Estimation of the primary, secondary and composite effects of malaria vaccines using data on multiple clinical malaria episodes

Yin Bun Cheung [a,b,c,*], Xiangmei Ma [b], K.F. Lam [b,d], Paul Milligan [e]

[a] Programme in Health Services & Systems Research, Duke-NUS Medical School, 20 College Road, Singapore 169856, Singapore
[b] Centre for Quantitative Medicine, Duke-NUS Medical School, 20 College Road, Singapore 169856, Singapore
[c] Center for Child Health Research, University of Tampere and Tampere University Hospital, Arvo Ylpön katu 34, Tampere 33520, Finland
[d] Department of Statistics and Actuarial Science, University of Hong Kong, Pokfulam Road, Hong Kong, China
[e] Faculty of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, Keppel Street, London WC1E 7HT, UK

ARTICLE INFO

ABSTRACT

Background: An effective malaria vaccine affects the risk of malaria directly, through the vaccine-induced immune response (the primary effect), and indirectly, as a consequence of reduced exposure to malaria infection and disease, leading to slower acquisition of natural immunity (the secondary effect). The beneficial primary effect may be offset by a negative secondary effect, resulting in a smaller or nil composite effect. Reports of malaria vaccine trials usually present only the composite effect. We aimed to demonstrate how the primary and secondary effects can also be estimated from trial data.
Methods: We propose an enhancement to the conditional frailty model for the estimation of primary effect using data on disease episodes. We use the Andersen-Gill model to estimate the composite effect. We consider taking the ratio of the hazard ratios to estimate the secondary effect. We used directed acyclic graphs and data from a randomized trial of the RTS,S/AS02 malaria vaccine to illustrate the problems and solutions. Time-varying effects were estimated by partitioning the follow-up into four time periods.
Results: The primary effect estimates from our proposed model were consistently stronger than the conditional frailty model in the existing literature. The primary effect of the vaccine was consistently stronger than the composite effect across all time periods. Both the primary and composite effects were stronger in the first three months, with hazard ratios (95% confidence interval) 0.62 (0.49–0.79) and 0.68 (0.54–0.84), respectively; the hazard ratios weakened over time. The secondary effect appeared mild, with hazard ratio 1.09 (1.02–1.16) in the first three months.
Conclusions: The proposed analytic strategy facilitates a more comprehensive interpretation of trial data on multiple disease episodes. The RTS,S/AS02 vaccine had modest primary and secondary effects that waned over time, but the composite effect in preventing clinical malaria remained positive up to the end of the study.
Clinical trials registration: ClinicalTrials.gov NCT00197041.

1. Introduction

Individuals in malaria-endemic areas gradually build up partial immunity to clinical malaria [1]. Interventions effective in preventing malaria disease episodes also slow the rate of acquisition of natural immunity [2]. Thus, an effective malaria vaccine affects the risk of malaria directly, through the vaccine-induced immune response, and indirectly as a consequence of reduced exposure to malaria infection and disease, leading to slower acquisition of natural immunity. In malaria vaccine trials, this slower acquisition of natural immunity among vaccine recipients could have several consequences. Booster doses could appear to have diminishing effectiveness and, once the vaccine-induced immunity has waned, incidence of malaria in the vaccine group could exceed that in the control group, because of the differences in the level of natural immunity the participants have acquired. Vaccine efficacy may appear lower in areas of higher transmission intensity because participants in the control group acquire natural immunity faster than the

---

participants in the malaria vaccine group and this difference is greater in high transmission areas. Such trends have been observed in trials of the RTS,S/AS01 vaccine [3,4], a pre-erythrocytic vaccine based on *Plasmodium falciparum* circumsporozoite surface antigen, with the AS01 adjuvant. In a phase 2b trial of this vaccine, in a subgroup with high exposure to malaria, after vaccine efficacy waned the incidence of malaria in the vaccine group exceeded that in the control group [3]. A similar pattern was observed in the incidence of severe malaria in a phase 3 trial, where in children who received three doses of the malaria vaccine, the initial reduction in severe malaria cases was offset by a relative increase in incidence after the efficacy of the primary vaccine doses had waned [4]. Vaccine efficacy against clinical malaria in the phase 3 trial was greater in sites in low transmission areas than in high transmission areas, and was lower after the booster dose than after the primary doses [5].

The World Health Organization Malaria Vaccine Advisory Committee recognized the limitation of time-to-first event analyses of malaria vaccine trials and recommended analysis of all events (also called recurrent events) when possible. It also called for further methodological development for the analysis of recurrent events [6]. In the studies of recurrent events, the phenomenon that past disease history may affect one's present risk of having the disease is called "event dependence" [7,8]. A vaccine's effect on the hazard of the disease at time t, $h(t)$, in its recipient via immunogenicity is termed the "primary effect". In contrast, the vaccine's effect on the hazard via its impact on event history between time t_0 and t, W_{t-}, where t_0 denotes the time at initiation of the exposure or intervention, is termed the "secondary effect" [7,8]. The secondary effect can be negative (reflecting reduced acquisition of natural immunity) or positive (when averted episodes, had they occurred, would have made the individual more vulnerable to the disease) [8]. It is plausible that for a malaria vaccine the secondary effect is negative and the composite effect (the net or total effect) is smaller than its primary effect. Concerns that the primary effect of an intervention could be outweighed by the secondary effect have been a major consideration in malaria control [9–11]. Secondary effects are expected to have the greatest impact when the primary effect of the intervention is large, transmission intensity is high, and natural immunity is durable [12,13]. However, the overall public health impact may still often be beneficial [12,14], especially for interventions that improve survival.

We used a directed acyclic graph (DAG) [15,16] to illustrate the issues in the estimation of the vaccine effects in a Cox-type model for recurrent events [7,8,17]. To focus on the core issues at hand, we do not include observed covariates in this DAG. Panel (i) of Fig. 1 depicts the true model (adapted from Fig. 1 of [15]). If an exposure x can affect the hazard of the outcome event at time t, $h(t)$, it would likely affect event history, W_{t-}, as well. Similarly, unobserved frailty (ω), also called omitted variables or heterogeneity, may affect both $h(t)$ and W_{t-}. These causal relationships are indicated by one-headed arrows. The potential of event dependence is indicated by the one-headed arrow from W_{t-} to $h(t)$.

In panel (i), there are three paths that connect x and $h(t)$: (a) $x \to h(t)$; (b) $x \to W_{t-} \to h(t)$, and (c) $x \to W_{t-} \leftarrow \omega \to h(t)$. Note that path (a) represents the primary effect and path (b) represents the secondary effect. In the terminology of DAG, both (a) and (b) are "directed" paths, as shown by the single direction the arrows point to. In contrast, path (c) is a non-directed path, and W_{t-} is a "collider" (being pointed to by two arrows) and ω is a "common ancestor" (where two arrows arise from). Note that a collider status is path-specific: W_{t-} is a collider in path (c) but an intermediate variable in path (b). In order to estimate the primary effect through directed path (a), the estimation needs to condition on W_{t-} (a non-collider) to block the secondary effect through the directed path (b). However, conditioning on a collider opens a path, as opposed to conditioning on a non-collider, which blocks a path [15,16]. Hence, the conditioning on W_{t-} opens the non-directed path (c) and thus generates a bias. Having conditioned on W_{t-} for the purpose of blocking path (b), it is essential to also condition on the non-collider ω in order to block path (c) to obtain unbiased estimate of the primary effect.

However, typical mixed-effects models assume that the observed variates and the frailty term are uncorrelated [18,19]. This contradicts the DAG in panel (i) that W_{t-} is affected by ω. This mis-specified model may be seen as assuming the unobserved frailty as comprising two sub-components, ω_1 and ω_2, as shown in panel (ii) of Fig. 1. Each of them has an arrow pointing to $h(t)$, but only ω_1 has an arrow pointing to W_{t-}. Typical mixed-effects models attempt to control for only ω_2 and therefore does not block the non-directed path (c).

Box-Steffensmeier and De Boef [20] proposed to estimate the primary effect by a conditional frailty (CF) model. This extension of the Cox model for recurrent events conditions on W_{t-} by stratification on the total number of previous episodes, and also conditions on ω by including a frailty term that is assumed to follow a parametric distribution. This formulation does not assume independence between ω and W_{t-}. Its conceptual framework follows the true model in panel (i) of Fig. 1. It properly accounts for the frailty and estimates the primary effect. This is achieved at the expense of not directly modelling event dependence. This method has been applied to the estimation of the primary effect of seasonal malaria chemoprevention [7,21].
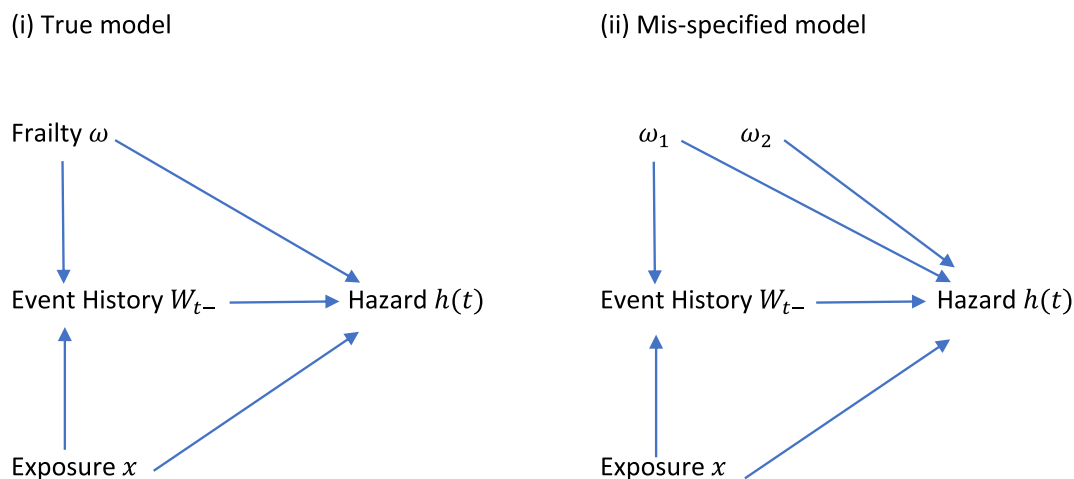
(i) True model

(ii) Mis-specified model

Fig. 1. Directed acyclic graph depicting recurrent time-to-event analysis.

On the other hand, to estimate the composite effect, one should not condition on the intermediate variable W_{t-}. Furthermore, the non-directed path (c) is blocked if the model does not condition on the collider W_{t-} [15,16]. As such, Cheung et al. proposed to use the Andersen-Gill (AG) model to estimate the composite effect [8]. The AG model is another extension of the Cox model for recurrent events [17]. We will provide further details about the AG and CF models in Appendix A.

An assumption of the CF model is that the total number of previous events sufficiently characterizes a person's event history. This is unlikely to be valid in the case of malaria. Partial immunity decays in the absence of exposure [1], and immune responses lapse more rapidly in young children [22]. Earlier exposures to malaria are therefore likely to have less influence on current risk than more recent exposures. Two persons whose time since last event are different are likely to have different levels of immunity and therefore hazard despite the same total number of previous events. Without sufficient control on event history, the CF model may under-estimate the primary effect. Furthermore, path analysis for dependent events that directly partitions the composite effect into primary and secondary effects is feasible only for linear models [23]. There has been little discussion in the literature on how to estimate the secondary effect and its standard error. In this article, we propose a simple modification to the CF model to improve the estimation of primary effect and a method to estimate the secondary effect and its standard error. We illustrate the proposed methods with data from a trial of the RTS,S/AS02 malaria vaccine.

2. Materials and methods

2.1. Design and participants

We use anonymized data from a randomized trial of the RTS,S/AS02 malaria vaccine conducted in Mozambican children to illustrate. The data was kindly made available through GlaxoSmithKline's data sharing platform. Details of the vaccine and study design have been published previously [24–26]. Briefly, the double-blind, randomized controlled trial recruited children aged 1–4 years in a moderate to high transmission area (entomological inoculation rate 38 infective bites per person per year) in southern Mozambique from 2003 to 2004. The RTS,S/AS02 is a pre-erythrocytic vaccine candidate based on *Plasmodium falciparum* circumsporozoite surface antigen, with the AS02 adjuvant. Episodes of clinical *Plasmodium falciparum* malaria were defined by axillary temperature $\geq$37.5 °C and *P falciparum* asexual parasitaemia >2500 per μL. After each episode of clinical malaria, a child was considered not susceptible for malaria for 28 days. After receiving malaria drug treatment, a child was considered not susceptible for 7–28 days, depending on which drug was taken. The number of children randomized to the control vaccine and malaria vaccine groups were 745 each. The analysis time started from 14 days after dose three of the vaccines. For brevity, we called this time since vaccination. The maximum analysis time per person was 42 months since vaccination. The original trial analysis adjusted for age at baseline, bednet use at baseline, distance from health facility, and geographical region as covariates. The analysis here included the same covariates in all models.

2.2. Statistical models

2.2.1. Andersen-Gill and conditional frailty models

Details of the AG and CF models are provided in Appendix A. Let β_C denote the log hazard ratio (HR) in the AG model and β_P denote the log HR in the CF model with stratification of the total number of previous events, as proposed by Box-Steffensmeier and De Boef

[20]. The subscripts C and P stand for composite and primary effects, respectively. The log HR estimates can be transformed back to the hazard ratios, HR_C and HR_P, respectively.

To remove the CF model's assumption that the persons within a stratum defined by number of previous events are homogeneous in risk level, we propose to further stratify within each stratum according to time since the latest event. Previous researchers have suggested that the use of tertiles is usually sufficient to control confounding [27]. Therefore, we consider tertiles. In the stratum of observations with k previous events, we calculate the tertiles of time between the kth and $(k + 1)^{th}$ observed events. A stratum in the CF model becomes three strata in the modified CF model. Let β_P^* and HR_P^* denote the log HR and HR in the modified CF model.

The AG model can be estimated by the *stcox* program in Stata. The CF and modified CF models can be estimated by the *strmcure* program in Stata [28].

To avoid data sparsity, it is advisable to pool the last few strata that have small number of observations [28]. In the analysis of this dataset, in both the CF and modified CF models we pooled the observations for times to the seventh or later events into one single stratum.

Previous researchers suggested that the duration of protection of the vaccine may be as short as only three months [29]. We partitioned the time since vaccination into $\leq$3, 3–12, 12–24, and >24 months and estimated the time-varying effects.

2.2.2. Estimation of secondary effect

While the importance of secondary effects is recognized [12,14,30], there has been little discussion in the literature on how to estimate it. In the statistics literature, it is known that except for linear models, there is no single model that can directly partition the composite effect into primary and secondary effects [23]. Assuming a multiplicative model for the hazard of recurrent events, $HR_C = HR_P^* \times HR_S$, or equivalently $\beta_C = \beta_P^* + \beta_S$, we propose to estimate the secondary effect by $\widehat{\beta}_S = \widehat{\beta}_C - \widehat{\beta}_P^*$, where $\widehat{\beta}_C$ and $\widehat{\beta}_P^*$ are the estimates obtained from fitting the AG and modified CF models, respectively. An alternative estimate of secondary effect based on $\widehat{\beta}_P$ can be obtained similarly but, for brevity, we do not discuss it in this manuscript.

We used the bootstrapping method, with persons as the re-sampling units, and 200 replicates to obtain 95% confidence intervals (CI) of all the log HR estimates. Within each bootstrap sample, we fitted two models to estimate the primary and composite effect log HR's and then took their difference to obtain the secondary effect log HR. The results were transformed back to give the 95% CI of HR's.

3. Results

The total child-months in the control and malaria vaccine groups were 25,687 and 26,307, respectively. The number of clinical malaria episodes were 774 and 658 in the control and malaria vaccine group, respectively (Table 1). Without adjustment for covariates and variation in follow-up time, a Mann-Whitney *U* test showed a statistically significant difference between the two groups (P = 0.002).

Table 2 summarizes the distribution of time since the latest event, by event order. The median time to first event was 7.7 months (33.3th and 66.7th percentiles 4.1 and 18.1 months, respectively). Except some minor irregularity, there was a trend that the more events a child had experienced, the shorter the time to the next event.

Table 3 shows the estimates of composite, primary and secondary effects based on the AG, CF and modified CF models. HR_C did not show a monotonic trend over time. It was strongest in

Table 1
Number (%) of children in the control and malaria vaccine groups, by total number of clinical malaria episodes during the study follow-up period.

No. of malaria episodes	Control (n = 745)	Malaria vaccine (n = 745)
0	375 (50.3)	438 (58.8)
1	185 (24.8)	157 (21.1)
2	85 (11.4)	67 (9.0)
3	49 (6.6)	32 (4.3)
4	21 (2.8)	18 (2.4)
5	7 (0.9)	15 (2.0)
6	16 (2.2)	7 (0.9)
7	2 (0.3)	7 (0.9)
8	3 (0.4)	3 (0.4)
9	1 (0.1)	1 (0.1)
10	1 (0.1)	0 (0)

the first three months, at 0.68. It then weakened to 0.81 in the 3–12 month interval, rebounded to 0.77 in the 12–24 month interval, and weakened again to 0.86 afterward.

HR_P from the first to the fourth time intervals were 0.64, 0.76, 0.76 and 0.84, respectively. In all time intervals, HR_P^* consistently showed a stronger reduction in hazard than HR_P.

In the first three months, HR_S was 1.09. Then it weakened over time, with 95% CI including the null value.

4. Discussion

In the phase 3 trial of the RTS,S/AS01 vaccine, vaccine efficacy varied with transmission intensity, with lower efficacy in sites with higher intensity [4]. A similar pattern has been observed in rotavirus vaccine [31]. The interpretation of this pattern may affect whether the products are deployed where they are most needed. One possible interpretation is that the primary effect of the intervention is actually constant, but the control groups in areas with high intensity acquired immunity more rapidly than the intervention groups in the same areas. Therefore, the composite effect may be smaller where disease burden is high. It is likely that malaria vaccines will require booster doses; the need for a booster and the timing should be based on primary effects rather than composite effects.

There has been an emerging consensus on the use of the AG model to estimate the composite effect [7,8,32]. But the estimation of the primary and secondary effects has been less discussed. With

unobserved frailty, the estimation is not straight-forward. The CF model provides a valid framework for the estimation of the primary effect [20]. Nevertheless, its accuracy may be affected by the assumption that, having conditioned on the total number of previous events, time since last event does not affect the outcome. We have proposed a modification to the CF model, by further stratification for time since last event.

Applying this approach to the RTS,S/AS02 trial data, the time-varying composite effect, HR_C, was strongest in the first three months as expected [29]. But the dip in the 3–12 month interval was unexpected. In contrast, the primary effect estimates showed smoother trajectory over the time periods. If there is a secondary effect, we would expect it to be relatively strong around the time the primary effect is strong. The dip in HR_C in the 3–12 month period appears to be the result of the secondary effect. The median (33th, 67th percentiles) time to first event was 7.7 (4.1, 18.1), as shown in Table 2. Hence the secondary effect may have less impact in the first time period than in the second. Across the whole study duration, the secondary effect, HR_S, was mild. It is not surprising, given that the primary effect of the RTS,S/AS02 vaccine was modest to begin with. The secondary effect could be stronger if the primary effect is strong and the disease incidence is high.

In the analysis of the RTS,S/AS02 trial data, the CF and modified CF models did not give results that differ hugely (Table 3). Nevertheless, they consistently differed in the expected direction, with stronger estimate of the primary effect with control of time since last event. In this dataset, as shown in Table 2, the number of previous events was associated with the time to the next event. As such, stratification for the former also indirectly stratified for the latter. This explained the lack of major difference between them. Nevertheless, the two methods may give larger difference in the evaluation of other infectious disease interventions or under other disease patterns. Since waning of immunity over time is a common phenomenon, the proposed method is useful in the studies of many other infectious diseases and prevention measures.

Although finer stratification on event history may more properly remove the influence of event history in the estimation of primary effect, caution is needed to avoid data sparsity. In Cox-type models, having only a small number of observations in a stratum may lead to a situation that nobody other than the person who has an event is at risk at that event time. This would cause an empty risk-set and exclusion of that event from the analysis. The higher the number of previous events, the more likely this problem

Table 2
Percentiles of time since last event (in months).

No. of malaria episodes	No. of children	33-th percent	Median	67-th percent
1	677	4.1	7.7	18.1
2	335	3.0	5.8	9.5
3	183	1.8	3.5	6.2
4	102	1.6	2.9	4.2
5	63	0.8	1.8	3.2
6	41	1.2	2.6	4.4
≥7	18	0.6	1.8	3.7

Table 3
Estimates of composite, primary and secondary effect *.

Target	Estimates	≤3	mo	3 to ≤	12 mo	12 to	≤24 mo	≥24	mo
Effects		HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI
Composite	HR_C	0.68	(0.54, 0.84)	0.81	(0.66, 0.99)	0.77	(0.65, 0.92)	0.86	(0.72, 1.02)
Primary	HR_P	0.64	(0.51, 0.80)	0.76	(0.61, 0.95)	0.76	(0.63, 0.93)	0.84	(0.69, 1.02)
Primary	HR_P^*	0.62	(0.49, 0.79)	0.75	(0.60, 0.95)	0.74	(0.60, 0.91)	0.83	(0.68, 1.02)
Secondary	HR_S	1.09	(1.02, 1.16)	1.07	(1.00, 1.16)	1.05	(0.96, 1.14)	1.04	(0.96, 1.13)

* All models adjusted for age at baseline, bednet use at baseline, distance from health facility, and geographical region.

will occur. For an intervention that is highly efficacious in reducing event rates, this means the excluded events are mainly from the non-intervention group. This may cause a bias towards underestimation of the vaccine effects. In our application of the CF model, two events were excluded due to this reason. Both occurred in the stratum for time to the sixth event; one out of two was in the control group. In the modified CF model, six events were excluded. They occurred in the strata for times to the fourth, fifth or sixth events; four out of six were in the control group. With such a small number of exclusion (out of 1432 events) and weak efficacy of the RTS,S/AS02 vaccine, the impact would be tiny. But this needs careful evaluation according to each study's data pattern; pooling of strata may be needed [28]. This was also the reason of our using the tertiles instead of finer stratification in the modified CF model.

We used time since vaccination as the time-scale. In the analysis of recurrent event, some investigators may use time since last event as the time-scale, with the exception that for the first event the time-scale is time since study enrolment [21,33]. This formulation re-sets the time to zero after each event. This time-scale cannot be justified if the model does not stratify for the number of previous events, because in that case a person may become at risk of the $(k + 1)^{th}$ event before having experienced the kth or even earlier events [33,34]. Therefore it cannot be used in the estimation of the composite effect. In order to make comparison between the estimates of the composite and primary effects, this time-scale should not be used in the estimation of the primary effect either.

5. Conclusions

The proposed analytic strategy can be used to estimate the primary, secondary and composite effects of an intervention. It offers a more comprehensive understanding of trial data on all disease episodes. The RTS,S/AS02 vaccine had moderate primary and secondary effects that waned over time, but the composite effect in preventing clinical malaria remained positive up to the end of the trial at 42 months.

Declarations

Ethics approval and consent to participate

This is an analysis of anonymized human trial data. This is approved by the National University of Singapore Institutional Review Board (B-16-064E). The participants' parent or guardian gave informed consent when they participated in the trial.

Consent for publication

Not applicable.

Availability of data and materials

The data was accessed under a data sharing agreement with GlaxoSmithKline. The dataset is not publicly available due to ownership not belonging to the authors. The data are available from the first author upon reasonable request.

Funding

CRediT authorship contribution statement

Yin Bun Cheung: Conceptualization, Methodology, Writing - original draft, Supervision. **Xiangmei M:** Formal analysis, Data curation, Writing - review & editing. **K.F. Lam:** Conceptualization, Methodology, Writing - review & editing. **Paul Milligan:** Conceptualization, Methodology, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Appendix A. Statistical models

Andersen-Gill model

For individual $i (i = 1, 2, \cdots, N)$, we observed $n_i - 1$ event times $\{t_{ij}, j = 1, 2, \cdots, n_i - 1\}$ with the event indicators $\delta_{ij} = 1; j = 1, 2, \cdots, n_i - 1$. The individual exits the study at time τ_i with no event occurred at the time $\tau_i (t_{in_i} = \tau_i, \delta_{in_i} = 0)$. That is, the last spell represented censoring. In the AG model, the hazard at time t for individual i with a binary key exposure variable x_i and covariate vector z_i is modelled as

$$h_i(t|Y_i(t), x_i, z_i) = Y_i(t)h_0(t) \exp(\beta_C x_i + \gamma_C z_i), \qquad (1)$$

where $h_0(t)$ is an unspecified baseline hazard function and $Y_i(\cdot)$ is the at-risk function for individual i [8,17]. In the present context, $x_i = 0$ and $x_i = 1$ indicates a person in the control and malaria vaccine group, respectively. The partial likelihood function is:

$$L(\beta_C, \gamma_C) = \prod_{i=1}^{N} \prod_{j=1}^{n_i-1} \frac{\exp(\beta_C x_i + \gamma_C z_i)}{\sum_{k=1}^{N} Y_k(t_{ij}) \exp(\beta_C x_k + \gamma_C z_k)}. \qquad (2)$$

Conditional frailty model

In the CF model [20], the hazard function of the jth event $(j = 1, 2, \cdots, n_i - 1)$ occurring at time t for individual i with the binary key exposure variable x_i and covariate vector z_i is:

$$h_{ij}(t|Y_{ij}(t), \omega_i, x_i, z_i) = Y_{ij}(t)\omega_i h_{0j}(t) \exp(\beta_P x_i + \gamma_P z_i) \qquad (3)$$

where $Y_{ij}(\cdot)$ is the at-risk process specific for individual i and event order j, that is, $Y_{ij}(t) = 1$ if at time t individual i has experienced $(j - 1)^{th}$ event and is at risk for the jth event, and $Y_{ij}(t) = 0$ otherwise. $h_{0j}(t)$ is an unspecified baseline hazard function for the jth event. The frailty term ω_i is assumed uncorrelated with x_i and z_i and independently and identically follows a gamma distribution with $E(\omega_i) = 1$ and $Var(\omega_i) = \psi$. The partial likelihood function for the model is:

$$L(\beta_P, \gamma_P | \omega_i, i = 1, \cdots N) = \prod_{i=1}^{N} \prod_{j=1}^{n_i-1}$$
$$\times \frac{\omega_i \exp(\beta_P x_i + \gamma_P z_i)}{\sum_{i'=1}^{N} Y_{i'j}(t_{ij}) \omega_{i'} \exp(\beta_P x_{i'} + \gamma_P z_{i'})}. \qquad (4)$$

For higher value of j, the number of persons who are at risk of the jth event may be small. To avoid sparse data within some strata, the last few strata may be pooled as one stratum.

Modified conditional frailty model

The CF model assumes that the persons within a stratum defined by number of previous events are homogeneous in risk level. To remove this assumption, we propose to further stratify within each stratum according to time since the latest event. Specify a set of cut-off values $\min_i(t_{ij}) = \zeta_{j0} < \zeta_{j1} < \cdots < \zeta_{j,K(j)-1} < \zeta_{j,K(j)} = \max_i(t_{ij})$ that partitions the time since the $(j-1)^{th}$ event into K(j) intervals, regardless of key exposure and covariate values. We consider tertiles and set K(j) = 3 for all j with one-third of observations within each of the three intervals ζ_{j0} to ζ_{j1}, ζ_{j1} to ζ_{j2}, and ζ_{j2} to ζ_{j3}. The model becomes

$$h_{ijk}(t|Y_{ijk}(t), \omega_i, x_i, z_i) = Y_{ijk}(t)\omega_i h_{0jk}(t)\exp(\beta_P^* x_i + \gamma_P^* z_i) \qquad (5)$$

where $Y_{ijk}(\cdot)$ and $h_{0jk}(t)$ are the at-risk indicator and unspecified baseline hazard function for individual i, event order j and time since latest event interval k. The partial likelihood function becomes

$$L(\beta_P^*, \gamma_P^*|\omega_i, i=1, \cdots N) = \prod_{i=1}^{N}\prod_{j=1}^{n_i-1}$$
$$\times \prod_{k=1}^{K(j)}\left[\frac{\omega_i \exp\left(\beta_P^* x_i + \gamma_P^* z_i\right)}{\sum_{i'=1}^{N} Y_{i'jk}(t_{ij})\omega_{i'} \exp\left(\beta_P^* x_{i'} + \gamma_P^* z_{i'}\right)}\right]^{\delta_{ijk}^*},$$
$$\qquad (7)$$

where $\delta_{ijk}^* = I(\zeta_{j,k-1} < t_{ij} \leq \zeta_{jk})$ and, specifically, $\delta_{ij1} = 1$ if $t_{ij} = \zeta_{j0}$.

Coefficients and estimation

The coefficients β_C in the AG model, β_P in the CF and β_P^* in the modified CF models are the log hazard ratios (HR) representing the estimates of total and primary effects, respectively. The AG model is estimated by maximizing the partial likelihood function by the Newton-Raphson algorithm [35] and the CF and modified CF models by the expectation-maximization (EM) algorithm via the *strmcure* macro in Stata [28]. Assuming a multiplicative model for the hazard of recurrent events, $HR_C = HR_P^* \times HR_S$, or equivalently $\beta_C = \beta_P^* + \beta_S$, we estimate the secondary effect by $\widehat{\beta}_S = \widehat{\beta}_C - \widehat{\beta}_P^*$, where $\widehat{\beta}_C$ and $\widehat{\beta}_P^*$ are the estimates obtained from fitting the AG and modified CF models, respectively. Confidence intervals are obtained by bootstrapping, with persons as the re-sampling units. Within each bootstrap sample, we fitted two models to estimate the primary and composite effect log HR's and then take their difference to obtain the secondary effect log HR. The 95% confidence intervals are then transformed back to the HR scale.

References

[1] Doolan DL, Dobano C, Baird JK. Acquired immunity to malaria. Clin Microbiol Rev 2009;22:13–36.
[2] Dicko A, Greenwood BM. Malaria vaccination and rebound malaria. *Lancet Infect Dis* 2019;19(8):790–1.
[3] Olotu A, Fegan G, Wambua J, et al. Seven-year efficacy of RTS, S/AS01 malaria vaccine among young African children. N Engl J Med 2016;374:2519–29.
[4] RTS, S Clinical Trials Partnership. Efficacy and safety of RTS,S/AS01 malaria vaccine with or without a booster dose in infants and children in Africa: Final results of a phase 3, individually randomized, controlled trial. Lancet 2015;386 (9988):31–45.
[5] Joint Technical Expert Group on Malaria Vaccines and World Health Organization Secretariat. Background Paper on the RTS,S/AS01 Malaria Vaccine. Geneva: World Health Organisation; 2015.
[6] Moorthy VS, Reed Z, Smith PG. WHO Malaria Vaccine Advisory Committee. MALVAC 2008: Measures of efficacy of malaria vaccines in phase 2b and phase 3 trials–scientific, regulatory and public health perspectives. Vaccine 2009;27 (5):624–8.
[7] Cairns M, Cheung YB, Xu T, et al. Analysis of malaria cohort studies: exploring partial and complete protection, and total and primary intervention effects. Am J Epidemiol 2015;181:1008–17.
[8] Cheung YB, Xu Y, Tan SH, Cutts F, Milligan P. Estimation of intervention effects using first or multiple episodes in clinical trials: The Andersen-Gill model re-examined. Stat Med 2010;29:328–36.
[9] Snow RW, Marsh K. Will reducing Plasmodium falciparum transmission alter malaria mortality among African children?. Parasitol Today 1995;11:188–90.
[10] Smith TA, Leuenberger R, Lengeler C. Child mortality and malaria transmission intensity in Africa. Trends Parasitol 2001;17:145–9.
[11] Nahlen BL, Clark JP, Alnwick D. Insecticide-treated bednets. Am J Trop Med Hyg 2003;68(4 Suppl):1–2.
[12] Pemberton-Ross P, Smith TA, Hodel EM, et al. Age-shifting in malaria incidence as a result of induced immunological deficit: a simulation study. Malar J 2015;14:287.
[13] Milligan PJM, Downham DY. Models of superinfection and acquired immunity to multiple parasite strains. J Appl Probability 1996;33:915–32.
[14] Greenwood BM, David PH, Otoo-Forbes LN, et al. Mortality and morbidity from malaria after stopping malaria chemoprophylaxis. Trans R Soc Trop Med Hyg 1995;89:629–33.
[15] Jahn-Eimermacher, Ingel K, Preussler S, et al. A DAG-based comparison of intervention effect underestimation between composite endpoint and multi-state analysis in cardiovascular trials. BMC Med Res Method 2017;17(1):92.
[16] Rothman KJ, Greenland S, Lash TL. Modern epidemiology. 3rd ed. Philadelphia, PA: Lippincott, Williams & Wilkins; 2008.
[17] Therneau TM, Grambsch PM. *Modeling survival data: extendingthe cox model*. New York: Springer; 2000.
[18] Allison P. Fixed-effects regression models. Thousand Oaks: Sage; 2011.
[19] Hsiao C. Analysis of panel data. Cambridge: Cambridge U.P; 1986.
[20] Box-Steffensmeier JM, De Boef S. Repeated events survival models: the conditional frailty model. Stat Med 2006;25(20):3518–33.
[21] Xu Y, Cheung YB, Lam KF, Milligan P. Estimation of summary protective efficacy using a frailty mixture model for recurrent event time data. Stat Med 2012;31:4023–39.
[22] Akpogheneta OJ, Duah NO, Tetteh KK, et al. Duration of naturally acquired antibody responses to blood-stage *Plasmodium falciparum* is age dependent and antigen specific. Infect Immun 2008;76:1748–55.
[23] Fosen J, Borgan O, Weedon-Fekjaer H, Aalen OO. Dynamic analysis of recurrent event data using the additive hazard model. Biom J 2006;48:381–98.
[24] Alonso PL, Sacarlal J, Aponte JJ, et al. Efficacy of the RTS, S/AS02A vaccine against Plasmodium falciparum infection and disease in young African children: randomised controlled trial. Lancet 2004;364:1411–20.
[25] Alonso PL, Sacarlal J, Aponte JJ, et al. Duration of protection with RTS, S/AS02A malaria vaccine in prevention of Plasmodium falciparum disease in Mozambican children: single-blind extended follow-up of a randomised controlled trial. Lancet 2005;366(9502):2012–8.
[26] Sacarlal J, Aide P, Aponte JJ, et al. Long-term safety and efficacy of the RTS, S/ AS02A malaria vaccine in Mozambican children. J Infect Dis 2009;200 (3):329–36.
[27] Clayton D, Hills M. Statistical models in epidemiology. Oxford: Oxford U.P; 1993.
[28] Xu Y, Cheung YB. Frailty models and frailty mixture models for recurrent event times. Stata J 2015;15:135–54.
[29] Smith P, Milligan P. Malaria vaccine: 3 or 6 months' protection?. Lancet 2005;365(9458):472–3.
[30] Penny MA, Maire N, Studer A, et al. What should vaccine developers ask? Simulation of the effectiveness of malaria vaccines. PLoS ONE 2008;3:e3193.
[31] Armah GE, Sow SO, Breiman RF, et al. Efficacy of pentavalent rotavirus vaccine against severe rotavirus gastroenteritis in infants in developing countries in sub-Saharan Africa: a randomised, double-blind, placebo-controlled trial. Lancet 2010;376(9741):606–14.
[32] Rauch G, Kieser M, Binder H, Bayes-Genis A, Jahn-Eimermacher A. Time-to-first-event versus recurrent-event analysis: points to consider for selecting a meaningful analysis strategy in clinical trials with composite endpoints. Clin Res Cardiol 2018 May;107(5):437–43.
[33] Kelly PJ, Lim LL. Survival analysis for recurrent event data: an application to childhood infectious diseases. Stat Med 2000;19:13–33.
[34] Metcalfe C, Thompson SG. Wei, Lin and Weissfeld's marginal analysis of multivariate failure time data: should it be applied to a recurrent events outcome?. Stat Methods Med Res 2007;16:103–22.
[35] Gould W, Pitblado J, Poi B. Maximum likelihood estimation with stata, 4th ed. College Station, TX: StataCorp; 2010.