Joint Modeling of Mixed *Plasmodium* Species Infections Using a Bivariate Poisson Lognormal Model

Kathryn L. Colborn,^{1*} Ivo Mueller,² and Terence P. Speed^{2,3,4}

¹Department of Biostatistics and Informatics, Colorado School of Public Health, University of Colorado Denver, Aurora, Colorado;

²Walter and Eliza Hall Institute, Melbourne, Australia; ³Department of Mathematics and Statistics, University of Melbourne, Melbourne, Australia; ⁴Department of Statistics, University of California, Berkeley, Berkeley, California

Abstract. Infectious diseases often present as coinfections that may affect each other in positive or negative ways. Understanding the relationship between two coinfecting pathogens is thus important to understand the risk of infection and burden of disease caused by each pathogen. Although coinfections with *Plasmodium falciparum* and *Plasmodium vivax* are very common outside Africa, it is yet unclear whether infections by the two parasite species are positively associated or if infection by one parasite suppresses the other. In this study, we use bivariate Poisson lognormal models (BPLM) to estimate covariate-adjusted associations between the incidence of infections (as measured by the force of blood-stage infections, molFOB) and clinical episodes caused by both *P. falciparum* and *P. vivax* in a cohort of Papua New Guinean children. A BPLM permits estimation of either positive or negative correlation, unlike most other multivariate Poisson models. Our results demonstrated a moderately positive association between *P. falciparum* and *P. vivax* infection rates, arguing against the hypothesis that *P. vivax* infections protect against *P. falciparum* infections. Our findings also suggest that the BPLM is only useful for counts with suitably large means and overdispersion.

INTRODUCTION

In research focused on infectious diseases, it is common to observe infection with a particular pathogen accompanied by infection with a second type of pathogen or a different species of the same pathogen (e.g., multiple dengue serotypes, human immunodeficiency virus (HIV) and hepatitis viruses, mixed bacterial and viral infections in children). Some typical approaches to estimating the risk of concurrent infections include simple χ^2 tests of association, correlation coefficients, linear models with one infection as the response and the other as the predictor, and multinomial response models.^{1–3} McCulloch provides a nice summary of the advantages of joint modeling of responses, such as the ability to estimate the overall effects of important covariates on multiple outcomes, avoidance of multiple testing, efficiency gains and estimation of association between outcomes.⁴ This last point, estimating a measure of association between two responses, was one of the primary objectives of a cohort study conducted in Papua New Guinea (PNG) investigating the effects of single and mixed species malaria infections in children.

In PNG, children are perennially exposed to both *Plasmodium falciparum* and *Plasmodium vivax* malaria parasites.⁵ Coinfections occur because children are repeatedly bitten by different mosquitoes carrying one or both of the parasites. It is unclear how the two species interact within the body of the infected child. Evidence from several studies comparing the frequency of infection in clinical and asymptomatic cases suggests that perhaps *P. vivax*, which typically causes less severe symptoms of malaria, suppresses *P. falciparum*, the parasite that often results in more severe symptoms or even death.^{6–8} However, more recent studies have shown that individuals infected with both species of malaria parasites experience more severe symptoms than those with single species infections.^{9–11} The PNG cohort study was designed to collect

information that would help investigators better understand the relationship between these two parasite species.¹²

In this study, we investigated two measures of malaria burden from the PNG study to motivate the use of bivariate models for estimating association between two coinfecting pathogens. The first measure of burden is the number of *P. falciparum* and *P. vivax* clinical episodes experienced over a period of time. The second measure of burden is the number of newly acquired genetically unique malaria infections over a period of observation, referred to as the molecular force of blood-stage infection (molFOB).^{13,14} We will relate these measures in separate bivariate models, comparing the counts for each of the two species over an interval of observation. We chose to fit a bivariate Poisson lognormal model (BPLM)¹⁵ to estimate a measure of association between these variables because it permits negative correlation and allows for inclusion of other important risk factors. There are several good examples of using a multivariate Poisson lognormal model to jointly model crash frequency and severity data in the literature.^{16–18} A BPLM has also been applied to model spatial and temporal fluctuations in the community structure of tropical butterflies.¹⁹

We begin by describing the PNG cohort. Next, we describe the BPLM. We then present the results of fitting a BPLM to clinical episodes and molFOB using two different methods. The first method uses Bayesian inference via Markov chain Monte Carlo (MCMC).²⁰ The second method uses maximum likelihood estimation (MLE) to fit the BPLM. The two separate methods are presented because when fitting the BPLM to clinical episodes, which were very small counts, we encountered numerical instability using the MLE. We conclude with a discussion of the results and provide suggestions for applying this model in practice.

MATERIALS AND METHODS

Study population. Details of the PNG cohort study have been provided elsewhere.^{12–14} Briefly, 264 children (1–3 years of age at enrollment) were examined for malaria infection every 8 weeks by fingerprick blood sampling and rapid diagnostic testing (RDT). Active case detection was carried out every 2 weeks, at which time children were examined for symptoms

* Address correspondence to Kathryn L. Colborn, Department of Biostatistics and Informatics, Colorado School of Public Health, University of Colorado Denver, Building 500, Room C3011, 13001 East 17th Place, Aurora, CO 80045. E-mail: kathryn.colborn@ucdenver.edu

of acute malaria infection, and blood samples were taken from children with a fever in the past 48 hours. Blood samples were also taken from all symptomatic children enrolled in the study who presented at clinics between regular visits. Antimalarial treatment was administered in RDT-positive cases with fever or if hemoglobin levels were less than 7.5 g/dL.

Each febrile illness that was accompanied by blood samples containing *P. falciparum* parasites (> 2,500 parasites/μL) and/or *P. vivax* parasites (> 500 parasites/μL) determined by light microscopy was classified as a clinical episode. molFOB was calculated by comparing the genetically unique parasites for each species between two longitudinal blood samples, identified by polymerase chain reaction (PCR).^{13,14} molFOB counts included only newly acquired infections since the previous blood sample.

We combined the data into four 16-week intervals for each child. The data described in previous publications were analyzed by eight 8-week intervals. We originally fit the BPLMs to the 8-week interval data, but the clinical episode counts were very small and exhibited almost no overdispersion. The BPLM comes with overdispersion, so when there is no overdispersion, there is no association. Thus, the data in this article were aggregated over 16 weeks so that we could model larger counts. Furthermore, we did not want to aggregate over the entire study period, because we hypothesized that the correlation between the two species might change depending on the season.

We were interested in the rates at which children were getting infected by parasites and sick with malaria, rather than simply the counts of infections and clinical episodes. Children were not considered to be at risk of acquiring new infections for approximately 14 days after they were treated with anti-malarials; therefore, we calculated each child's total time at risk during a 16-week interval.

The BPLM. In this section, we describe a BPLM to model the association between the two responses caused by the two species of malaria parasites. We will use counts of clinical episodes caused by *P. falciparum* and counts of clinical episodes caused by *P. vivax* as examples to describe the model.

When two outcomes are measured on the same unit over the same period of time, it is relatively straightforward to adapt their univariate distributions to a bivariate distribution. Let Y_1 and Y_2 represent the counts of clinical episodes caused by the two species over a 16-week interval. We view them as dependent overdispersed Poisson random variables by defining their log-normally distributed conditional means λ_1 and λ_2 given two i.i.d. $N(0,1)$ random variables Z_1 and Z_2 as follows: $\lambda_1 = e^{\mu_1 + \sigma_1 U_1}$ and $\lambda_2 = e^{\mu_2 + \sigma_2 U_2}$, where $U_1 = Z_1$ and $U_2 = \rho Z_1 + (1 - \rho^2)^{1/2} Z_2$. The bivariate Poisson lognormal distribution thus has five parameters, μ_1 , σ_1 , μ_2 , σ_2 and ρ , and can be written as

$$\begin{aligned}
 P(Y_1 = y_1, Y_2 = y_2 | \mu_1, \mu_2, \sigma_1, \sigma_2, \rho) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\exp(y_1(\mu_1 + \sigma_1 u_1) - e^{\mu_1 + \sigma_1 u_1})}{y_1!} \\
 &\times \frac{\exp(y_2(\mu_2 + \sigma_2 u_2) - e^{\mu_2 + \sigma_2 u_2})}{y_2!} \\
 &\times \frac{1}{2\pi\sqrt{1 - \rho^2}} \\
 &\times \exp\left\{-\frac{1}{2(1 - \rho^2)} [u_1^2 + u_2^2 - 2\rho u_1 u_2]\right\} du_1 du_2. \quad (1)
 \end{aligned}$$

The covariance is

$$\text{Cov}(Y_1, Y_2) = e^{\mu_1 + \mu_2 + \frac{1}{2}(\sigma_1^2 + \sigma_2^2)} (e^{\rho\sigma_1\sigma_2} - 1). \quad (2)$$

The marginal means are $E(Y_1) = e^{\mu_1 + \frac{\sigma_1^2}{2}}$ and $E(Y_2) = e^{\mu_2 + \frac{\sigma_2^2}{2}}$. The marginal variances are $\text{Var}(Y_1) = e^{2\mu_1 + \sigma_1^2} (e^{\sigma_1^2} - 1) + e^{\mu_1 + \frac{\sigma_1^2}{2}}$ and $\text{Var}(Y_2) = e^{2\mu_2 + \sigma_2^2} (e^{\sigma_2^2} - 1) + e^{\mu_2 + \frac{\sigma_2^2}{2}}$. Thus, under the assumed bivariate distribution, Y_1 and Y_2 are conditionally independent given U_1 and U_2 .

We will describe Equation (1) including covariates in the next section. These covariates replace μ_j with $\beta_{0j} + \beta_{\text{age}j} + \beta_{\text{ITN}j} + \beta_{\text{FOB}j}$ for clinical episodes, and with $\beta_{0j} + \beta_{\text{age}j} + \beta_{\text{ITN}j}$ for infections, where j is one for *P. falciparum* coefficients and two for *P. vivax* coefficients. In previous analyses of these data, it was determined that age, insecticide treated bednet (ITN) use, and molFOB were important predictors of clinical episodes caused by a particular malaria species.^{13,14} *P. vivax* parasites can remain dormant in the liver for long periods of time, sometimes up to 3 years, unlike *P. falciparum* parasites that usually appear in the blood within 1–2 weeks of a bite by an infective mosquito. Thus, *P. vivax* can establish new blood-stage infections in the absence of additional mosquito bites through the activation of such long-lasting liver stages, called hypnozoites. For this reason, *P. falciparum* molFOB is a more direct measure of recent exposure to infective mosquito bites than *P. vivax* molFOB . In PNG, both parasites are transmitted by the same mosquitoes, so it is permissible to use *P. falciparum* molFOB as a surrogate measure of exposure to mosquito bites and thus risk of experiencing clinical episodes caused by either species. The validity of this approach was demonstrated by a recent study that used seasonal variation in *P. falciparum* molFOB as a surrogate marker for seasonal patterns of infective mosquito bites for either species.²¹

We also present Kendall's τ correlation coefficient estimates to compare the difference between the correlation we achieved using the BPLM and that estimated by a method that does not permit adjustment for other risk factors.

Fitting the BPLM to the PNG data. Here we describe two methods for estimating the parameters of the BPLM, maximum likelihood, using adaptive Gaussian quadrature (AGQ),²² which is an integral approximation method, and a Bayesian method that uses MCMC to estimate the posterior distributions of the parameters of the BPLM. We will refer to these two methods throughout as MLE and Bayesian, respectively. Both methods are presented because the Bayesian method provides numerical stability when the MLE is badly behaved. We compare these two algorithms with respect to their frequentist properties.

The first algorithm was fit using the NLMIXED procedure in SAS.²³ The NLMIXED procedure maximizes an approximation to the likelihood integrated over the overdispersion parameters. We chose AGQ for the integral approximation method because it is the default for NLMIXED. The second method was fit using the MCMCglmm package²⁴ in R.²⁵ The MCMCglmm package fits various types of linear and generalized linear mixed models using MCMC techniques. The algorithm uses a combination of Gibbs sampling, slice sampling, and Metropolis–Hastings to move the chain through the parameter space. We specified a multivariate inverse Wishart prior for the overdispersion parameters, $\sigma_1, \sigma_2 \sim \text{IW}([0.002, 0.002], \Sigma)$, where the diagonal values of Σ were equal to one and the off-diagonal values were equal to

zero. This prior assumed a priori independence between the σ 's; however, a fully parameterized covariance matrix was estimated from the model fit by the following matrix equation.

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

A multivariate normal prior with zero means and large variances (10^6) was specified for the μ 's. For each simulation, we used 20,000 iterations with a burn-in period of 2,000 and a thinning interval of 20.

RESULTS

The clinical episode counts for each species were similar in intervals 1 and 2, but there were slightly more *P. falciparum* cases in intervals 3 and 4 compared with *P. vivax* (Table 1). The incidence of clinical episodes was slightly higher for *P. vivax* compared with *P. falciparum* in intervals 1 and 2, but the opposite was true in intervals 3 and 4. The number and incidence of genetically distinct infections were much higher for *P. vivax* compared with *P. falciparum*. Intervals 3 and 4 had the highest infection rates of the four intervals (Table 1).

Incidence rates of clinical episodes and molFOB are summarized by person-years-at-risk in Table 1. Confidence intervals for the rates were calculated using Haenszel et al.'s table.²⁶ Children most often experienced between one and three clinical episodes per year-at-risk for both species. However, children experienced much higher rates of new *P. vivax* infections compared with new *P. falciparum* infections.

To determine an unadjusted association between clinical episode–annualized rates and between molFOB annualized rates, we calculated Kendall's τ rank correlation coefficients. These are summarized in Table 2. There was no association between clinical episodes caused by *P. falciparum* and those caused by *P. vivax* according to Table 2. However, there was a weak to moderate positive association between *P. falciparum* and *P. vivax* molFOB . The estimates in Table 2 do not permit adjustments for covariates that are also related to the outcomes. The advantage of fitting a BPLM is that it allows for estimation of the association between the two outcomes and estimation of the effects of the covariates on both outcomes simultaneously.

A BPLM fit to malaria clinical episodes caused by *P. falciparum* and *P. vivax* parasites. We fit a BPLM to clinical episode counts with age, ITN use, and *P. falciparum*

molFOB as covariates for each interval using both MLE and Bayesian methods (Table 3). Extensive discussion of the relationships between the covariates and the marginal responses is covered in previous analyses.^{13,14} Briefly, as we saw in previous analyses, age and rate of exposure were positively associated with clinical episodes caused by *P. falciparum*, and ITN use appeared to have a protective effect (Table 3). Age and ITN use were negatively associated with clinical episodes caused by *P. vivax* (Table 3). The primary objective of this analysis was to obtain a covariate-adjusted estimate of the association between *P. falciparum* and *P. vivax* clinical episodes. Table 3 displays parameter estimates from the BPLM fit to clinical episodes. The estimates of interest are the overdispersion parameters (σ *P. falciparum* and σ *P. vivax*) and the association parameter, ρ . The estimates of the fixed effect coefficients and σ 's (Table 3) by MLE and the Bayesian method agreed somewhat. There were considerable differences in the estimates of ρ between the two algorithms, especially in the credible (or confidence) intervals. The credible intervals were so wide that it was not worthwhile to calculate the correlation and interpret the results. The NLMIXED procedure failed to produce estimates of the standard errors for some of the σ 's and ρ 's in Table 3 because of optimization errors.

One explanation for the poor fits might have been the presence of outliers, or single observations that strongly influenced the parameter estimates. Thus, several model diagnostics were used to investigate the residuals and influence of specific observations on the parameter estimates using the MCMCglmm package. Plots of residuals for each of the marginal fitted values for clinical episodes caused by *P. falciparum* and *P. vivax* are provided in Supplemental Figure 1. There were large residuals for both *P. falciparum* and *P. vivax* clinical episode counts in all four intervals, and most of them were in the negative direction, indicating that the fitted values were lower than the observed values. The range of fitted values across all intervals and species was zero to two clinical episodes, whereas the range of observed values was zero to five clinical episodes. The interquartile ranges of clinical episodes caused by both species were zero to one in the observed data; thus, the fitted values from the model were in an acceptable range of the majority of the distribution of observed episode counts, but the model did a poor job of estimating the tails of the distributions.

We also evaluated the parameter estimates from MCMCglmm using the bootstrap method.²⁷ To do this, we sampled the

TABLE 1

Median counts (interquartile range) and incidence rates per person-years-at-risk (95% confidence interval) for clinical episodes and molFOB

Variable	Interval 1	Interval 2	Interval 3	Interval 4
<i>Plasmodium falciparum</i> clinical episodes				
Count	0 (0, 1)	0 (0, 1)	1 (0, 1)	0 (0, 1)
Rate	1.5 (1.2, 1.9)	1.7 (1.4, 2.1)	3.1 (2.7, 3.6)	1.3 (1.1, 1.7)
<i>Plasmodium vivax</i> clinical episodes				
Count	0 (0, 1)	0 (0, 1)	0 (0, 1)	0 (0, 0)
Rate	1.8 (1.4, 2.2)	2.0 (1.6, 2.4)	1.9 (1.6, 2.3)	0.8 (0.6, 1.1)
<i>P. falciparum</i> molFOB				
Count	1 (0, 2)	1 (0, 2)	1 (0, 3)	1 (0, 2)
Rate	5.5 (5.0, 6.0)	4.4 (4.0, 4.8)	8.2 (7.5, 9.0)	3.9 (3.6, 4.3)
<i>P. vivax</i> molFOB				
Count	3 (1, 5)	4 (2, 6)	4 (2, 7)	3 (1, 5)
Rate	15.3 (14.2, 16.4)	17.0 (16.0, 18.1)	18.3 (17.2, 19.5)	11.2 (10.4, 12.0)

TABLE 2

Kendall's τ correlation coefficients (P values) for clinical episode and molFOB annualized rates caused by *Plasmodium falciparum* and *Plasmodium vivax* parasites

	Clinical episodes	molFOB
Interval 1	0.06 (0.3)	0.16 (0.001)
Interval 2	-0.003 (1.0)	0.20 (< 0.001)
Interval 3	-0.02 (0.6)	0.14 (0.001)
Interval 4	0.1 (0.09)	0.11 (0.02)

same number of observations that existed in the observed data for each interval with replacement and fit BPLMs to each sample. Supplemental Figure 2 displays the distributions of the parameter estimates from the 1,000 samples by interval. It is clear from the histograms that there was substantial variability in the parameter estimates across the 1,000 samples. We computed bootstrap confidence intervals using the 2.5th and 97.5th percentiles of the fitted distributions for each parameter. These values are provided in Table 3. The bootstrap confidence intervals were slightly larger than the Bayesian credible intervals for the μ 's and σ 's but were slightly smaller for the estimates of ρ .

A BPLM fit to *P. falciparum* and *P. vivax* molFOB . In this section, we present the results of fitting a BPLM to the rate of acquisition of *P. falciparum* and *P. vivax* parasites, or molFOB . As described previously, *P. vivax* molFOB is not a direct measure of recent exposure to infective mosquito bites; rather, it is a measure of blood-stage infection caused by both primary and relapsing infections. Therefore, by fitting a BPLM to *P. falciparum* and *P. vivax* molFOB , we were estimating the association between rates of new blood-stage infections of each parasite acquired during an interval. Each BPLM included age and ITN use as covariates and were again fit using NLMIXED and MCMCglmm. The results are summarized by interval in Table 4. Again, we reported extensively on the

marginal models in previous analyses,^{13,14} and here we see similar relationships, where age was positively associated with *P. falciparum* infections and ITN use was negatively associated with *P. falciparum* and *P. vivax* infections (Table 4).

There was a moderate positive association between *P. falciparum* and *P. vivax* molFOB in the first three 16-week intervals (Table 4, ρ). To get an estimate of the correlation between the two responses, we used the following equation

$$\text{Cor}(Y_1, Y_2) = \frac{\text{Cov}(Y_1, Y_2)}{\sqrt{\text{Var}(Y_1)}\sqrt{\text{Var}(Y_2)}}. \quad (3)$$

The results from applying this formula to the parameter estimates in Table 4 are summarized in Table 5. The age and ITN coefficients were multiplied by the mean age and mean ITN use over all children during each specific interval, and these values were added to the intercepts to obtain the μ 's.

The estimates of correlation presented in Table 5 were much higher than those obtained using Kendall's τ statistic (Table 2) and were significantly greater than zero for the first three intervals. The sign and magnitude of the correlation coefficients presented in Table 5 suggest that *P. falciparum* and *P. vivax* molFOB were moderately positively correlated. The estimates of the overdispersion parameters for molFOB counts in Table 4 were very similar between the two estimation algorithms and their confidence intervals were much narrower than those estimated for clinical episodes in Table 3.

DISCUSSION

Coinfection with *P. falciparum* and *P. vivax* parasites was common in this cohort of PNG children. In this study, we found no indication that *P. vivax* is protective against *P. falciparum* infections, as indicated by the positive correlation coefficients in Table 5. Our findings align with a number of more recent studies that have shown that *P. vivax* does not appear to have a

TABLE 3
Results from using MLE and Bayesian methods to fit BPLMs to clinical episodes by 16 week intervals

Method	Interval 1	Interval 2	Interval 3	Interval 4
MLE				
Age: Pf	37 (3, 72)	7 (-23, 37)	6 (-17, 28)	-7 (-34, 20)
Age: Pv	-26 (-63, 10)	-28 (-60, 5)	-54 (-87, -21)	-50 (-94, -7)
ITN use: Pf	-38 (-89, 14)	-5 (-51, 41)	-33 (-70, 4)	-14 (-66, 38)
ITN use: Pv	52 (-1, 104)	26 (-22, 74)	39 (-9, 87)	7 (-62, 77)
molFOB : Pf	5 (3, 8)	10 (8, 13)	4 (3, 6)	8 (5, 10)
molFOB : Pv	4 (1, 7)	-4 (-9, 1)	0 (-3, 3)	-5 (-13, 3)
σ : Pf	33 (3, 103)	2 (-, -)	2 (-, -)	30 (11, 111)
σ : Pv	68 (37, 144)	71 (45, 156)	61 (34, 140)	119 (83, 228)
ρ	59 (-148, 267)	0 (-, -)	40 (-, -)	100 (-, -)
Bayesian				
Age: Pf	33 (9, 60)	2 (-23, 26)	8 (-12, 30)	-10 (-38, 21)
Age: Pv	-26 (-50, 5)	-28 (-58, 3)	-47 (-74, -16)	-44 (-88, -4)
ITN: Pf	-34 (-77, 14)	0 (-47, 45)	-28 (-67, 14)	-17 (-56, 37)
ITN: Pv	44 (2, 96)	29 (-21, 82)	44 (-5, 87)	18 (-76, 105)
molFOB : Pf	5 (3, 7)	11 (8, 13)	4 (3, 6)	8 (5, 11)
molFOB : Pv	4 (1, 7)	-2 (-7, 2)	0 (-3, 3)	-6 (-15, 2)
σ : Pf	24 (3, 54)	16 (3, 30)	17 (4, 33)	32 (3, 77)
σ : Pv	53 (3, 107)	81 (40, 123)	54 (11, 86)	127 (82, 177)
ρ	7 (-97, 100)	31 (-96, 100)	-15 (-98, 100)	74 (-81, 100)
Bayesian with bootstrap				
σ : Pf	35 (13, 73)	19 (9, 35)	17 (9, 33)	34 (13, 66)
σ : Pv	61 (19, 105)	70 (27, 103)	51 (15, 90)	117 (35, 174)
ρ	31 (-89, 98)	14 (-86, 93)	5 (-88, 88)	72 (-32, 99)

BPLMs = bivariate Poisson lognormal models; ITN = insecticide treated bednet; MLE = maximum likelihood estimation; Pf = *Plasmodium falciparum*; Pv = *Plasmodium vivax*; - = failed to converge. All values are multiplied by 100.

TABLE 4
Results from using MLE and Bayesian methods to fit BPLMs to m_{ol} FOB counts by 16-week intervals

Method	Interval 1	Interval 2	Interval 3	Interval 4
MLE				
Age: Pf	26 (0, 52)	51 (30, 72)	41 (25, 57)	21 (-3, 44)
Age: Pv	12 (-7, 30)	12 (-4, 27)	14 (0, 28)	13 (-3, 28)
ITN: Pf	-21 (-59, 18)	-122 (-157, -8)	-88 (-114, -62)	-135 (-174, -96)
ITN: Pv	-54 (-82, -26)	-64 (-88, -40)	-19 (-41, 3)	-59 (-83, -35)
σ : Pf	84 (64, 103)	54 (33, 75)	48 (34, 62)	70 (51, 88)
σ : Pv	70 (57, 83)	65 (55, 76)	58 (48, 68)	59 (47, 71)
ρ	35 (8, 62)	62 (28, 95)	43 (13, 74)	2 (-31, 35)
Bayesian				
Age: Pf	26 (0, 52)	51 (30, 69)	41 (25, 55)	21 (-5, 43)
Age: Pv	12 (-6, 33)	12 (-4, 27)	14 (0, 27)	13 (-2, 29)
ITN: Pf	-22 (-61, 17)	-121 (-154, -87)	-87 (-113, -60)	-135 (-172, -98)
ITN: Pv	-54 (-82, -24)	-64 (-88, -43)	-19 (-40, 3)	-58 (-85, -35)
σ : Pf	85 (66, 104)	52 (32, 76)	47 (34, 61)	69 (50, 91)
σ : Pv	70 (58, 84)	66 (56, 78)	59 (50, 70)	60 (47, 71)
ρ	36 (9, 62)	72 (37, 99)	46 (15, 78)	3 (-34, 35)

BPLMs = bivariate Poisson lognormal models; ITN = insecticide treated net; MLE = maximum likelihood estimation; Pf = *Plasmodium falciparum*; Pv = *Plasmodium vivax*. All values are multiplied by 100.

protective effect against *P. falciparum*.^{9,10,11} There are several key reasons why our study was well-suited to investigate this relationship. First, blood samples were acquired from participants in this study at regular intervals, and individuals were only treated if they had a clinical case of malaria, not simply if they had parasitemia. This made it possible for us to detect exposure to both parasites longitudinally in a natural environment. Second, we used PCR-based detection and differentiation of *P. falciparum* and *P. vivax*, whereas other studies have used light microscopy, which can miss coinfections, especially when one species is more predominant in a population.^{9,28,29} Although these diagnostic difficulties also apply to PCR-based detection, they do to a lesser degree.^{30,31} Finally, the statistical model that we applied to this problem is a novel application, and to our knowledge, it is the most appropriate model for estimating the correlation between *P. falciparum* and *P. vivax* exposure while adjusting for other risk factors.

The strength of correlation between the infection rates varied across the four intervals (Table 5). It was not clear why the association between *P. falciparum* m_{ol} FOB and *P. vivax* m_{ol} FOB was so much lower in interval 4 compared with the other intervals. One explanation might be that incidence of newly acquired *P. falciparum* parasites is more strongly associated with season, because these parasites are seen more immediately in the blood after an infective bite by a mosquito. A recent study reported that 70–90% of *P. vivax* blood-stage infections are the result of relapsing clones from long-lasting, dormant liver stages.²¹ These relapsing infections can occur without the need to acquire new infections from mosquito bites, thus reducing the amount of seasonality in incidence of *P. vivax* blood-stage infections. The rates of *P. falciparum* infections were lowest in interval 4, which coincided with the

start of the dry season. This followed the wet-season interval with the highest rate of clinical episodes caused by *P. falciparum*. As reported earlier, there were higher rates of *P. vivax* infections (15.3 per child per year) in these children compared with *P. falciparum* (5.4 per child per year), yet there were more clinical episodes caused by *P. falciparum* (1.9 per child per year) than those caused by *P. vivax* (1.6 per child per year) over the entire cohort. This is likely due to faster acquisition of immunity in the presence of high *P. vivax* m_{ol} FOB.¹³

The BPLM for jointly modeling *P. falciparum* and *P. vivax* burden performed much better for m_{ol} FOB counts. These were estimated consistently by MLE, which was likely because both of these variables had higher counts and exhibited more positive skewness than the clinical episodes (Table 1). A BPLM turned out not to be a suitable choice to model clinical episodes in these data because the counts were too small and exhibited almost no overdispersion.

We conducted several simulation experiments to investigate the numerical instability seen in the small counts of clinical episodes (not shown). Judging by the simulation results, low bias and reasonable standard errors could not be achieved by either method until the means and standard deviations were sufficiently large, creating overdispersion. The Bayesian method did a reasonable job of estimating ρ when it was smaller, but it tended to overestimate ρ when it had a larger absolute value (simulation results not shown). The multivariate crash data analyzed by Park and Lord¹⁸ included five dependent crash severity measures. Two of these measures had small means (< 0.5), but they exhibited more positive skew than the clinical episode counts in the PNG data. Park and Lord implemented a Bayesian model; however, they did not provide credible intervals for their estimates of ρ , so it is difficult to assess stability from the results presented. Finally, it is impossible to test every possible scenario through simulations, but based on the simulations we conducted, we support simulating data to mimic your counts before trusting the model estimates, especially when the means and standard deviations are small.

Finally, a BPLM fit by MLE is a useful method for estimating a covariate-adjusted correlation between two overdispersed Poisson variables, so long as the means and standard deviations are sufficiently large, especially when the two variables

TABLE 5

Estimates of the association between *Plasmodium falciparum* and *Plasmodium vivax* m_{ol} FOB using the parameter estimates from the BPLMs (Table 4) and Equation (3)

Method	Interval 1	Interval 2	Interval 3	Interval 4
MLE	25 (4, 50)	39 (7, 87)	30 (5, 65)	1 (-7, 27)
Bayesian	26 (4, 48)	44 (11, 84)	33 (6, 67)	2 (-7, 29)

BPLMs = bivariate Poisson lognormal models; MLE = maximum likelihood estimation. All values multiplied by 100.

might be negatively correlated. This method should be considered when analyzing counts of simultaneously measured outcomes that are assumed to be correlated, which is the case with important biological features of infectious diseases.

Received June 30, 2017. Accepted for publication October 14, 2017.

Published online November 27, 2017.

Supplemental figures appear at www.ajtmh.org.

Acknowledgments: We would like to thank the participants and their families from the PNG study and the laboratory and field teams.

Financial support: This work was funded in part by Australian National Health and Medical Research Council (NHMRC) program grant 490037, Swiss National Science Foundation Grants 310030-134889 and 31003A-112196, and National Institutes of Health Grants AI063135, AI-46919, and TW007872.

Authors' addresses: Kathryn L. Colborn, Department of Biostatistics and Informatics, Colorado School of Public Health, University of Colorado Denver, Aurora, CO, E-mail: kathryn.colborn@ucdenver.edu. Ivo Mueller, Walter and Eliza Hall Institute, Melbourne, Australia, E-mail: ivomueller@fastmail.fm. Terence P. Speed, Walter and Eliza Hall Institute, Melbourne, Australia, Department of Mathematics and Statistics, University of Melbourne, Melbourne, Australia, and Department of Statistics, University of California, Berkeley, Berkeley, CA, E-mail: terry@wehi.edu.au.

REFERENCES

- Coovadia H, Wilkinson P, 1998. Childhood human immunodeficiency virus and tuberculosis co-infections: reconciling conflicting data. *Int J Tuberc Lung Dis* 2: 844–851.
- Maurer A, Sturchler D, 2000. A waterborne outbreak of small round structured virus, *Campylobacter* and shigella co-infections in La Neuveville, Switzerland, 1998. *Epidemiol Infect* 125: 325–332.
- Backus L, Boothroyd D, Deyton L, 2005. HIV, hepatitis C and HIV/hepatitis C virus co-infection in vulnerable populations. *AIDS* 19: S13–S19.
- McCulloch C, 2008. Joint modeling of mixed outcome types using latent variables. *Stat Methods Med Res* 17: 53–73.
- Müller I, Bockarie M, Alpers M, Smith T, 2003. The epidemiology of malaria in Papua New Guinea. *Trends Parasitol* 2003: 253–259.
- Luxemburger C, Ricci F, Nosten F, Raimond D, Bathet S, White NJ, 1997. The epidemiology of severe malaria in an area of low transmission in Thailand. *Trans R Soc Trop Med Hyg* 91: 256–262.
- Maitland K, Williams T, Newbold C, 1997. *Plasmodium vivax* and *P. falciparum*: biological interactions and the possibility of cross-species immunity. *Parasitol Today* 13: 227–231.
- Maitland K, Williams TN, Peto TE, Day KP, Clegg JB, Weatherall DJ, Bowden DK, 1997. Absence of malaria-specific mortality in children in an area of hyperendemic malaria. *Trans R Soc Trop Med Hyg* 91: 562–566.
- Genton B, D'Acremont V, Rare L, Baea K, Reeder JC, Alpers MP, Müller I, 2008. *Plasmodium vivax* and mixed infections are associated with severe malaria in children: a prospective cohort study from Papua New Guinea. *PLoS Med* 5: 881–889.
- McKenzie FE, Smith DL, O'Meara WP, Forney JR, Magill AJ, Permpanich B, Erhart LM, Sirichaisinthop J, Wongsrichanalai C, Gasser RA Jr, 2006. Fever in patients with mixed-species malaria. *Clin Infect Dis* 42: 1713–1718.
- Price R, Tjitra E, Guerra CA, Yeung S, White NJ, Anstey NM, 2007. Vivax malaria: neglected and not benign. *Am J Trop Med Hyg* 77: 79–87.
- Lin E et al., 2010. Differential patterns of infection and disease with *P. falciparum* and *P. vivax* in young Papua New Guinean children. *PLoS One* 5: e9047.
- Koepfli C, Colborn KL, Kiniboro B, Lin E, Speed TP, Siba PM, Felger I, Mueller I, 2013. A high force of *Plasmodium vivax* blood-stage infection drives the rapid acquisition of immunity in Papua New Guinean children. *PLoS Negl Trop Dis* 7: e2403.
- Mueller I et al., 2012. Force of infection is key to understanding the epidemiology of *Plasmodium falciparum* malaria in Papua New Guinean children. *Proc Natl Acad Sci USA* 109: 10030–10035.
- Aitchison J, Ho C, 1989. The multivariate Poisson-log normal distribution. *Biometrika* 76: 643–653.
- Ma X, Chen S, Chen F, 2016. Correlated random-effects bivariate poisson lognormal model to study single-vehicle and multi-vehicle crashes. *J Transp Eng* 142: 04016049.
- Ma J, Kockelman KM, Damien P, 2008. A multivariate Poisson-lognormal regression model for prediction of crash counts by severity, using Bayesian methods. *Accid Anal Prev* 40: 964–975.
- Park E, Lord D, 2007. Multivariate Poisson-lognormal models for jointly modeling crash frequency by severity. *Transp Res Rec* 2019: 1–6.
- Engen S, Lande R, DeVries PJ, 2002. Analyzing spatial structure of communities using the two-dimensional Poisson lognormal species abundance model. *Am Nat* 160: 60–73.
- Chib S, Winkelmann R, 2012. Markov chain Monte Carlo analysis of correlated count data. *J Bus Econ Stat* 19: 428–435.
- Ross A, Koepfli C, Schoepflin S, Timinao L, Siba P, Smith T, Mueller I, Felger I, Tanner M, 2016. The incidence and differential seasonal patterns of *Plasmodium vivax* primary infections and relapses in a cohort of children in Papua New Guinea. *PLoS Negl Trop Dis* 10: e0004582.
- Pinheiro J, Bates D, 1995. Approximations to the log-likelihood function in the nonlinear mixed-effects model. *J Comput Graph Stat* 4: 12–35.
- SAS Institute, 2013. Copyright 2013 SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC.
- Hadfield JD, 2010. MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *J Stat Softw* 33: 1–22.
- R Core Team, 2015. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.
- Haenszel W, Loveland D, Sirken M, 1962. Lung-cancer mortality as related to residence and smoking histories. I. White males. *J Natl Cancer Inst* 28: 947–1001.
- Efron B, Tibshirani R, 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Stat Sci* 1: 54–75.
- Phimpraphi W et al., 2008. Longitudinal study of *Plasmodium falciparum* and *Plasmodium vivax* in a Karen population in Thailand. *Malar J* 7: 99.
- Smith T, Genton B, Baea K, Gibson N, Narara A, Alpers MP, 2001. Prospective risk of morbidity in relation to malaria infection in an area of high endemicity of multiple species of *Plasmodium*. *Am J Trop Med Hyg* 2001: 262–267.
- McNamara DT, Kasehagen LJ, Grimberg BT, Cole-Tobian J, Collins WE, Zimmerman PA, 2006. Diagnosing infection levels of four human malaria parasite species by a polymerase chain reaction/ligase detection reaction fluorescent microsphere-based assay. *Am J Trop Med Hyg* 2006: 413–421.
- Snounou G, Viriyakosol S, Zhu XP, Jarra W, Pinheiro L, do Rosario VE, Thaithong S, Brown KN, 1993. High sensitivity of detection of human malaria parasites by the use of nested polymerase chain reaction. *Mol Biochem Parasitol* 1993: 315–320.