JOINT MODELLING OF RECURRENT CLINICAL MALARIA EPISODES AND LONGITUDINAL PARASITEMIA DATA

BY

CHRISTOPHER CHIKHOSI STANLEY (Student Number: 1482592)

PhD Thesis Abstract

Background: Plasmodium falciparum infection is one of the most common parasitic infections in humans, with a disproportionately higher share of the problem in Sub-Saharan Africa (SSA). In Malawi, malaria is endemic and the prevalence of *Plasmodium falciparum* parasite is also high. Individuals in high-transmission settings are exposed to bites of infected Anopheles mosquitoes and develop frequent infections. With each infection, acquired immunity develops making subsequent disease episodes less likely. Therefore, longitudinal cohorts are ideal to study risk of malaria over time especially in high transmission areas because repeated infection is common. When modelling the risk of malaria in a cohort study, methods ought to account for this naturally acquired immunity which develops in individuals over time. Methods that focus on first incident alone (single-event joint model) excluding subsequent episodes may be inefficient since much information is lost and may underestimate the disease burden. A gap exists in literature on the use of joint modeling to analyze recurrent clinical malaria and parasitemia data collected during cohort studies.

Aim: To determine factors for recurrent malaria episodes in Malawi's Mfera cohort study through joint modelling of the recurrent episodes and longitudinal parasitemia data.

Methods: In order to investigate the trends of use of appropriate statistical methods among articles analyzing prospective longitudinal cohort data for clinical malaria episodes or Plasmodium infection published from 1996 until December 2017, a review of statistical methods was conducted. Original articles in English, published over the 22-year period were reviewed in chronological order. Particular interest was on standard longitudinal analysis techniques including generalized estimating equations (GEE) and mixed-effects models which are considered appropriate for analyses of repeatedly measured data. For repeated time-to-episodes, recurrent Cox proportional hazards (PH)-based models such as frailty models, Andersen-Gill (AG), Prentice and Williams and Peterson (PWP), and the marginal means/rates models were considered appropriate. Other survival methods such as the Kaplan–Meier estimator, log-rank test and Cox PH model were considered only appropriate for analyses of time-to-single-episode.

Using a cohort data of participants enrolled with uncomplicated malaria in Malawi, a conventional Cox Proportional Hazards (PH) model of time-to-first clinical malaria episode with time-dependent parasitemia was compared with three competing joint models. The cohort enrolled 120 participants who presented with uncomplicated malaria at the Mfera health center in the Chikwawa district between June 2014 and March 2015. Participants underwent passive and active surveillance on a monthly basis and whenever sick for up to two years to assess the incidence of clinical malaria and *P. falciparum* infection. Four models of time-to-new clinical malaria disease were fitted. These included the reference model (model 1), a conventional Cox PH model fitted with time-dependent parasitemia measured at each visit, and three competing

joint models. The three joint models had a common formulation only differing in the assumed association structure linking the hazard of clinical malaria with parasitemia. The hazard of clinical malaria disease at any time t was linked with: 1) the current underlying value of parasitemia at the same time point (model 2), 2) the rate of change in parasitemia trajectory at time t (model 3), and 3) the cumulative trajectory or area under profile of parasitemia from baseline up to time t (model 4). Each joint model was formed of a mixed-effects sub-model of parasitemia longitudinal and Cox PH sub-model for time-to-first clinical malaria episode. Parameters were estimated using Markov chain Monte Carlo (MCMC) through the random walk Metropolis–Hastings (M-H) algorithm. Diffused normal priors were assumed for the covariates. Diagnostic assessments were conducted to assess the convergence of the MCMC samples using trace and kernel density estimator plots.

To assess whether jointly modelling recurrent clinical malaria episodes and parasitemia is more accurate than single-event joint models where the subsequent episodes are ignored, the joint model was applied to a real dataset from the Mfera malaria cohort. The models fitted were shared Gamma frailty model for the recurrent events and the mixed-effects model taking competing distributions for the longitudinal process: Poisson, quasi-Poisson and negative binomial (NB). Model fit was compared based on the Akaike information criterion (AIC). A model with lowest value of AIC is considered to be the best fitting model. In order to assess how the optimal recurrent event joint model can perform under different conditions, a simulation study was also conducted. In the simulations, data were generated with parameters that resemble the Mfera cohort. Age in years was assumed to be normally distributed on the log-scale, and values were transformed back to original scales by taking an exponential function to maintain the

skewness that would reflect real data. The baseline hazard function was assumed to follow a Weibull distribution. Participants were assumed to be censored according to a uniform distribution on time of follow-up. Parasitemia data measurements were simulated from a mixed effects model with function of follow-up time. The model performance in terms of bias was assessed under different scenarios of: study sample sizes; level of censoring; length of follow-up period; Gamma distributed frailty term variances; and correlation level of association between longitudinal and recurrent processes.

Results: In the statistical review, 197 (10.0%) articles met the inclusion criteria and were included in the analyses. The number of articles increased over the 22-year study period, with the highest number found between 2012 and 2016. Out of 197 articles reviewed, the most commonly reported methods included contingency tables which comprised Pearson Chi-square, Fisher exact and McNemar's tests (n = 102, 51.8%), Student's t-tests (n = 82, 41.6%), followed by Cox PH models (n = 62, 31.5%) and Kaplan–Meier estimators (n = 59, 30.0%). The longitudinal analysis methods GEE and mixed-effects models were reported in 41 (20.8%) and 24 (12.2%) articles, respectively, and increased in use over time. Six articles analyzed repeated time-to-episodes using appropriate recurrent models: Andersen-Gill (AG) (4); and frailty models (2).

There were 120 participants in the Mfera cohort, of which 69 (57.5%) were females. The overall median age was 7.5 years [inter-quartile range (IQR): 4.7 – 18.1], 6.3 years (IQR: 3.2 – 13.1) for males and 9.2 years (IQR: 5.3 – 18.5) females. Among the 115 participants who had at least one follow-up visit post enrolment included in the survival analyses, 372 recurrent clinical malaria

episodes were experienced over the two-year follow-up period, of which 219 (58.9%) occurred among females. Out of the 372 episodes, 125 (33.6%) were among children under 5 years, 193 (51.9%) in school aged children of 5–15 years and 54 (14.5%) were experienced by adults above 15 years old. Overall, there was a decreasing trend in number of recurrent clinical malaria episodes during follow-up from 106 episodes in first month to 75 episodes in 24^{th} month, for a drop of 3 participants in the period. In terms of precision, the recurrent events joint model outperformed single event joint model which underestimated the hazard ratios of clinical malaria in most cases with larger standard errors. Based on the recurrent event joint model, the hazard of recurrent clinical malaria decreased with increasing participants' age (hazard ratio [HR]=0.960 [95% confidence interval [CI]: 0.944, 0.976]), for 1-year increase in age. The hazard was also higher among participants who reported not to use LLINs every night compared to those who used the nets nightly (HR=1.424, [95% CI: 1.218, 2.032]). Compared to dry season, the hazard was higher during the rainy season (HR=1.355, [95% CI: 1.048, 1.753]).

In the simulation study, the performance of the recurrent event joint model improved as measured from decreasing bias with increased study sample size overall from 100, 200 to 400 participants, and by increasing length of study follow-up from 2 years to 3 and 4 years. The performance of the joint model worsened by increasing level of censoring from 10% to 20% and 50% in that order worsened the performance of the joint model.

Conclusions: Despite similar study designs across the reports, the statistical methods varied substantially and often represented overly simplistic models of risk. The results underscore the

need for more effort to be channeled towards adopting standardized longitudinal methods to analyze prospective cohort studies of malaria infection and disease.

Modelling the risk of clinical malaria using the recurrent event joint models with improved precision may enhance study efficiency and allow for design of clinical trials with relatively lower sample sizes with increased power. Underestimation of risk by the single event joint model may lead to incorrect inferences and wrong conclusions. The simulation study show that the recurrent events joint model can give parameter estimates with low bias. These results underline the need for models to utilize all collected follow-up data when estimating the risk of recurrent clinical malaria.

Keywords: Plasmodium falciparum, Longitudinal studies, Longitudinal analysis, Cohort studies, Prospective studies, longitudinal data, malaria parasitemia, time-to-event analysis, clinical malaria, Cox proportional hazards model, joint modelling, recurrent events, shared frailty model