)))*var_VE;*use delta
methods;
ll_eta=eta-(1.96*sqrt(var_eta));
ul_eta=eta+(1.96*sqrt(var_eta));
ve=exp(eta)/(1+exp(eta));* using transformation of expit
FUNCTION OF ve;
ll_ve=exp(ll_eta)/(1+exp(ll_eta));
ul_ve=exp(ul_eta)/(1+exp(ul_eta));
*define true parameter of VE precision;
tp=(&p_naive*&ve11)+ ((1-&p_naive)*&ve10);
lenght=ul_ve-ll_ve;
if ll_ve<=tp and ul_ve >=tp then covtot=1; else covtot=0;
*define the coverage;
keep ve se_ve ll_ve ul_ve lenght tp covtot; run;

%if &i = 1 %then %do;
     data outall_ar_hbr;
       set result_hbr_ar;
     %end;
   %else %do;
     proc append base=outall_ar_hbr data=result_hbr_ar;quit;
     %end;

%mend sim_AR;
```

C3.2 Person Time Approach (FU)

```
%macro sim_FU;

*###### RAW VE_FU #############;
proc sort data=sim; by group ;run;
proc means data=sim noprint;
output out=out2 sum(case)=sum_case sum(time)=sum_time;
by group ;run;

data raw_group0_fu (where=(group=0) rename=(_freq_=n0
sum_case=case0 sum_time=t0 ))
    raw_group1_fu (where=(group=1)rename=(_freq_=n1
sum_case=case1 sum_time=t1));
set out2;run;

data result_raw_fu;merge raw_group0_fu  raw_group1_fu ;
r=t1/t0;
p=case1/(case0+case1);
rr=p/(r*(1-p));
VE=1-rr;
var_ve=rr*rr*(1/(case1)+ 1/(case0)+ 1/t1 + 1/t0);
se_ve=sqrt(var_ve);
ll=1/(1+(2*(case0+1))/(2*case1*finv(0.025,2*case1,2*(case
0+1))));
ul=1/(1+(2*case0)/(2*(case1+1)*finv(0.975,2*(case1+1),2*c
ase0)));
ll_ve=1-ul/(r*(1-ul));
ul_ve=1-ll/(r*(1-ll));
tp=(&p_naive*&ve11)+ ((1-&p_naive)*&ve10);*define true
parameter of VE general weighted;
lenght=ul_ve-ll_ve;
if ll_ve<=tp and ul_ve >=tp then covtot=1; else covtot=0;
*define the coverage;
keep ve se_ve ll_ve ul_ve tp lenght covtot; run;

%if &i = 1 %then %do;
     data outall_fu_raw;
       set result_raw_fu;
     %end;
   %else %do;
     proc append base=outall_fu_raw
data=result_raw_fu;quit;
     %end;

*################ FU halloran ###############;
proc sort data=sim; by naive group ;run;
proc means data=sim noprint;
output out=out1 sum(case)=sum_case sum(time)=sum_time;
by naive group ;run;

data hal_fu_group0 (where=(group=0))
    hal_fu_group1 (where=(group=1));
set out1;
R=sum_case/sum_time;
drop _type_;run;

data hal_group0na0_fu (where=(naive=0)
rename=(_freq_=n0 sum_case=case0 sum_time=t0 r=r0))
             hal_group0na1_fu (where=(naive=1)
rename=(_freq_=n1 sum_case=case1 sum_time=t1 r=r1));
set hal_fu_group0; run;

data hal_group1na0_fu (where=(naive=0)
rename=(_freq_=n0 sum_case=case0 sum_time=t0 r=r0))
             hal_group1na1_fu (where=(naive=1)
rename=(_freq_=n1 sum_case=case1 sum_time=t1 r=r1));
set hal_fu_group1; run;

data hal_gr0_fu;merge hal_group0na0_fu  hal_group0na1_fu
;
we_na0=n0/(n0+n1);
we_na1=n1/(n0+n1);
den=we_na0*r0 +we_na1*r1;
var_r0=r0*r0*(1/(case0) +1/t0);
var_r1=r1*r1*(1/(case1) +1/t1);
var_den=(we_na0*we_na0*var_r0)+(we_na1*we_na1*var_r
1);
keep den var_den;run;

data hal_gr1_fu;merge hal_group1na0_fu  hal_group1na1_fu
;
we_na0=n0/(n0+n1);
we_na1=n1/(n0+n1);
num=we_na0*r0 +we_na1*r1;
var_r0=r0*r0*(1/(case0) +1/t0);
var_r1=r1*r1*(1/(case1) +1/t1);
var_num=(we_na0*we_na0*var_r0)+(we_na1*we_na1*var_
r1);
keep num var_num;run;

data result_hal_fu;merge hal_gr0_fu hal_gr1_fu;
rr=num/den;
VE=1-rr;
var_rr=rr*rr* ( var_num/(num*num)+ var_den/(den*den));
var_ve=var_rr;
se_ve=sqrt(var_ve);
*********calculating CI **********;
eta=log(ve/(1-ve)); * using transformation of logit
FUNCTION OF ve;
var_eta=(1/(ve*(1-ve)))*(1/(ve*(1-ve)))*var_VE;*use delta
methods;
ll_eta=eta-(1.96*sqrt(var_eta));
ul_eta=eta+(1.96*sqrt(var_eta));
```

```
ve=exp(eta)/(1+exp(eta));* using transformation of expit
FUNCTION OF ve;
ll_ve=exp(ll_eta)/(1+exp(ll_eta));
ul_ve=exp(ul_eta)/(1+exp(ul_eta));
*define true parameter of VE hal assume in here N the same
as T;
numtp=&p_naive*((1-&ve11)*&p_c01) + (1-
&p_naive)*((1-&ve10)*&p_c00);
dentp=&p_naive*&p_c01 +(1-&p_naive)*&p_c00;
tp=1- numtp/dentp ;
lenght=ul_ve-ll_ve;
if ll_ve<=tp and ul_ve >=tp then covtot=1; else covtot=0;
*define the coverage;
keep ve se_ve ll_ve ul_ve tp lenght covtot; run;
run;

%if &i = 1 %then %do;
     data outall_fu_hal;
       set result_hal_fu;
     %end;
   %else %do;
     proc append base=outall_fu_hal data=result_hal_fu;quit;
     %end;

* FU general weighted - number of cases and follow up time;

data naive0 (where=(naive=0))
    naive1 (where=(naive=1));
set out1;drop _type_;run;

data group0na0 (where=(group=0) rename=(_freq_=n0
sum_case=case0 sum_time=t0))
             group1na0 (where=(group=1) rename=(_freq_=n1
sum_case=case1 sum_time=t1));
set naive0; run;

data group0na1 (where=(group=0) rename=(_freq_=N0
sum_case=case0 sum_time=t0))
             group1na1 (where=(group=1) rename=(_freq_=N1
sum_case=case1 sum_time=t1));
set naive1; run;

data na0;merge group0na0  group1na0 ;
r=t1/t0;
p=case1/(case0+case1);
rr=p/(r*(1-p));

VE_non=1-rr;
var_ve_non=rr*rr*(1/(case0) + 1/(case1) + 1/t0 + 1/t1 );
ll=1/(1+(2*(case0+1))/(2*case1*finv(0.025,2*case1,2*(case
0+1))));
ul=1/(1+(2*case0)/(2*(case1+1)*finv(0.975,2*(case1+1),2*c
ase0)));
VE_non_ll=1-ul/(r*(1-ul));
VE_non_ul=1-ll/(r*(1-ll));
N_non=N0+N1;
t_non=t0+t1;run;

data na1;merge group0na1  group1na1 ;
r=t1/t0;
p=case1/(case0+case1);
rr=p/(r*(1-p)) ;
VE_naive=1-rr;
var_ve_naive=rr*rr*(1/(case0) + 1/(case1) + 1/t0 + 1/t1 );
ll=1/(1+(2*(case0+1))/(2*case1*finv(0.025,2*case1,2*(case
0+1))));
ul=1/(1+(2*case0)/(2*(case1+1)*finv(0.975,2*(case1+1),2*c
ase0)));
VE_naive_ll=1-ul/(r*(1-ul));
VE_naive_ul=1-ll/(r*(1-ll));
N_naive=N0+N1;
t_naive=t0+t1;run;

/*weighted based on number of cases*/
data g_fu_N;
merge na0 na1;
name='we_N';
w_na1_N=N_naive/(N_naive+N_non);
w_na0_N=N_non/(N_naive+N_non);
vE=(w_na1_N*VE_naive)+(w_na0_N*VE_non);
var_ve=(w_na1_N*w_na1_N*var_ve_naive)+(w_na0_N*w_
na0_N*var_ve_non);
se_ve=sqrt(var_ve);
ll_ve=(w_na1_N*VE_naive_ll)+(w_na0_N*VE_non_ll);
ul_ve=(w_na1_N*VE_naive_ul)+(w_na0_N*VE_non_ul);
lenght=ul_ve-ll_ve;
tp=(&p_naive*&ve11)+ ((1-&p_naive)*&ve10);*define true
parameter of VE general weighted;
if ll_ve<=tp and ul_ve >=tp then covtot=1; else covtot=0;
*define the coverage;
keep name ve se_ve ll_ve ul_ve lenght tp covtot; run;

/*weighted based on total number of follow up time*/
data g_fu_t;
merge na0 na1;
name='we_t';
we_naive_t=t_naive/(t_naive+t_non);
we_non_t=t_non/(t_naive+t_non);
VE=(we_naive_t*VE_naive)+(we_non_t*VE_non);
var_ve=(we_naive_t*we_naive_t*var_ve_naive)+(we_non_t
*we_non_t*var_ve_non);
se_ve=sqrt(var_ve);
ll_ve=(we_naive_t*VE_naive_ll)+(we_non_t*VE_non_ll);
ul_ve=(we_naive_t*VE_naive_ul)+(we_non_t*VE_non_ul);
lenght=ul_ve-ll_ve;
tp=(&p_naive*&ve11)+ ((1-&p_naive)*&ve10);*define true
parameter of VE general weighted;
if ll_ve<=tp and ul_ve >=tp then covtot=1; else covtot=0;
*define the coverage;
keep name ve se_ve ll_ve ul_ve lenght tp covtot; run;

data result_g_fu;set g_fu_N g_fu_t;run;

%if &i = 1 %then %do;
     data outall_fu_g;
       set result_g_fu;
     %end;
   %else %do;
     proc append base=outall_fu_g data=result_g_fu;quit;
     %end;

*######### FU - by naive  ###############;

data fu_na1;set na1;name='naive';
ve=VE_naive;
se_ve=sqrt(var_ve_naive);
ll_ve=VE_naive_ll;
ul_ve=VE_naive_ul;
tp=&ve11;*define true parameter of VE naive;
lenght=ul_ve-ll_ve;
if ll_ve<=tp and ul_ve >=tp then covtot=1; else covtot=0;
*define the coverage;
keep name ve se_ve ll_ve ul_ve tp lenght covtot; run;
```

```
data fu_na0;set na0;name='non-n';
ve=VE_non;
se_ve=sqrt(var_ve_non);
ll_ve=VE_non_ll;
ul_ve=VE_non_ul;
tp=&ve10;*define true parameter of VE naive;
lenght=ul_ve-ll_ve;
if ll_ve<=tp and ul_ve >=tp then covtot=1; else covtot=0;
*define the coverage;
keep name ve se_ve ll_ve ul_ve tp lenght covtot; run;

data result_fu_bynaive;set fu_na0 fu_na1;run;

%if &i = 1 %then %do;
     data outall_fu_bynaive;
       set result_fu_bynaive;
     %end;
   %else %do;
     proc append base=outall_fu_bynaive
data=result_fu_bynaive;quit;
     %end;

*###### FU - precision approach haber  #############;

*non-naive*;
data hbr_fu_na0;merge group0na0  group1na0 ;
r=t1/t0;
p=case1/(case0+case1);
rr=p/(r*(1-p)) ;
VE_n0=1-rr;
var_ve_n0=rr*rr*(1/(case0) + 1/(case1) + 1/t0 + 1/t1 );
ll=1/(1+(2*(case0+1))/(2*case1*finv(0.025,2*case1,2*(case
0+1))));
ul=1/(1+(2*case0)/(2*(case1+1)*finv(0.975,2*(case1+1),2*c
ase0)));
VE_n0_ll=1-ul/(r*(1-ul));
VE_n0_ul=1-ll/(r*(1-ll));
We_na0=1/ var_ve_n0;
**********calculating the variance of we_na0 ***********;
Var_log_rr=(1/(case0) + 1/(case1));
var_log_var=((1/var_log_rr)**2)*( (1-
case1/t1)*(case1)/((case1)**4)+ (1-
case0/t0)*(case0)/((case0)**4) );
var_log_we0=4* Var_log_rr+
var_log_var+4*sqrt(Var_log_rr* var_log_var);
var_we0=(we_na0**2)* var_log_we0;
wexVE_na0= We_na0* VE_n0;
var_log_wexVE_na0= var_log_we0+ ((1/ VE_n0)**2)*
var_ve_n0 +2*sqrt(var_log_we0*((1/ VE_n0)**2)*
var_ve_n0);
var_wexVE_na0= (wexVE_na0**2)* var_log_wexVE_na0;
VE_llxwe_na0= VE_n0_ll* We_na0;
VE_ulxwe_na0= VE_n0_ul* We_na0;
keep wexVE_na0 var_wexVE_na0 var_we0 we_na0
VE_llxwe_na0 VE_ulxwe_na0;run;

*naive*;
data hbr_fu_na1;merge group0na1  group1na1 ;
r=t1/t0;
p=case1/(case0+case1);
rr=p/(r*(1-p)) ;
VE_n1=1-rr;
var_ve_n1=rr*rr*(1/(case0) + 1/(case1) + 1/t0 + 1/t1 );
ll=1/(1+(2*(case0+1))/(2*case1*finv(0.025,2*case1,2*(case
0+1))));
ul=1/(1+(2*case0)/(2*(case1+1)*finv(0.975,2*(case1+1),2*c
ase0)));
VE_n1_ll=1-ul/(r*(1-ul));
VE_n1_ul=1-ll/(r*(1-ll));
We_na1=1/ var_ve_n1;
**********calculating the variance of we_na1 ***********;
Var_log_rr=(1/(case0) + 1/(case1));
var_log_var=((1/var_log_rr)**2)*( (1-
case1/t1)*(case1)/((case1)**4)+ (1-
case0/t0)*(case0)/((case0)**4) );
var_log_we1=4* Var_log_rr+
var_log_var+4*sqrt(Var_log_rr* var_log_var);
var_we1=(we_na1**2)* var_log_we1;
wexVE_na1= We_na1* VE_n1;
var_log_wexVE_na1= var_log_we1+ ((1/ VE_n1)**2)*
var_ve_n1 +2*sqrt(var_log_we1*((1/ VE_n1)**2)*
var_ve_n1);
var_wexVE_na1= (wexVE_na1**2)* var_log_wexVE_na1;
VE_llxwe_na1= VE_n1_ll* We_na1;
VE_ulxwe_na1= VE_n1_ul* We_na1;
keep wexVE_na1 var_wexVE_na1 var_we1 we_na1
VE_llxwe_na1 VE_ulxwe_na1;run;

data result_hbr_fu;merge hbr_fu_na1 hbr_fu_na0;
num= wexVE_na1+ wexVE_na0;
den= we_na1+we_na0;
Var_num= var_wexVE_na1+ var_wexVE_na0;
Var_den= var_we1+ var_we0;
VE=num/den;
Var_ve_log=((1/num)**2)*var_num+ ((1/den)**2)*var_den-
2*sqrt( ((1/num)**2)*var_num * ((1/den)**2)*var_den );
Var_ve=(ve**2)* Var_ve_log;
se_ve=sqrt(var_ve);
ll_ve= (VE_llxwe_na1+ VE_llxwe_na0)/den;
ul_ve= (VE_ulxwe_na1+ VE_ulxwe_na0)/den;
*define true parameter of VE precision;
tp=(&p_naive*&ve11)+ ((1-&p_naive)*&ve10);
lenght=ul_ve-ll_ve;
if ll_ve<=tp and ul_ve >=tp then covtot=1; else covtot=0;
*define the coverage;
keep ve se_ve ll_ve ul_ve tp lenght covtot; run;

%if &i = 1 %then %do;
     data outall_fu_hbr;
       set result_hbr_fu;
     %end;
   %else %do;
     proc append base=outall_fu_hbr data=result_hbr_fu;quit;
     %end;

%mend sim_FU;
```

C3.3 Proportional Hazard (HR)

```
%macro sim_hazard;

*###### RAW VE_hazard #############;
ods output parameterestimates=hr_raw_parm;
ods select parameterestimates;
proc phreg data=sim;
class group(ref='0');
model time*censored(1)=group/risklimits covb;
run;

data result_hr_raw;set hr_raw_parm;
ve=1-HazardRatio;
```

```
varHR=(HazardRatio*HazardRatio)*(StdErr*StdErr);*use
delta methods;
varVE=varHR;*use delta methods;
se_ve=sqrt(varVe);
ll_ve=1-HRUpperCL;
ul_ve=1-HRLowerCL;
tp=(&p_naive*&ve11)+ ((1-&p_naive)*&ve10);*define true
parameter of VE ;
lenght=ul_ve-ll_ve;
if ll_ve<=tp and ul_ve >=tp then covtot=1; else covtot=0;
*define the coverage;
keep ve se_ve ll_ve ul_ve tp lenght covtot; run;

%if &i = 1 %then %do;
     data outall_hr_raw;
       set result_hr_raw;
     %end;
   %else %do;
     proc append base=outall_hr_raw
data=result_hr_raw;quit;
     %end;

*############## bynaive VE_hazard #############;

            proc sort data=sim ;by naive;run;
            ods select parameterestimates;
            ods output parameterestimates=hr_bynaive_parm;
            proc phreg data=sim ;
            class group(ref='0');
            model time*censored(1)=group/risklimits covb;
            by naive;run;

data result_hr_bynaive;set hr_bynaive_parm;
if naive=1 then name='naive';else name='non-n';
ve=1-HazardRatio;
varHR=(HazardRatio*HazardRatio)*(StdErr*StdErr);*use
delta methods;
varVE=varHR;*use delta methods;
se_ve=sqrt(varVe);
ll_ve=1-HRUpperCL;
ul_ve=1-HRLowerCL;
if naive=1 then tp=&ve11; else tp=&ve10;*define true
parameter of VE RAW;
lenght=ul_ve-ll_ve;
if ll_ve<=tp and ul_ve >=tp then covtot=1;else covtot=0;
*define the coverage;
keep name ve se_ve ll_ve ul_ve tp lenght covtot; run;

%if &i = 1 %then %do;
     data outall_hr_bynaive;
       set result_hr_bynaive;
     %end;
   %else %do;
     proc append base=outall_hr_bynaive
data=result_hr_bynaive;quit;
     %end;

*####  general weighted VE_hazard- number of cases and
follow up time #######;

proc sort data=sim; by naive ;run;
proc means data=sim noprint;
output out=outhr sum(case)=sum_case sum(time)=sum_time;
by naive ;
run;

data hr_g_naive0 (where=(naive=0)rename=(_freq_=n0
sum_case=case0 sum_time=t0))
    hr_g_naive1 (where=(naive=1)rename=(_freq_=n1
sum_case=case1 sum_time=t1));
set outhr;drop _type_;run;

data hr_g_n0 (where=(name='non-n')rename=(ve=ve0
se_ve=se0 ll_ve=ll0 ul_ve=ul0))
    hr_g_n1 (where=(name='naive')rename=(ve=ve1
se_ve=se1 ll_ve=ll1 ul_ve=ul1));
set result_hr_bynaive;run;

***weighted by N;
data hr_g_n;merge hr_g_n0 hr_g_n1 hr_g_naive0
hr_g_naive1;
name='we_N';
w0=n0/(n0+n1);
w1=n1/(n0+n1);
ve=w0*ve0+w1*ve1;
var_ve= w0*w0*se0*se0 + w1*w1*se1*se1 ;
se_ve=sqrt(var_ve);
ll_ve=w0*ll0+w1*ll1;
ul_ve=w0*ul0+w1*ul1;
lenght=ul_ve-ll_ve;
tp=(&p_naive*&ve11)+ ((1-&p_naive)*&ve10);*define true
parameter of VE general weighted;
if ll_ve<=tp and ul_ve >=tp then covtot=1; else covtot=0;
*define the coverage;
keep name ve se_ve ll_ve ul_ve lenght tp covtot; run;

***weighted by t;
data hr_g_t;merge hr_g_n0 hr_g_n1 hr_g_naive0
hr_g_naive1;
name='we_T';
w0=t0/(t0+t1);
w1=t1/(t0+t1);
ve=w0*ve0+w1*ve1;
var_ve= w0*w0*se0*se0 + w1*w1*se1*se1 ;
se_ve=sqrt(var_ve);
ll_ve=w0*ll0+w1*ll1;
ul_ve=w0*ul0+w1*ul1;
lenght=ul_ve-ll_ve;
tp=(&p_naive*&ve11)+ ((1-&p_naive)*&ve10);*define true
parameter of VE general weighted;
if ll_ve<=tp and ul_ve >=tp then covtot=1; else covtot=0;
*define the coverage;
keep name ve se_ve ll_ve ul_ve lenght tp covtot; run;

data result_hr_g;set hr_g_t hr_g_n;run;

%if &i = 1 %then %do;
     data outall_hr_g;
       set result_hr_g;
     %end;
   %else %do;
     proc append base=outall_hr_g data=result_hr_g;quit;
     %end;

*######## haber VE_hazard #############;

data hr_hbr_parm;set hr_bynaive_parm;
ve=1-HazardRatio;
ll_ve=1-HRUpperCL;
ul_ve=1-HRLowerCL;
varHR=(HazardRatio*HazardRatio)*(StdErr*StdErr);*use
delta methods;
varVE=varHR;*use delta methods;
```

```
weight=1/varVE;
VExw=weight*VE;
VExw_ll=weight*ll_ve;
VExw_ul=weight*ul_ve;run;

proc means data=hr_hbr_parm noprint;output
out=hr_hbr_parm2
sum(VExw)=sum_wexVE
sum(VExw_ll)=sum_wexVE_ll
sum(VExw_ul)=sum_wexVE_ul
sum(weight)=sum_we;run;

data result_hr_hbr;set hr_hbr_parm2 ;
VE=sum_wexVE/sum_we;
ll_ve=sum_wexVE_ll/sum_we;
ul_ve=sum_wexVE_ul/sum_we;
*define true value (tp);
tp=(&p_naive*&ve11)+ ((1-&p_naive)*&ve10);*define true
parameter of VE general weighted;
lenght=ul_ve-ll_ve;
if ll_ve<=tp and ul_ve >=tp then covtot=1; else covtot=0;
*define the coverage;
keep ve ll_ve ul_ve tp lenght covtot; run;

%if &i = 1 %then %do;
     data outall_hr_hbr;
       set result_hr_hbr;
     %end;
   %else %do;
     proc append base=outall_hr_hbr data=result_hr_hbr;quit;
     %end;

%mend sim_hazard;
```

C3. 4 Poisson Regression

```
%macro sim_poisson;

*######## RAW VE_poisson ############;
data sim2;set sim;logt=log(time);run;

ods select parameterestimates;
ods output parameterestimates=pois_raw_parm;
proc genmod data=sim2;
class group(ref='0')/param=ref;
model case=group/offset=logt dist=poisson link=log
dscale;run;

data result_poiss_raw;set pois_raw_parm;
if parameter ne 'group' then delete;
VE=1-exp(Estimate);
varRR=(exp(Estimate)*exp(Estimate))*(StdErr*StdErr);*use
delta methods;
varVE=varRR;*use delta methods;
se_ve=sqrt(varVe);
ll_ve=1-exp(UpperWaldCL);
ul_ve=1-exp(LowerWaldCL);
tp=(&p_naive*&ve11)+ ((1-&p_naive)*&ve10);*define true
parameter of VE general weighted;
lenght=ul_ve-ll_ve;
if ll_ve<=tp and ul_ve >=tp then covtot=1; else covtot=0;
*define the coverage;
keep ve se_ve ll_ve ul_ve tp lenght covtot; run;

%if &i = 1 %then %do;
     data outall_poiss_raw;
       set result_poiss_raw;
     %end;
   %else %do;
     proc append base=outall_poiss_raw
data=result_poiss_raw;quit;
     %end;

*######### by naive VE_poisson ###############;
proc sort data=sim2;by naive;run;
ods select parameterestimates;
ods output parameterestimates=pois_bynaive_parm;
proc genmod data=sim2;
class group(ref='0')/param=ref;
model case=group/offset=logt dist=poisson link=log dscale;
by naive;
run;

data result_poiss_bynaive;set pois_bynaive_parm;
if parameter ne 'group' then delete;
if naive=1 then name='naive';else name='non-n';
VE=1-exp(Estimate);
varRR=(exp(Estimate)*exp(Estimate))*(StdErr*StdErr);*use
delta methods;
varVE=varRR;*use delta methods;
se_ve=sqrt(varVe);
ll_ve=1-exp(UpperWaldCL);
ul_ve=1-exp(LowerWaldCL);
if naive=1 then tp=&ve11; else tp=&ve10;*define true
parameter of VE RAW;
lenght=ul_ve-ll_ve;
if ll_ve<=tp and ul_ve >=tp then covtot=1;else covtot=0;
*define the coverage;
keep name ve se_ve ll_ve ul_ve tp lenght covtot; run;

%if &i = 1 %then %do;
     data outall_poiss_bynaive;
       set result_poiss_bynaive;
     %end;
   %else %do;
     proc append base=outall_poiss_bynaive
data=result_poiss_bynaive;quit;
     %end;

*###  general weighted VE_poisson- number of cases ####;

proc sort data=sim; by naive ;run;
proc means data=sim noprint;
output out=outpoiss sum(case)=sum_case
sum(time)=sum_time ;
by naive ;
run;

data poiss_g_naive0 (where=(naive=0)rename=(_freq_=n0
sum_case=case0 sum_time=t0))
    poiss_g_naive1 (where=(naive=1)rename=(_freq_=n1
sum_case=case1 sum_time=t1));
set outpoiss;drop _type_;run;

data poiss_g_n0 (where=(name='non-n')rename=(ve=ve0
se_ve=se0 ll_ve=ll0 ul_ve=ul0))
    poiss_g_n1 (where=(name='naive')rename=(ve=ve1
se_ve=se1 ll_ve=ll1 ul_ve=ul1));
set result_poiss_bynaive;run;

***weighted by N;
data poiss_g_n;
```

```
merge poiss_g_naive0 poiss_g_naive1 poiss_g_n0
poiss_g_n1 ;
name='we_N';
w0=n0/(n0+n1);
w1=n1/(n0+n1);
ve=w0*ve0+w1*ve1;
var_ve= w0*w0*se0*se0 + w1*w1*se1*se1 ;
se_ve=sqrt(var_ve);
ll_ve=w0*ll0+w1*ll1;
ul_ve=w0*ul0+w1*ul1;
lenght=ul_ve-ll_ve;
tp=(&p_naive*&ve11)+ ((1-&p_naive)*&ve10);*define true
parameter of VE general weighted;
if ll_ve<=tp and ul_ve >=tp then covtot=1; else covtot=0;
*define the coverage;
keep name ve se_ve ll_ve ul_ve lenght tp covtot; run;

***weighted by t;
data poiss_g_t;
merge poiss_g_naive0 poiss_g_naive1 poiss_g_n0
poiss_g_n1 ;
name='we_T';
w0=t0/(t0+t1);
w1=t1/(t0+t1);
ve=w0*ve0+w1*ve1;
var_ve= w0*w0*se0*se0 + w1*w1*se1*se1 ;
se_ve=sqrt(var_ve);
ll_ve=w0*ll0+w1*ll1;
ul_ve=w0*ul0+w1*ul1;
lenght=ul_ve-ll_ve;
tp=(&p_naive*&ve11)+ ((1-&p_naive)*&ve10);*define true
parameter of VE general weighted;
if ll_ve<=tp and ul_ve >=tp then covtot=1; else covtot=0;
*define the coverage;
keep name ve se_ve ll_ve ul_ve lenght tp covtot; run;

data result_poiss_g;set poiss_g_t poiss_g_n;run;

%if &i = 1 %then %do;
    data outall_poiss_g;
      set result_poiss_g;
    %end;
  %else %do;
    proc append base=outall_poiss_g
data=result_poiss_g;quit;
    %end;

*###### haber VE_poiss ########;

data poiss_hbr_parm;set pois_bynaive_parm;
if parameter ne 'group' then delete;
rr=exp(Estimate);
VE=1-rr;
VE_ll=1-exp(UpperWaldCL);
VE_ul=1-exp(LowerWaldCL);
varRR=(rr*rr)*(StdErr*StdErr);*use delta methods;
varVE=varRR;*use delta methods;
weight=1/varVE;
VExw=weight*VE;
VExw_ll=weight*VE_ll;
VExw_ul=weight*VE_ul;run;

proc means data=poiss_hbr_parm noprint;
output out=poiss_hbr_parm2
sum(VExw)=sum_wexVE
sum(VExw_ll)=sum_wexVE_ll
sum(VExw_ul)=sum_wexVE_ul
sum(weight)=sum_we;run;

data result_poiss_hbr;set poiss_hbr_parm2 ;
VE=sum_wexVE/sum_we;
ll_ve=sum_wexVE_ll/sum_we;
ul_ve=sum_wexVE_ul/sum_we;
*define true value (tp);
tp=(&p_naive*&ve11)+ ((1-&p_naive)*&ve10);
lenght=ul_ve-ll_ve;
if ll_ve<=tp and ul_ve >=tp then covtot=1; else covtot=0;
*define the coverage;
keep ve se_ve ll_ve ul_ve tp lenght covtot; run;

%if &i = 1 %then %do;
    data outall_poiss_hbr;
      set result_poiss_hbr;
    %end;
  %else %do;
    proc append base=outall_poiss_hbr
data=result_poiss_hbr;quit;
    %end;

%mend sim_poisson;
```

C4. Macro Compile

C4.1 Attack Rates Approach (AR)

```
%macro compile_AR;

*#AR_raw final output#;
data outall_ar_raw; set outall_ar_raw ;
if se_ve=. then delete;if lenght=. then delete;run;

proc means data= outall_ar_raw noprint;
        output out= outall_ar_raw2    mean(tp)=tp
                  mean(ve)=ve stddev(ve)=std_ve
                   mean(se_ve)=se_ve
                   mean(ll_ve)=ll_ve mean(ul_ve)=ul_ve
                   mean(lenght)=lenght
                   sum(covtot)=covtot;run;

data outall_ar_raw2; set outall_ar_raw2;
method='AR';name='raw';
mse=(ve-tp)*(ve-tp)+(std_ve)*(std_ve);
coverage=covtot/_freq_*100;
keep method name _freq_ tp ve std_ve se_ve ll_ve ul_ve
lenght coverage mse ;run;

*#AR_hal final output#;
data outall_ar_hal; set outall_ar_hal ;
if se_ve=. then delete;if lenght=. then delete;run;

proc means data= outall_ar_hal noprint;
        output out= outall_ar_hal2 mean(tp)=tp
                  mean(ve)=ve stddev(ve)=std_ve
                   mean(se_ve)=se_ve
                   mean(ll_ve)=ll_ve mean(ul_ve)=ul_ve
                   mean(lenght)=lenght
                   sum(covtot)=covtot;run;

data outall_ar_hal2; set outall_ar_hal2;
method='AR';name='hal';
mse=(ve-tp)*(ve-tp)+(std_ve)*(std_ve);
coverage=covtot/_freq_*100;
```

```
keep method name _freq_ tp ve std_ve se_ve ll_ve ul_ve
lenght coverage mse ;run;

*#AR_g final output#;
data outall_ar_g; set outall_ar_g ;
if se_ve=. then delete;if lenght=. then delete;run;

proc means data= outall_ar_g noprint;
            output out= outall_ar_g2 mean(tp)=tp
            mean(ve)=ve stddev(ve)=std_ve
            mean(se_ve)=se_ve
            mean(ll_ve)=ll_ve mean(ul_ve)=ul_ve
            mean(lenght)=lenght
            sum(covtot)=covtot;run;
data outall_ar_g2; set outall_ar_g2;
method='AR';name='we_N';
mse=(ve-tp)*(ve-tp)+(std_ve)*(std_ve);
coverage=covtot/_freq_*100;
keep method name _freq_ tp ve std_ve se_ve ll_ve ul_ve
lenght coverage mse ;run;

*#AR_bynaive final output#;
data outall_ar_bynaive; set outall_ar_bynaive ;
if se_ve=. then delete;if lenght=. then delete;run;

proc sort data=outall_ar_bynaive; by name;run;
proc means data= outall_ar_bynaive noprint;
            output out= outall_ar_bynaive2 mean(tp)=tp
            mean(ve)=ve stddev(ve)=std_ve
              mean(se_ve)=se_ve
              mean(ll_ve)=ll_ve mean(ul_ve)=ul_ve
              mean(lenght)=lenght
             sum(covtot)=covtot;by name;run;

data outall_ar_bynaive2; set outall_ar_bynaive2;
method='AR';
mse=(ve-tp)*(ve-tp)+(std_ve)*(std_ve);
coverage=covtot/_freq_*100;
keep method name _freq_ tp ve std_ve se_ve ll_ve ul_ve
lenght coverage mse ;run;

*#AR_hbr final output#;
data outall_ar_hbr; set outall_ar_hbr ;
if se_ve=. then delete;if lenght=. then delete;run;

proc means data= outall_ar_hbr noprint;
            output out= outall_ar_hbr2 mean(tp)=tp
                        mean(ve)=ve stddev(ve)=std_ve
              mean(se_ve)=se_ve
              mean(ll_ve)=ll_ve mean(ul_ve)=ul_ve
              mean(lenght)=lenght
             sum(covtot)=covtot;run;
data outall_ar_hbr2; set outall_ar_hbr2;
method='AR';name='pres';
mse=(ve-tp)*(ve-tp)+(std_ve)*(std_ve);
coverage=covtot/_freq_*100;
keep method name _freq_ tp ve std_ve se_ve ll_ve ul_ve
lenght coverage mse ;run;

data final_ar;
set outall_ar_bynaive2 outall_ar_raw2 outall_ar_g2
outall_ar_hal2  outall_ar_hbr2;
n=&n;
r_naive=&p_naive;
IRc_naive=&p_c01;
IRc_non=&p_c00;
VE_naive=&ve11;
ve_non=&ve10;
run;

%mend compile_AR;
```

C4.2 Person Time Approach (FU)

```
%macro compile_FU;

*#FU_raw final output#;
data outall_fu_raw; set outall_fu_raw ;
if se_ve=. then delete;if lenght=. then delete;run;

proc means data= outall_fu_raw noprint;
            output out= outall_fu_raw2 mean(tp)=tp
            mean(ve)=ve stddev(ve)=std_ve
              mean(se_ve)=se_ve
            mean(ll_ve)=ll_ve mean(ul_ve)=ul_ve
            mean(lenght)=lenght
              sum(covtot)=covtot; run;
data outall_fu_raw2; set outall_fu_raw2;
method='FU';name='raw';
mse=(ve-tp)*(ve-tp)+(std_ve)*(std_ve);
coverage=covtot/_freq_*100;
keep method name _freq_ tp ve std_ve se_ve ll_ve ul_ve
lenght coverage mse ;run;

*#fu_hal final output#;
data outall_fu_hal; set outall_fu_hal ;
if se_ve=. then delete;if lenght=. then delete;run;

proc means data= outall_fu_hal noprint;
            output out= outall_fu_hal2 mean(tp)=tp
            mean(ve)=ve stddev(ve)=std_ve
            mean(se_ve)=se_ve
            mean(ll_ve)=ll_ve mean(ul_ve)=ul_ve
            mean(lenght)=lenght
            sum(covtot)=covtot;run;

data outall_fu_hal2; set outall_fu_hal2;
method='FU';name='hal';
mse=(ve-tp)*(ve-tp)+(std_ve)*(std_ve);
coverage=covtot/_freq_*100;
keep method name _freq_ tp ve std_ve se_ve ll_ve ul_ve
lenght coverage mse ;run;

*#FU_g final output#;
data outall_fu_g; set outall_fu_g ;
if se_ve=. then delete;if lenght=. then delete;run;

proc sort data=outall_fu_g; by name;run;
proc means data= outall_fu_g noprint;
            output out= outall_fu_g2 mean(tp)=tp
            mean(ve)=ve stddev(ve)=std_ve
            mean(se_ve)=se_ve
            mean(ll_ve)=ll_ve mean(ul_ve)=ul_ve
            mean(lenght)=lenght
             sum(covtot)=covtot;by name;run;
data outall_fu_g2;   set outall_fu_g2;
method='FU';
mse=(ve-tp)*(ve-tp)+(std_ve)*(std_ve);
coverage=covtot/_freq_*100;
keep method name _freq_ tp ve std_ve se_ve ll_ve ul_ve
lenght coverage mse ;run;

*#FU_bynaive final output#;
```

```
data outall_fu_bynaive; set outall_fu_bynaive ;
if se_ve=. then delete;if lenght=. then delete;run;

proc sort data=outall_fu_bynaive; by name;run;
proc means data= outall_fu_bynaive noprint;
            output out= outall_fu_bynaive2 mean(tp)=tp
            mean(ve)=ve stddev(ve)=std_ve
              mean(se_ve)=se_ve
              mean(ll_ve)=ll_ve mean(ul_ve)=ul_ve
              mean(lenght)=lenght
             sum(covtot)=covtot;by name;run;
data outall_fu_bynaive2; set outall_fu_bynaive2;
method='FU';
mse=(ve-tp)*(ve-tp)+(std_ve)*(std_ve);
coverage=covtot/_freq_*100;
keep method name _freq_ tp ve std_ve se_ve ll_ve ul_ve
lenght coverage mse;run;

*#FU_precision hbr final output#;
data outall_fu_hbr; set outall_fu_hbr ;
if se_ve=. then delete;if lenght=. then delete;run;

proc means data= outall_fu_hbr noprint;
            output out= outall_fu_hbr2 mean(tp)=tp
            mean(ve)=ve stddev(ve)=std_ve
              mean(se_ve)=se_ve
              mean(ll_ve)=ll_ve mean(ul_ve)=ul_ve
              mean(lenght)=lenght
             sum(covtot)=covtot;run;
data outall_fu_hbr2; set outall_fu_hbr2;
method='FU';name='pres';
mse=(ve-tp)*(ve-tp)+(std_ve)*(std_ve);
coverage=covtot/_freq_*100;
keep method name _freq_ tp ve std_ve se_ve ll_ve ul_ve
lenght coverage mse ;run;

data final_FU;
set outall_fu_bynaive2 outall_fu_raw2 outall_fu_g2
outall_fu_hal2  outall_fu_hbr2;
n=&n;
r_naive=&p_naive;
IRc_naive=&p_c01;
IRc_non=&p_c00;
VE_naive=&ve11;
ve_non=&ve10;
run;

%mend compile_FU;
```

C4.3 Proportional Hazard (HR)

```
%macro compile_hazard;

*#HR_raw final output#;
data outall_hr_raw; set outall_hr_raw ;
if ve<-200 then delete;if se_ve=. then delete;if lenght=. then
delete;run;

proc means data= outall_hr_raw noprint;
            output out=outall_hr_raw2 mean(tp)=tp
            mean(ve)=ve stddev(ve)=std_ve
              mean(se_ve)=se_ve
              mean(ll_ve)=ll_ve mean(ul_ve)=ul_ve
              mean(lenght)=lenght
             sum(covtot)=covtot;run;
data outall_hr_raw2; set outall_hr_raw2;
method='HR';name='raw';
mse=(ve-tp)*(ve-tp)+(std_ve)*(std_ve);
coverage=covtot/_freq_*100;
keep method name _freq_ tp ve std_ve se_ve ll_ve ul_ve
lenght coverage mse ;run;

*#HR_bynaive final output#;
data outall_hr_bynaive; set outall_hr_bynaive ;
if ve<-200 then delete;if se_ve=. then delete;if lenght=. then
delete;run;

proc sort data=outall_hr_bynaive; by name;run;
proc means data= outall_hr_bynaive noprint;
            output out=outall_hr_bynaive2 mean(tp)=tp
            mean(ve)=ve stddev(ve)=std_ve
              mean(se_ve)=se_ve
             mean(ll_ve)=ll_ve mean(ul_ve)=ul_ve
              mean(lenght)=lenght
              sum(covtot)=covtot;by name;run;
data outall_hr_bynaive2; set outall_hr_bynaive2;
method='HR';
mse=(ve-tp)*(ve-tp)+(std_ve)*(std_ve);
coverage=covtot/_freq_*100;
keep method name _freq_ tp ve std_ve se_ve ll_ve ul_ve
lenght coverage mse ;run;

*#HR_general weighted final output#;
data outall_hr_g; set outall_hr_g ;
if ve<-200 then delete;if se_ve=. then delete;if lenght=. then
delete;run;

proc sort data=outall_hr_g; by name;run;
proc means data= outall_hr_g noprint;
            output out=outall_hr_g2         mean(tp)=tp
            mean(ve)=ve stddev(ve)=std_ve
             mean(se_ve)=se_ve
              mean(ll_ve)=ll_ve mean(ul_ve)=ul_ve
            mean(lenght)=lenght
             sum(covtot)=covtot;by name;run;
data outall_hr_g2; set outall_hr_g2;
method='HR';
mse=(ve-tp)*(ve-tp)+(std_ve)*(std_ve);
coverage=covtot/_freq_*100;
keep method name _freq_ tp ve std_ve se_ve ll_ve ul_ve
lenght coverage mse ;run;

*#HR_precision hbr final output#;
data outall_hr_hbr; set outall_hr_hbr ;
if ve<-200 then delete;if lenght=. then delete;run;

proc means data= outall_hr_hbr noprint;
            output out=outall_hr_hbr2 mean(tp)=tp
            mean(ve)=ve stddev(ve)=std_ve
            mean(ll_ve)=ll_ve mean(ul_ve)=ul_ve
            mean(lenght)=lenght
             sum(covtot)=covtot;run;
data outall_hr_hbr2; set outall_hr_hbr2;
 method='HR';name='pres';
mse=(ve-tp)*(ve-tp)+(std_ve)*(std_ve);
coverage=covtot/_freq_*100;
keep method name _freq_ tp ve std_ve ll_ve ul_ve lenght
coverage mse ;run;

data final_HR;
set outall_hr_bynaive2 outall_hr_raw2 outall_hr_g2
outall_hr_hbr2;
n=&n;
r_naive=&p_naive;
```

```
IRc_naive=&p_c01;
IRc_non=&p_c00;
VE_naive=&ve11;
ve_non=&ve10;
run;

%mend compile_hazard;
```

C4.4 Poisson Regression

```
%macro compile_poisson;
*#poiss_raw final output#;
data outall_poiss_raw; set outall_poiss_raw ;
if ve<-200 then delete;if se_ve=. then delete;if lenght=. then
delete;run;

proc means data= outall_poiss_raw noprint;
            output out=outall_poiss_raw2 mean(tp)=tp
            mean(ve)=ve stddev(ve)=std_ve
              mean(se_ve)=se_ve
              mean(ll_ve)=ll_ve mean(ul_ve)=ul_ve
              mean(lenght)=lenght
             sum(covtot)=covtot; run;
data outall_poiss_raw2; set outall_poiss_raw2;
method='PO';name='raw';
mse=(ve-tp)*(ve-tp)+(std_ve)*(std_ve);
coverage=covtot/_freq_*100;
keep method name _freq_ tp ve std_ve se_ve ll_ve ul_ve
lenght coverage mse ;run;

*#poiss_bynaive final output#;
data outall_poiss_bynaive; set outall_poiss_bynaive ;
if ve<-200 then delete;if se_ve=. then delete;if lenght=. then
delete;run;

proc sort data=outall_poiss_bynaive; by name;run;
proc means data= outall_poiss_bynaive noprint;
            output out=outall_poiss_bynaive2 mean(tp)=tp
            mean(ve)=ve stddev(ve)=std_ve
              mean(se_ve)=se_ve
              mean(ll_ve)=ll_ve mean(ul_ve)=ul_ve
              mean(lenght)=lenght
             sum(covtot)=covtot;by name;run;
data outall_poiss_bynaive2; set outall_poiss_bynaive2;
method='PO';
mse=(ve-tp)*(ve-tp)+(std_ve)*(std_ve);
coverage=covtot/_freq_*100;
keep method name _freq_ tp ve std_ve se_ve ll_ve ul_ve
lenght coverage mse ;run;

*#poiss_general weighted final output#;
data outall_poiss_g; set outall_poiss_g ;
if ve<-200 then delete;if se_ve=. then delete;if lenght=. then
delete;run;

proc sort data=outall_poiss_g; by name;run;
proc means data= outall_poiss_g noprint;
            output out=outall_poiss_g2 mean(tp)=tp
            mean(ve)=ve stddev(ve)=std_ve
             mean(se_ve)=se_ve
              mean(ll_ve)=ll_ve mean(ul_ve)=ul_ve
              mean(lenght)=lenght
             sum(covtot)=covtot;by name;run;
data outall_poiss_g2;          set outall_poiss_g2;
method='PO';
mse=(ve-tp)*(ve-tp)+(std_ve)*(std_ve);
coverage=covtot/_freq_*100;
keep method name _freq_ tp ve std_ve se_ve ll_ve ul_ve
lenght coverage mse ;run;

*#poiss_haber final output#;
data outall_poiss_hbr; set outall_poiss_hbr ;
if ve<-200 then delete;if lenght=. then delete;run;

proc means data= outall_poiss_hbr noprint;
            output out=outall_poiss_hbr2 mean(tp)=tp
            mean(ve)=ve stddev(ve)=std_ve
            mean(ll_ve)=ll_ve mean(ul_ve)=ul_ve
            mean(lenght)=lenght
            sum(covtot)=covtot;run;
data outall_poiss_hbr2; set outall_poiss_hbr2;
method='PO';name='pres';
mse=(ve-tp)*(ve-tp)+(std_ve)*(std_ve);
coverage=covtot/_freq_*100;
keep method name _freq_ tp ve std_ve ll_ve ul_ve lenght
coverage mse ;run;

data final_Poisson;
set outall_poiss_bynaive2 outall_poiss_raw2 outall_poiss_g2
outall_poiss_hbr2;
n=&n;
r_naive=&p_naive;
IRc_naive=&p_c01;
IRc_non=&p_c00;
VE_naive=&ve11;
ve_non=&ve10;
run;

%mend compile_poisson;
```

Auteursrechtelijke overeenkomst

Ik/wij verlenen het wereldwijde auteursrecht voor de ingediende eindverhandeling:
Evaluation of Different Strategies for Weighted Average Vaccine Efficacy

Richting: **Master of Statistics-Biostatistics**
Jaar: **2012**

in alle mogelijke mediaformaten, - bestaande en in de toekomst te ontwikkelen - , aan de Universiteit Hasselt.

Niet tegenstaand deze toekenning van het auteursrecht aan de Universiteit Hasselt behoud ik als auteur het recht om de eindverhandeling, - in zijn geheel of gedeeltelijk -, vrij te reproduceren, (her)publiceren of distribueren zonder de toelating te moeten verkrijgen van de Universiteit Hasselt.

Ik bevestig dat de eindverhandeling mijn origineel werk is, en dat ik het recht heb om de rechten te verlenen die in deze overeenkomst worden beschreven. Ik verklaar tevens dat de eindverhandeling, naar mijn weten, het auteursrecht van anderen niet overtreedt.

Ik verklaar tevens dat ik voor het materiaal in de eindverhandeling dat beschermd wordt door het auteursrecht, de nodige toelatingen heb verkregen zodat ik deze ook aan de Universiteit Hasselt kan overdragen en dat dit duidelijk in de tekst en inhoud van de eindverhandeling werd genotificeerd.

Universiteit Hasselt zal mij als auteur(s) van de eindverhandeling identificeren en zal geen wijzigingen aanbrengen aan de eindverhandeling, uitgezonderd deze toegelaten door deze overeenkomst.

Voor akkoord,

Rakhmawati, Trias Wahyuni

Datum: **14/09/2012**