# Evaluating the predictive performance of Malaria antibodies and FCGR3B Gene Polymorphisms on Plasmodium falciparum infection outcome: a prospective cohort study

**Duah Dwomoh** (duahdwomoh@yahoo.com)
University of Ghana
**Bright Adu**
University of Ghana
**Daniel Dodoo**
University of Ghana
**Michael Theisen**
University of Copenhagen
**Samuel Iddi**
University of Ghana School of Medicine and Dentistry
**Thomas A. Gerds**
University of Copenhagen

**Research**

**Keywords:** Evaluating the predictive performance of Malaria antibodies, FCGR3B Gene Polymorphisms, on Plasmodium falciparum infection outcome,

**Posted Date:** December 13th, 2019

**DOI:** https://doi.org/10.21203/rs.2.18670/v1



**Version of Record:** A version of this preprint was published at Malaria Journal on August 27th, 2020. See the published version at https://doi.org/10.1186/s12936-020-03381-8.

# Evaluating the predictive performance of Malaria antibodies and *FCGR3B* Gene Polymorphisms on *Plasmodium falciparum* infection outcome: a prospective cohort study

Duah Dwomoh[1], Bright Adu[2], Daniel Dodoo[2], Michael Theisen[3,4], Samuel Iddi[5], Thomas A. Gerds[6]

1. Department of Biostatistics, School of Public Health, University of Ghana, Accra, Ghana.
2. Department of Immunology, Noguchi Memorial Institute of Medical Research, College of Health Sciences, University of Ghana, Accra, Ghana.
3. Department for Congenital disorders, Statens Serum Institut, Copenhagen, Denmark.
4. Centre for Medical Parasitology at Department of International Health, Immunology and Microbiology, University of Copenhagen and Department of Infectious Diseases, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark
5. Department of Statistics and Actuarial Sciences, University of Ghana, Accra, Ghana.
6. Biostatistics Section, University of Copenhagen, Copenhagen, Denmark.

Bright Adu email: badu@noguchi.ug.edu.gh
Daniel Dodoo: ddodoo@noguchi.ug.edu.gh
Michael Theisen: mth@ssi.dk
Samuel Iddi: isamuel.gh@gmail.com
Thomas Alexander Gerds: tag@biostat.ku.dk

Corresponding authour : duahdwomoh@yahoo.com

# Abstract

## Background

Malaria antigen-specific antibodies and polymorphisms in host receptors involved in antibody functionality have been associated with different outcomes of *Plasmodium falciparum* infections. Thus, to identify key prospective malaria antigens for vaccine development, there is the need to evaluate the associations between malaria antibodies and antibody dependent host factors with more rigorous statistical methods. In this study, we employ suitable statistical models to evaluate the predictive performance of malaria-specific antibodies and host gene polymorphisms on *P. falciparum* infection in a longitudinal cohort study involving Ghanaian children.

## Methods

Models with different functional forms were built using known predictors (age, sickle cell status, blood group status, parasite density, and mosquito bed net use) and malaria antigen-specific antibodies and *FCGR3B* polymorphisms shown to mediate antibody-dependent neutrophil function. The models were evaluated through visualization and assessment of differences between the Area Under the Receiver Operating Characteristic Curve and Brier Score estimated by suitable internal cross-validation designs.

## Results

This study has found that the *FCGR3B*-c.233C>A genotype and IgG against AMA1 were relatively better compared to the other antibodies and *FCGR3B* genotypes in classifying or predicting malaria risk among children.

## Conclusions

Apical Membrane Antigen 1 could be a key potential vaccine antigen while *FCGR3B*-c.233C>A under the additive and dominant models of inheritance could be an important modifier of the effect of malaria protective antibodies.

# Background

Malaria remains a major public health concern globally and is considered as one of the most prevalent and lethal human infectious diseases among children in sub-Saharan Africa [1]. Despite the drastic reduction in the number of malaria cases and deaths in all ages globally, it still accounted for 10% of child deaths in sub-Saharan Africa [1] and mortality is mostly higher among children below the age of five years. Individuals in endemic regions increasingly develop resistance to infection and disease with age and this is conventionally thought to reflect a slow and gradual acquisition of protective immunity [2]. It has recently been shown that interaction between naturally acquired antibodies to *Plasmodium falciparum* and polymorphisms in host genes (*FCGR3B*) plays a key role in immunity against malaria [3]. It is conceivable that other host genes may also modify the protective effect of malaria antibodies. This emphasizes the need for robust modeling approaches to effectively address such confounders in malaria vaccine studies. It is quite plausible that the long delay in attaining an effective malaria vaccine may partly be due to inadequacies of traditional statistical approaches used in malaria immuno-epidemiological studies to determine the performance of predictors in classifying or predicting malaria risk [4–7]. Traditionally, most studies use generalized linear models (GLM) depending on the measurement of clinical malaria which provides an extensive class of tools for modeling the effect of predictors. Statistical prognostic modeling techniques have been applied primarily in the area of non-communicable diseases such as cardiovascular diseases and lung cancer. For instance, Gail et al. [8] developed a model of breast cancer risk prediction and implications for chemoprevention which was later validated by Beverly et al. [9]. Several risk prediction models for other cancers and cardiovascular diseases [10–18] have also been developed. For clinical malaria, on the other hand, personalized risk estimation has not been extensively studied. As indicated by several authors [4–7, 19], markers such as polymorphisms and antigen-specific antibodies proposed for classifying or predicting risk in individual subjects must be held to a much higher standard than just assessing associations based on odds ratio estimates. Pepe et al. [19] showed that strong statistical associations (odds ratio, relative risk, etc) between disease and host-specific factors found in literature do not necessarily imply that those factors could discriminate between a subject who is likely or not have the disease in a specified time. A risk prediction model exploits the joint predictive power of several variables on the risk of an even or disease. A robust malaria risk prediction model based on epidemiological predictors may contribute to finding possible answers to the question of which parasite antigens and host factors should be the main research focus in the efforts to find optimal control strategies and vaccines. This is particularly important as the number of malaria-specific antibodies and host gene polymorphisms found to be associated with clinical malaria have increased significantly over the past few years but with little impact on malaria control. Using a more rigorous prediction modeling approach, this study aims to evaluate the predictive performance of malaria antibodies and *FCGR3B* gene polymorphisms on *P. falciparum* infection outcome by estimating Brier scores and Area Under the Receiver Operating Characteristic Curve (AUROC) through appropriate bootstrap cross-validation design. The identified model was obtained by comparing Brier score estimates and the AUROC curve of several models that integrate malaria antibodies and host gene polymorphisms.

# Methods

## Data source

Data used for modeling the risk of malaria was secondary data obtained from a prospective longitudinal malaria cohort study which was conducted from May 2008 to January 2009 among children under 13 years of age in five different communities in the Shai Osudoku (formerly Dangme West) district of Ghana (Asutsuare, Kewum, Mafikorpe, Osuwem, and

Volivo) [20, 21]. The study recruited 799 children of which 393 (49.2%) were males and 406 (50.8%) were females. These children were observed both actively and passively for malaria case detection. The primary outcome measure was clinical malaria defined as fever with any level of *P. falciparum* parasitemia plus at least one clinical symptom of malaria such as vomiting, joint pains, diarrhea e.t.c. In this study, the term "Susceptible" and "Protected" are used to represent clinical malaria and no malaria case detection respectively over the study period. The proportion of children that developed malaria in the one year follow up was 15.0% (incidence proportion) and there were approximately 1.7 malaria cases per 100 children per month (incidence rate) [20]. The predictors of clinical malaria included age in years, sex, sickle cell status, blood group, hemoglobin level, malaria antigen-specific immunoglobulin (Ig) G and subclasses (IgG1, IgG2, IgG3, and IgG4) and *FCGR3B* polymorphisms. We present a summary of climatic variables at the time of the study to serve as a guide for future studies that may wish to compare their findings to this study. Changes in relative humidity (RH) which is the ratio of the partial pressure of water vapor relative to saturated vapour at the specified temperature was assessed over the study period. The minimum and maximum relative humidity over the study period were 71.0% and 98.0% respectively. The total monthly rainfall ranged between 0 and 273.5mm with the month of May 2008 recording the highest total rainfall. The total rainfall between June and December 2008 ranged from 19.7mm to 125.4mm. The maximum daily rainfall was recorded in the month of June 2008 (60mm). There was no rainfall in the month of January 2009. The minimum and maximum air temperature ranged from $19.3^{o}C$ to $36.8^{o}C$.

## Malaria risk prediction model and performance measures

Let $D_n = \{Y_i, \boldsymbol{X}_i\}_{i=1,\dots,n}$ be a malaria data set with $n$ number of children aged less than 13 years. $\boldsymbol{X}_i = (X_{ip})_{p=1,\dots,P, i=1,\dots,n}$ be an input matrix of $P$ predictors of malaria (age, antibodies, e.t.c) and

$$Y_i = \begin{cases} 1 \; if \; the \; child \; is \; positive \\ 0 \; if the \; child \; is \; negative \end{cases}$$

Let $G \subseteq \{1,\cdots G\}$ be a subset of the available predictors where $p = p_1, p_2, \dots, p_g$ . Let $\beta = \beta_1, \beta_2, \dots, \beta_g$ be the regression coefficient to be estimated. The predicted risk of malaria is modeled using the logistic regression model. Specifically, for the trained prediction model, $\hat{\tau}_n$ which assigns to each child the probability of developing clinical malaria, the estimated malaria risk prediction model is given by:

$$P(Y_i = 1|\boldsymbol{X}_i) = \hat{\tau}_n(\boldsymbol{X}_i) = \hat{\tau}_n = \left[1 + \exp\{\hat{\beta}_0 - \sum_{g \, \epsilon \, G} \hat{\beta}_g(X_p)\}\right]^{-1}$$

where $\hat{\beta}$ is a vector of estimated parameters of the model associated with predictor variables, $\hat{\beta}_0$ is the estimated intercept of the logistic regression model considered as the baseline risk of malaria. and $\hat{\tau}_n(\boldsymbol{X}_i)$ is the predicted risk of malaria for the $i$th child with baseline characteristics $\boldsymbol{X}_i$.

The modeling strategy was to first identify a baseline model out of several other competing models relating to the prevalence of malaria to baseline covariates. Comprehensive discrimination and calibration assessments of the fitted models were explored based on the AUROC curve, Brier score, root mean squared error and the total explained variations in predicted probabilities ($R^2$) using bootstrap cross-validation from 200 bootstrap samples with replacement. In selecting predictor variables to be included in the final model, candidate predictor variables were screened using tools for discriminative and calibration abilities such as the area under the receiver operating characteristics curve, the Brier score and explained variation in predicted risk. The model's ability to predict accurately was further assessed by means of the calibration plot. According to Gerds et al. [22], the values of the Brier score could be interpreted as the loss or regret which is incurred when the

prediction model $\hat{\tau}_n$ is applied to a child whose true malaria status is $Y_i$. Model performance was assessed via bootstrap cross-validation. Selection of other candidate predictor variables (age, sickle cell, blood group, hemoglobin, parasite density and, mosquito net use) for the baseline model was premised on subject matter knowledge and prior evidence of association with the risk of malaria from literature. Continuous predictors were fitted via restricted cubic splines. To quantify over-fitting and to recalibrate the model, the heuristic shrinkage estimator $\hat{\gamma} = \frac{model\chi^2 - p}{model\chi^2}$ was used where $p$ is the number of predictors (regression parameters including both linear and non-linear and possible interaction terms), $\chi^2$ is the likelihood ratio $\chi^2$ test statistic computed using the full set of $p$ parameters to determine whether any of the predictor(s) is/are associated with log-odds of developing clinical malaria. For the model to be calibrated well for future data, $\hat{\gamma}$ was multiply by $X\hat{\beta}$ and that defines shrinkage. The penalty factor was determined by means of repeated cross-validation of the data. All antigen-specific antibodies were log-transformed to base $e$ in subsequent analysis. All models were fitted with R programming software version 3.2.4 with the following specialized packages: Design [23], Penalized [24], Data cleaning and all other forms of data preparations were done with Stata SE version 13. A $p$-value of $< .05$ was considered statistically significant.

# Results

## Description of study participants and clinical malaria distribution

Complete information on candidate predictors was available for 395 children. The overall median age for this study sample was 5.0 years (Interquartile range =3.0-8.0). The cumulative incidence of malaria was 13% (53 out of 395 children). The bed net use among the children was 40.5%. The analysis of sociodemographic characteristics and baseline biomarkers on the risk of clinical malaria indicated that the cumulative incidence of malaria did not differ significantly among the baseline predictors that were studied ($p$-value $\geq 0.05$). Distribution of other predictors and clinical malaria status can be found in Table 1.

204
205
206 Table 1. Bivariate analysis of malaria predictors and clinical malaria

| Predictor | Levels | Protected N=342 | Susceptible N=53 | Combined N=395 | p-value |
|---|---|---|---|---|---|
| Age in years | | 5(3,8) | 5 (3,7) | 5 (3,8) | 0.12[e] |
| Mosquito net use | No | 60% (204) | 58% (31) | 59% (235) | 0.87[f] |
| Blood group | A | 18% (63) | 19% (10) | 18% (73) | 0.93[f] |
| | AB | 6% (22) | 8% (4) | 7% (26) | |
| | B | 29% (98) | 25% (13) | 28% (111) | |
| | O | 46% (159) | 49% (26) | 47% (185) | |
| Sickle cell status | Positive | 15% (51) | 19% (10) | 15% (61) | 0.46[f] |
| Parasite count(categorized) | positive | 7% (25) | 4% (2) | 7% (27) | 0.34[f] |
| Hemoglobin at enrollment(gram per dL) | | 12 (11,13) | 11 (11,12) | 12 (11,13) | 0.11[e] |
| Additive c.108C > G | CC | 29% (100) | 26% (14) | 29% (114) | 0.4[f] |
| | CG | 42% (144) | 36% (19) | 41% (163) | |
| | GG | 29% (98) | 38% (20) | 30% (118) | |
| Additive c.114T > C | CC | 27% (92) | 26% (14) | 27% (106) | 0.82[f] |
| | CT | 43% (147) | 47% (25) | 44% (172) | |
| | TT | 30% (103) | 26% (14) | 30% (117) | |
| Additive c.233C > A | AA | 9% (31) | 2% (1) | 8% (32) | 0:2[f] |
| | AC | 27% (93) | 28% (15) | 27% (108) | |
| | CC | 64% (218) | 70% (37) | 65% (255) | |
| Additive c.244A > G | GG | 31% (106) | 32% (17) | 31% (123) | 0.36[f] |
| | AG | 39% (134) | 47% (25) | 40% (159) | |
| | AA | 30% (102) | 21% (11) | 29%(113) | |
| Additive c.316A > G | GG | 13% (43) | 17% (9) | 13% (52) | 0.65[f] |
| | AG | 32% (109) | 32% (17) | 32% (126) | |
| | AA | 56% (190) | 51% (27) | 55% (217) | |
| Additive c.194A > G | GG | 39% (135) | 30% (16) | 38% (151) | 0.43[f] |
| | AG | 39% (135) | 45% (24) | 40% (159) | |
| | AA | 21% (72) | 25% (13) | 22% (85) | |
| Dominant c.108C > G | CC/CG | 71% (244) | 62% (33) | 70%(277) | 0:18[f] |
| Dominant c.114T > C | TT/CT | 73% (250) | 74% (39) | 73% (289) | 0:94[f] |
| Dominant c.194A > G | AA/GG | 61% (207) | 70% (37) | 62% (244) | 0:2[f] |
| Dominant c.233C > A | AC/CC | 91% (311) | 98% (52) | 92% (363) | 0:075[f] |
| Dominant c.244A > G | AG/AA | 69% (236) | 68% (36) | 69% (272) | 0:87[f] |
| Dominant c.316A > G | AG/AA | 87% (299) | 83% (44) | 87% (343) | 0:38[f] |
| Recessive c.108C > G | CC | 29%(100) | 26% (14) | 29% (114) | 0:67[f] |
| Recessive c.114T > C | TT | 30% (103) | 26% (14) | 30% (117) | 0:58[f] |
| Recessive c.194A > G | AA | 21% (72) | 25% (13) | 22% (85) | 0:57[f] |
| Recessive c.233C > A | CC | 64% (218) | 70% (37) | 65% (255) | 0:39[f] |
| Recessive c.244A > G | AA | 30% (102) | 21% (11) | 29% (113) | 0:17[f] |
| Recessive c.316A > G | AA | 56% (190) | 51% (27) | 55% (217) | 0:53[f] |
| log.IgG-MSP1 | | 2.2 (1.6,3.7) | 2.4 (1.6,3.2) | 2.3 (1.6,3.6) | 0:77[e] |
| log.IgG1-MSP1 | | 3.2 (2.4,5.4) | 3.1 (2.6,4.4) | 3.2 (2.4,5.4) | 0:6[e] |
| log.IgG2-MSP1 | | 1.9 (1.6,2.8) | 2.0 (1.6,2.7) | 1.9 (1.6,2.8) | 0:71[e] |
| log.IgG3-MSP1 | | 3.3 (1.9,6.0) | 3.1 (2.0,6.0) | 3.3 (1.9,6.0) | 0:83[e] |
| log.IgG4-MSP1 | | 1.6 (1.4,2.0) | 1.6 (1.3,1.9) | 1.6 (1.4,2.0) | 0:64[e] |
| log.IgG-MSP3 | | 3.2 (2.5,4.6) | 3.2 (2.5,4.8) | 3.2 (2.5,4.6) | 0:66[e] |
| log.IgG-1MSP3 | | 3.2 (2.6,4.7) | 2.9 (2.6,4.2) | 3.2 (2.6,4.7) | 0:36[e] |
| log.IgG-2MSP3 | | 1.8 (1.6,2.2) | 1.8 (1.5,2.1) | 1.8 (1.6,2.2) | 0:52[e] |
| log.IgG-3MSP3 | | 2.7 (1.8,4.5) | 2.6 (2.0,4.1) | 2.7 (1.8,4.5) | 0:89[e] |
| log.IgG-4MSP3 | | 1.7 (1.4,2.1) | 1.7 (1.5,2.1) | 1.7(1.4,2.1) | 0:82[e] |
| log.IgG-GLURPR0 | | 3.4 (2.4,4.7) | 3.8 (2.6,4.8) | 3.4 (2.4,4.7) | 0:36[e] |
| log.IgG1-GLURPR0 | | 3.4 (2.5,4.9) | 3.7 (2.7,5.1) | 3.4 (2.5,4.9) | 0:40[e] |
| log.IgG2-GLURPR0 | | 1.9(1.6,2.3) | 1.8(1.6,2.9) | 1.9 (1.6,2.4) | 0:61[e] |
| log.IgG3-GLURPR0 | | 2.1 (1.6,3.4) | 2.0 (1.7,2.7) | 2.1 (1.6,3.4) | 0:78[e] |
| log.IgG4-GLURPR0 | | 1.5 (1.3,1.7) | 1.5 (1.2,1.7) | 1.5 (1.3,1.7) | 0:44[e] |
| log.IgG-GLURPR2 | | 3.9 (2.2,5.9) | 4.2 (2.8,6.0) | 4.0 (2.2,5.9) | 0:26[e] |
| log.IgG1-GLURPR2 | | 5.8 (3.8,8.0) | 5.8 (4.3,7.5) | 5.8 (3.9,8.0) | 0:93[e] |
| log.IgG2-GLURPR2 | | 3.2 (2.1,6.3) | 2.9 (2.1,6.4) | 3.1 (2.1,6.3) | 0:56[e] |
| log.IgG3-GLURPR2 | | 5.9 (3.5,7.8) | 6.1 (4.1,7.3) | 5.9 (3.5,7.6) | 0:96[e] |
| log.IgG4-GLURPR2 | | 2.1 (1.8,3.1) | 2.1 (1.8,2.4) | 2.1 (1.8,2.9) | 0:62[e] |
| log.igG-AMA1 | | 6.2 (3.6,9.2) | 6.4 (4.5,8.4) | 6.7 (3.8,9.1) | 0:45[e] |
| log.IgG1-AMA1 | | 9.5 (6.2,10.2) | 8.9 (6.6,10.0) | 9.4 (6.3,10.2) | 0:61[e] |

| | | | | |
|---|---|---|---|---|
| log.IgG2-AMA1 | 3.2 (2.1,4.5) | 2.9 (2.1,4.2) | 3.2 (2.1,4.4) | 0:57[e] |
| log.IgG3-AMA1 | 4.7 (3.1,6.4) | 4.8 (3.2,6.8) | 4.7 (3.2,6.5) | 0:54[e] |
| log.igG4-AMA1 | 4.0 (2.5,5.1) | 3.6 (2.9,4.7) | 3.9 (2.6,5.1) | 0:62[e] |

a (b, c), represent the median, lower quartile, and the upper quartile for continuous variables. $e$-Wilcoxon Ranksum test, $f$-Fishers Exact Test. Numbers after percents are frequencies. Tests used: Wilcoxon test; Pearson test, MSP: Merozoite surface protein, GLURP: Glutamate Rich Protein, AMA: Apical membrane antigen. Note: natural log transformation was used.

## Modeling results on baseline covariates and socio-demographic characteristics

Four different initial models were rigorously assessed and the best model was chosen to serve as the baseline model using the aforementioned indices. These models are presented in Table 2. 'Model.standard' is a linear additive model of age in years, hemoglobin level in (g/dl), sickle cell, blood group, bed net use and parasite density categorized into positive and negative for presence and absence of parasite in blood at enrollment, respectively. The 'Model.spline' is the additive model (Model.standard) but we modeled the nonlinear effect of age and hemoglobin using 5 knots restricted cubic splines, and also included an interaction term of bed net use, parasite density. The 'Model.slope.optimism' (SCFM) is the 'Model.spline' but with regression coefficients shrunk by slope corrected optimism. Finally, we fitted the 'PMLE.model', which is the 'Model.spline' but estimates of the $\beta$ coefficient were based on the Penalized Maximum Likelihood Estimation procedure [25]. Upon careful consideration based on the above model prediction performance measures, the 'PMLE.model' was chosen as the baseline model. The bootstrap cross-validated estimates of AUROC, Brier score, root mean squared error and explained variations in predicted probabilities of the 'PMLE.model' are 50.0%, 10.0%, 0.32 and 17.0%, respectively. Although these performance indices were generally poor, it was better as compared to the other models (Table 2). The probability that PMLE.model will assign a higher predicted risk to a randomly chosen child with malaria compared with a randomly chosen child with no malaria is 50.0% (discrimination ability = 50.0%; Table 2 ) which is not better than random prediction. The Brier score of 10.0% is the expected loss or regret incurred when predicted risk from PMLE.model is issued to a child whose true malaria status is either susceptible or protected. The variations in predicted risk of malaria explained by the chosen model is 17.0% (Table 2).

Table 2. Predictive effect of socio-demographic indices and baseline covariates

| Model specification | LR, p-value | BCV AUROC (%) | BCV BS(%) | RMSE | $R^2$ |
|---|---|---|---|---|---|
| Standard model (S) | 5.34, 0.7209 | 51.0 | 12.0 | 0.35 | 0.01 |
| Spline model | 8.08, 0.8387 | 49.0 | 11.0 | 0.33 | 0.10 |
| Slope corrected optimism model (SCFM) | 7.32,0.7844 | 49.0 | 11.0 | 0.33 | 0.10 |
| PMLE model | 5.76, 0.7360 | 50.0 | 10.0 | 0.32 | 0.17 |

LR: Likelihood Ratio test statistic; BCV: Bootstrap cross-validation; RMSE: Root Mean Squared Error; $R^2$: Proportion of explained variation in predicted risk; AUCROC: Area Under the Receiver Operating Characteristic curve, BS: Brier score.

Furthermore, the calibration plot in Fig 1 shows that the PMLE model underestimates children at low risk of malaria and overestimate children at higher risk of malaria.

**Figure 1. Calibration plots comparing the four baseline models**

## Modeling results on antigen-specific antibodies and *FCGR3B* polymorphisms

The predictive effect of each antibody (IgG and subclasses) and *FCGR3B* was evaluated by introducing them one after the other in the selected baseline penalized maximum likelihood model (PMLE.model) as shown in Table 3. All antibodies were modeled via 5 knots restricted cubic spline. Admittedly, none of the antibodies nor the *FCGR3B* could significantly improve the performance of the PMLE.model after introduction but it was observed that IgGAMA1, IgG1AMA1, and *FCGR3Bc.*233*C* > *A* were better than all the other predictors in relation to their bootstrap cross-validated estimates of AUROC, Brier score, and $R^2$.

Table 3. Assessing the predictive effect of each malaria antibodies and *FCGR3B* polymorphisms on the risk of malaria

| Model specification | LR, p-value | BCV AUROC | BCV BS (%) | RMSE | $R^2$ |
|---|---|---|---|---|---|
| F+log.IgGAMA1 | 12.23, 0.0301 | 55 | 10 | 0.3511 | 2.0 |
| F+log.IgG1AMA1 | 14.09 , 0.0223 | 57 | 10 | 0.3501 | 2.0 |
| F+log.IgG2AMA1 | 5.76, 0.8058 | 49 | 12 | 0.3400 | 1.0 |
| F+log.IgG3AMA1 | 10.41, 0.7668 | 54 | 12 | 0.3525 | 2.0 |
| F+log.IgG4AMA1 | 7.03, 0.7668 | 50 | 12 | 0.3439 | 3.0 |
| F+log.IgGGLURPR0 | 7.41, 0.6501 | 51 | 12 | 0.3514 | 1.0 |
| F+log.IgG1GLURPR0 | 6.93, 0.6987 | 51 | 12 | 0.3518 | 0.0 |
| F+log.IgG2GLURPR0 | 6.38, 0.7528 | 50 | 12 | 0.3512 | 1.0 |
| F+log.IgG3GLURPR0 | 5.99, 0.7872 | 49 | 12 | 0.3532 | 1.0 |
| F+log.IgG4GLURPR0 | 6.13, 0.7716 | 50 | 12 | 0.3517 | 1.0 |
| F+log.IgGGLURPR2 | 7.98, 0.5923 | 53 | 12 | 0.3518 | 1.0 |
| F+log.IgG1GLURPR2 | 8.62, 0.6135 | 52 | 12 | 0.3512 | 1.0 |
| F+log.IgG2GLURPR2 | 6.81, 0.7672 | 49 | 12 | 0.3514 | 1.0 |
| F+log.IgG3GLURPR2 | 6.68, 0.7754 | 50 | 12 | 0.3812 | 1.0 |
| F+log.IgG4GLURPR2 | 6.18, 0.7671 | 50 | 12 | 0.3909 | 2.0 |
| F+log.IgGMSP1 | 6.60, 0.7302 | 51 | 12 | 0.3609 | 0.0 |
| F+log.IgG1MSP1 | 6.03, 0.7812 | 50 | 12 | 0.3512 | 1.0 |
| F+log.IgG2MSP1 | 6.53, 0.7369 | 50 | 12 | 0.3517 | 0.0 |
| F+log.IgG3MSP1 | 5.78, 0.8059 | 49 | 12 | 0.3518 | 0.0 |
| F+log.IgG4MSP1 | 5.81, 0.8032 | 49 | 12 | 0.3547 | 1.0 |
| F+log.IgGMSP3 | 8.08, 0.5816 | 52 | 12 | 0.3558 | 0.0 |
| F+log.IgG1MSP3 | 8.41, 0.6146 | 52 | 12 | 0.3678 | 1.0 |
| F+log.IgG2MSP3 | 5.96, 0.7897 | 50 | 12 | 0.3579 | 2.0 |
| F+log.IgG3MSP3 | 6.68, 0.7217 | 50 | 12 | 0.3510 | 1.0 |
| F+log.IgG4MSP3 | 5.81, 0.8029 | 50 | 12 | 0.3512 | 2.0 |
| F+ Additive c.108C>G | 7.81, 0.6743 | 51 | 12 | 0.3484 | 0.5 |
| F+Additive c.114T>C | 6.3, 0.8108 | 49 | 12 | 0.3501 | 0.5 |
| F+Additive c.194A>G | 7.42, 0.7103 | 50 | 12 | 0.3474 | 1.1 |
| F+Additive c.233C>A | 8.76, 0.5557 | 51 | 12 | 0.3470 | 1.3 |
| F+Additive c.244A>G | 8.51, 0.6066 | 51 | 12 | 0.3464 | 1.6 |
| F+Additive c.316A>G | 6.51, 0.7894 | 50 | 12 | 0.3483 | 0.6 |
| F+Dominant c.108C>G | 7.83, 0.6025 | 52 | 12 | 0.3455 | 2.1 |
| F+Dominant c.114T>C | 5.76, 0.8029 | 50 | 12 | 0.3458 | 2.0 |
| F+Dominant c.194A>G | 7.44, 0.6411 | 51 | 12 | 0.3467 | 1.5 |
| F+Dominant c.233C>A | 8.99, 0.0456 | 58 | 11 | 0.3166 | 2.4 |
| F+Dominant c.244A>G | 5.79, 0.8011 | 49 | 12 | 0.3491 | 0.1 |
| F+Dominant c.316A>G | 6.46, 0.7319 | 51 | 12 | 0.3476 | 1.0 |
| F+Recessive c.108C>G | 6.05, 0.7769 | 49 | 12 | 0.3454 | 2.2 |
| F+Recessive c.114T>C | 6.23, 0.7597 | 49 | 12 | 0.3487 | 0.3 |
| F+Recessive c.194A>G | 6.07, 0.7731 | 49 | 12 | 0.3480 | 0.7 |
| F+Recessive c.233C>A | 6.36, 0.7468 | 50 | 12 | 0.3500 | 0.4 |
| F+Recessive c.244A>G | 8.26, 0.5566 | 52 | 12 | 0.3470 | 1.3 |
| F+Recessive c.316A>G | 6.1, 0.7735 | 49 | 12 | 0.3479 | 0.8 |

F: Penalized maximum likelihood model (PMLE.model), LR: Likelihood Ratio test statistic; BCV: Bootstrap cross-validation; RMSE: Root Mean Squared Error; $R^2$: Proportion of explained variation in predicted risk; AUCROC: Area Under the Receiver Operating Characteristic curve, BS: Brier score.

## Evaluating the joint effect of IgGAMA1, IgG1AMA1 and dominant gene *c.233C > A*

The selection of IgGAMA1, IgG1AMA1, and dominant gene c.233C > A in the subsequent model building were based on the fact that they had a higher AUROC and $R^2$, and a smaller Brier Score as previously shown in Table 3. First, we fitted a model which was basically the baseline model (PMLE.model) together with the three predictors (IgGAMA1, IgG1AMA1 and dominant gene *c.233C > A*) but with no shrinkage adjustment to the regression coefficients. This model was denoted as the final model with no shrinkage (FMNS). Second, we fitted another model with a baseline model (PMLE.model) together with the three predictors (IgGAMA1, IgG1AMA1 and dominant gene *c.233C > A* ) but with shrinkage of the regression coefficient using van Houwelingen-Le Cessie heuristic estimate. This was done to improve the calibration ability of the model. This was denoted as slope corrected optimism with the van Houwelingen-Le Cessie heuristic (SCOV) model. Finally, a model was fitted and evaluated with the only IgGAMA1, IgG1AMA1 and dominant gene c.233C>A. We denote this model as the no baseline line covariate (NBC) model. The predictive performance indices of these three final models were evaluated and the results showed that the NBC model with only IgGAMA1, IgG1AMA1, and dominant gene c.233C>A predicted malaria incidence better as this model had the highest bootstrap cross-validated AUROC and $R^2$ with a smaller Brier score as shown in Table 4.

Table 4. Prediction performance measures of the final selected models

| Model specification | LR, p-value | BCV AUROC (97.5% CI) | BCV Brier score (%) | $R^2$ (%) |
|---|---|---|---|---|
| FMNS | 25.91, 0.0176 | 57.49(45.38-68.38) | 12.10 | 1.00 |
| SCOV | 31.08, 0.0175 | 58.38(45.38 -68.72) | 12.00 | 2.00 |
| NBC | 22.41, 0.0077 | 61.51(48.87-70.71) | 11.72 | 4.00 |

FMNS=PMLE.model+logIgGAMA1+logIgG1AMA1+dominant c.233 with 5 knots restricted cubic spline, Model
SCOV=Slope Corrected final model with van Houwelingen-Le Cessie heuristic estimate.
NBC=No baseline variable included: only dominant c.233 + logIgGAMA1 + logIgG1AMA1
Abbreviations: LR, BCV, RMSE, $R^2$, AUROC, BS, represent the Likelihood Ratio test statistic, bootstrap cross-validation, Root Mean Squared Error, the proportion of explained variation in predicted risk and Area Under the Receiver Operating Characteristic curve, Brier score, respectively.

## Assessing the prediction performance of the three models using calibration plots

The prediction performance of these models was explored by examining calibration plots, that is the plots of sensitivity versus (1-specificity) (AUROC). There was no significant difference in their respective areas under the curve before internal validation, although the model with penalty corrected parameter estimates (SCOV model) had a higher AUROC (75.3%). After internally validating these models, we observed that the AUROC of the model with only IgGAMA1, IgG1AMA1, and dominant gene c.233C>A (NBC model) performed better than the other two models (AUROC=61.5%) ( Fig 2).

**Figure 2. Discrimination and calibration ability of the three final models**

FMNS=Final model with no shrinkage, SCOV=Final model with slope corrected optimism using Van Howelligen estimator, NBV=Model with no baseline variables (only IgGAMA1, IgG1AMA1, and dominant gene c.233C>A); BCV=Bootstrap cross-validation.

The calibration plots in Fig 3 shows that the same model was well calibrated as it appears to be closer to the $45^o$ line compared to the two other models.

**Figure 3. Calibration plots comparing the three final selected models**

In relation to how these models assign the low and higher predicted risk of malaria using the box-whisker plots, there was not much difference between the model with slope corrected optimism and the model with only IgGAMA1, IgG1AMA1, and dominant gene *c.*233*C* > *A* although both were slightly better than the PMLE model updated with 3 more predictors ( Fig 4).

**Figure 4. Evaluating the discrimination ability of the three final selected models.**
Abbreviations: SCOV: slope corrected optimism model, NBC: no baseline covariate model, that is model with only IgGAMA1, IgG1AMA1, and *c.*233*C* > *A* genotype, FMNS: Parameter estimates via penalized maximum likelihood estimation.

## Discussion

The study investigated the predictive performance of several malaria-specific antibodies (IgG and subclasses) and *FCGR3B* polymorphisms on the malaria risk by comparing a baseline malaria model with a prediction model that integrates the antibody and genetic data. Malaria prognosis of a child is an estimate of the child's future malaria risk. The prognosis in this study is based on the child's baseline socio-demographic characteristics, blood group, sickle cell status, the use of bed net, presence of malaria parasite in blood and malaria-specific antibodies and *FCGR3B* genotype. Most of the socio-demographic factors, malaria antigen-specific antibodies and the *FCGR3B* variant used in this study as predictors have been found to be associated with clinical malaria in various malaria seroepidemiological studies [21, 26–32]. The discrimination and calibration performance of the baseline and integrated models measured via AUROC and the Brier score were far less than the generally recommended 80% or more for AUROC and smaller Brier score (closer to zero). Although most of the antibodies did not improve the performance of the baseline model, it is worth noting that AMA1 specific antibodies and *FCGR3B*-c.233C>A under the additive and dominant models of inheritance, could discriminate children of low and higher risk of malaria. Admittedly, the improvement in AUROC and Brier score was not very substantial from the baseline model but they showed signs of improving the performance of the baseline model. This finding was consistent with other studies [29, 32–36] which identified AMA1 as an important blood-stage malaria vaccine candidate.

Traditional statistical methods of estimating odds ratios, relative risk and hazard ratio's in epidemiological studies to assess associations between antibodies, genetic polymorphisms, and clinical malaria may not adequately determine the performance of each for predicting the risk of malaria for a child [19]. An antibody or gene variant associated with protection against malaria does not necessarily imply that it could significantly discriminate between randomly chosen children with a low or high risk of malaria and consequently, would not improve model prediction performance. This may have contributed to the poor predictive performance of most of the antibodies and genotypes and the model as a whole. For classification of children into

the high and low risk of malaria, statistical techniques should be used that directly address classification accuracy (e.g. Brier score, AUROC) rather than traditional regression models for assessing associations (reporting of the odds ratio, relative risk and hazard ratio's for time to event outcomes). Studies that link antibody responses and gene polymorphisms to clinical malaria either control for age or restrict the analysis to individuals who have been exposed to *P. falciparum* [29]. In situations where they have adjusted for most of the predictors of malaria, the predictive performance of certain antibody responses was reduced and only a few remained statistically significant [29]. Other important parameters such as hemoglobin level and sickle cell status should all be controlled for in a prediction model and its predictive performance assessed [28]. The inconsistencies in malaria risk prediction model performance indices may also be due to misclassification of the outcome variable (clinical malaria). There are several different case definitions for clinical malaria based on different parasitemia threshold values and what would have been recorded as a case (malaria) in a particular study may be recorded as control (no malaria) in a different study and vice versa. The number of events (malaria cases) studied is important for prognostic research [37] because there is the risk of overestimating the predictive performance of the model when the number of predictors is much larger than the number of outcome events (malaria). Besides, different prognostic studies have suggested that for each candidate predictor (antibodies, host genes, age, sickle cell, hemoglobin, parasite density, e.t.c) studied, at least 10 events (malaria cases) are required [38–41] but these numbers could be lower in certain circumstances [42]. If for instance the number of cases studied is 200 based on fever with a threshold of 2500 parasite per microliter of blood, then there will be differences in the number of events in a different study where clinical malaria was defined as fever with any level of parasitemia with the latter having a larger number of cases in general to improve the power of the test. So long as malaria case definitions are not clearly defined, there will be inconsistencies in results of which antibody, host genetic factor and or their interactions have higher predictive performance on the malaria risk prediction model. Thus, in practice, different modeling strategies may result in similar or very different results even if they are applied to data from the same study. This study did not quantify the predictive performance of climatic factors on the model performance because these children were exposed to the same climatic conditions over the study period. In other words, children enrolled in the study were clustered around the same endemic area with no significant difference in climatic variables. We recommend that prospective studies may consider including climatic variables in the malaria predictive model in situations where there are climatic differences in the geographic locations of subjects being studied. Not accounting for key predictors of malaria in the prediction model generally results in poor calibration and discrimination performance indices. The environmental or climatic factors that have been found from other studies but were not included in the model to quantify their predictive performance include relative humidity, temperature, rainfall, and vegetation index. These environmental factors could affect malaria incidence and to a larger extent, the performance prediction model. Minimum and maximum temperature affect the development of the malaria parasite and its mosquito vector which are key driving forces of the spread of malaria [43]. The time needed for the parasite to complete its development in the mosquito, decreases to less than 10 days as temperature increases from $21^{\circ}C$ to $27^{\circ}C$, with $27^{\circ}C$ being the optimum [44]. Differentials in the quantity of rainfall also promote the spread of malaria as anopheline mosquitoes breed in water [45, 46]. It is recommended here that prospective studies consider including climatic variables in the malaria predictive model

in situations where there are climatic differences in the geographic locations of subjects being studied.

Other factors that contribute to the incidence of malaria but were not measured to assess their predictive effect may also be of interest. These include drug resistance in parasites, vector species, parasite strain and insecticide resistance in mosquitoes [3, 30, 47–50]. The study's inability to observe these factors may have contributed to poor model performance [51]. The proportion of missing values on covariates was relatively high and can result in loss of statistical power, efficiency and precision of the predicted risk notwithstanding the fact that standard statistical procedures of handling missing data were adhered to. This is particularly so when advanced statistical techniques that handle missing completely at random, missing not at random and missing at random make critical but untestable assumptions about how the data went missing [37]. The proportion of missing observations was common among immunoglobulins and host genetic factors which coincidentally were key predictors of interest in the study. To accurately evaluate the effect of malaria antibodies and host genetic factors on the risk of malaria, steps should be taken to reduce to the barest minimum the proportion of missing values among candidate predictors of malaria risk as this may bias model performance metrics.

## Conclusion

This study has found that the *FCGR3B*-c.233C>A genotype and IgG against AMA1 were relatively better compared to the other antibodies and *FCGR3B* genotypes in classifying or predicting malaria risk among children. That is, Apical Membrane Antigen 1 could be a key potential vaccine antigen while *FCGR3B*-c.233C>A under the additive and dominant models of inheritance were also identified as an important modifier of the effect of malaria protective antibodies. Furthermore, the goal of malaria etiological risk factor studies may be quite different from studies where antibodies and host genes are used in classifying a child into high or low-risk groups. Hence the statistical methods between such studies differ to a large extent. If the latter is required, we recommend the use of model discrimination and calibration indices such as Brier score, RMSE, and AUROC.

## Declarations

### Ethics approval and consent to participate

Ethical approval for the study was given by the institutional review board of the Noguchi Memorial Institute for Medical Research (NMIMR) of the University of Ghana, Accra. Written informed consent was given by the parents and guardians of children before they were enrolled in the study.

### Consent for publication

Written informed consent was obtained from the parents for publication of this research article.

### Availability of data and materials

Data will be made available upon reasonable request.

### Competing interests

There are no competing interests.


## Funding

The study was supported by grants from the Danish Ministry of Foreign Affairs (DFC file no. 14-P01-GHA); African Malaria Network Trust (grant 008/2008AIA), and the European and Developing Countries Clinical Trials Partnership (grant TA.2007.40200.012).


## Authors' contributions

DD1: Conceptualization, data curation, formal analysis, investigation, methodology, writing the original draft
BA: Data curation, investigation, review of the manuscript, designed and conducted field work, conducted laboratory experiments.
DD2: Data curation, investigation, review of the manuscript, designed and conducted field work
MT: Data curation, investigation, review of the manuscript, designed and conducted field work
SI: Formal analysis, investigation, methodology, review of the manuscript
TG: Formal analysis, investigation, methodology, review of the manuscript


## Acknowledgments

We would like to thank the children and their parents and guardians from Asutsuare and its environs who volunteered to participate in the study without whose cooperation this study would have been impossible. The Afro Immuno Assay 2 (AIA2) Field Assistants, Medical Assistants team, nurses, and laboratory Technicians at the Osudoku and Osuwem Community Health Centres are acknowledged for their enormous support during the fieldwork. We thank the research assistants of the Immunology Department and staff of the Transport Department of Noguchi Memorial Institute for Medical Research, Ghana, for both field and laboratory assistance.

## Abbreviations

AMA: Apical Merozoite Antigen
FCR3G: Fc Gamma Receptor III gene
GLURP: Glutamate Rich Protein
IgG: Immunoglobulin G
MSP: Merozoite Surface Proteininsecticide-treated bed nets
PMLE:Penalized Maximum Likelihood Estimate
RMSE: Root Mean Squared Error
AUROC: Area Under the Receiver Operating Characteristic Curve
artemisinin-based combination therapy
IRB:Institutional Review Board
CI:Confidence interval
AIA2: Afro Immuno Assay 2
FMNS=Final model with no shrinkage
SCOV: slope corrected optimism model
NBC: no baseline covariate model,
NBV=Model with no baseline variables
BCV=Bootstrap cross-validation.

## References

1. Organization WH. World malaria report 2015. World Health Organization; 2016.
2. Hviid L. Clinical disease, immunity and protection against Plasmodium falciparum malaria in populations living in endemic areas. Expert Rev Mol Med. 1998;1998:1–10. doi:10.1017/S1462399498000179.
3. Ahmed Ismail H, Ribacke U, Reiling L, Normark J, Egwang T, Kironde F, et al. Acquired antibodies to merozoite antigens in children from Uganda with uncomplicated or severe Plasmodium falciparum malaria. Clin Vaccine Immunol. 2013;20:1170–80. doi:10.1128/CVI.00156-13.
4. Baker SG, Kramer BS, Srivastava S. Markers for early detection of cancer: statistical guidelines for nested case-control studies. BMC Med Res Methodol. 2002;2:4. doi:10.1186/1471-2288-2-4.
5. Boyko EJ, Alderman BW. The use of risk factors in medical diagnosis: opportunities and cautions. J Clin Epidemiol. 1990;43:851–8.

526 6. Kattan MW. Judging new markers by their ability to improve predictive accuracy. 2003.
527 7. Emir B, Wieand S, Su JQ, Cha S. Analysis of repeated markers used to predict progression of
528 cancer. Stat Med. 1998;17:2563–78.
529 8. Gail MH, Costantino JP. Validating and improving models for projecting the absolute risk of
530 breast cancer. J Natl Cancer Inst. 2001;93:334–5. doi:10.1093/jnci/93.5.334.
531 9. Rockhill B, Spiegelman D, Byrne C, Hunter DJ, Colditz GA. Validation of the Gail et al. model
532 of breast cancer risk prediction and implications for chemoprevention. J Natl Cancer Inst.
533 2001;93:358–66. doi:10.1093/jnci/93.5.358.
534 10. Park Y, Freedman AN, Gail MH, Pee D, Hollenbeck A, Schatzkin A, et al. Validation of a
535 colorectal cancer risk prediction model among white patients age 50 years and older. J Clin Oncol.
536 2009;27:694.
537 11. Moons KGM, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, et al. Risk
538 prediction models: II. External validation, model updating, and impact assessment. Heart.
539 2012;98:691–8.
540 12. Ridker PM, Paynter NP, Rifai N, Gaziano JM, Cook NR. C-reactive protein and parental
541 history improve global cardiovascular risk prediction: the Reynolds Risk Score for men.
542 Circulation. 2008;118:2243--51, 4p following 2251.
543 doi:10.1161/CIRCULATIONAHA.108.814251.
544 13. Wilson PWF, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction
545 of coronary heart disease using risk factor categories. Circulation. 1998;97:1837–47.
546 14. Etzel CJ, Kachroo S, Liu M, D'Amelio A, Dong Q, Cote ML, et al. Development and validation
547 of a lung cancer risk prediction model for African-Americans. Cancer Prev Res. 2008;1:255–65.
548 15. Folsom AR, Chambless LE, Ballantyne CM, Coresh J, Heiss G, Wu KK, et al. An assessment
549 of incremental coronary risk prediction using C-reactive protein and other novel risk markers: the
550 atherosclerosis risk in communities study. Arch Intern Med. 2006;166:1368–73.
551 16. Tice JA, Cummings SR, Ziv E, Kerlikowske K. Mammographic breast density and the Gail
552 model for breast cancer risk prediction in a screening population. Breast Cancer Res Treat.
553 2005;94:115–22.
554 17. Hubert HB, Feinleib M, McNamara PM, Castelli WP. Obesity as an independent risk factor
555 for cardiovascular disease: a 26-year follow-up of participants in the Framingham Heart Study.
556 Circulation. 1983;67:968–77.
557 18. Benjamin EJ, Levy D, Vaziri SM, D'Agostino RB, Belanger AJ, Wolf PA. Independent risk
558 factors for atrial fibrillation in a population-based cohort: the Framingham Heart Study. Jama.
559 1994;271:840–4.
560 19. Pepe MS, Janes H, Longton G, Leisenring W, Newcomb P. Limitations of the odds ratio in
561 gauging the performance of a diagnostic, prognostic, or screening marker. Am J Epidemiol.
562 2004;159:882–90.
563 20. Adu B, Jepsen MPG, Gerds TA, Kyei-Baafour E, Christiansen M, Dodoo D, et al. Fc gamma
564 receptor 3B (FCGR3B-c. 233C>A-rs5030738) polymorphism modifies the protective effect of
565 malaria specific antibodies in Ghanaian children. J Infect Dis. 2013;209:285–9.
566 21. Adu B, Dodoo D, Adukpo S, Hedley PL, Arthur FKN, Gerds TA, et al. Fc gamma receptor
567 IIIB (Fc$γ$RIIIB) polymorphisms are associated with clinical malaria in Ghanaian children. PLoS
568 One. 2012;7:e46197.
569 22. Gerds TA, Cai T, Schumacher M. The performance of risk prediction models. Biom J.
570 2008;50:457–79. doi:10.1002/bimj.200810443.

571 23. Harrell Jr FE, Dupont MC, Hmisc D. The design package. R Packag version. 2007;2.
572 24. Goeman JJ. L1 penalized estimation in the Cox proportional hazards model. Biom J.
573 2010;52:70–84. doi:10.1002/bimj.200900028.
574 25. Verweij PJ, Van Houwelingen HC. Penalized likelihood in Cox regression. Stat Med.
575 1994;13:2427–36. doi:10.1002/sim.4780132307.
576 26. Mockenhaupt FP, Ehrhardt S, Cramer JP, Otchwemah RN, Anemana SD, Goltz K, et al.
577 Hemoglobin C and resistance to severe malaria in Ghanaian children. J Infect Dis. 2004;190:1006–
578 9.
579 27. Williams TN, Mwangi TW, Wambua S, Alexander ND, Kortok M, Snow RW, et al. Sickle
580 cell trait and the risk of Plasmodium falciparum malaria and other childhood diseases. J Infect Dis.
581 2005;192:178–86. doi:10.1086/430744.
582 28. Aidoo M, Terlouw DJ, Kolczak MS, McElroy PD, ter Kuile FO, Kariuki S, et al. Protective
583 effects of the sickle cell gene against malaria morbidity and mortality. Lancet. 2002;359:1311–2.
584 doi:10.1016/S0140-6736(02)08273-9.
585 29. Greenhouse B, Ho B, Hubbard A, Njama-Meya D, Narum DL, Lanar DE, et al. Antibodies to
586 Plasmodium falciparum antigens predict a higher risk of malaria but protection from symptoms
587 once parasitemic. J Infect Dis. 2011;204:19–26. doi:10.1093/infdis/jir223.
588 30. Tiendrebeogo RW, Adu B, Singh SK, Dziegiel MH, Nébié I, Sirima SB, et al. Antibody-
589 dependent cellular inhibition is associated with reduced risk against febrile malaria in a
590 longitudinal cohort study involving Ghanaian children. In: Open Forum Infect Dis. Oxford
591 University Press; 2015.
592 31. Polley SD, Conway DJ, Cavanagh DR, McBride JS, Lowe BS, Williams TN, et al. High levels
593 of serum antibodies to merozoite surface protein 2 of Plasmodium falciparum are associated with
594 reduced risk of clinical malaria in coastal Kenya. Vaccine. 2006;24:4233–46.
595 doi:10.1016/j.vaccine.2005.06.030.
596 32. Mitchell GH, Thomas AW, Margos G, Dluzewski AR, Bannister LH. Apical membrane
597 antigen 1, a major malaria vaccine candidate, mediates the close attachment of invasive merozoites
598 to host red blood cells. Infect Immun. 2004;72:154–8. doi:10.1128/iai.72.1.154-158.2004.
599 33. Remarque EJ, Faber BW, Kocken CHM, Thomas AW. Apical membrane antigen 1: a malaria
600 vaccine candidate in review. Trends Parasitol. 2008;24:74–84.
601 34. Malkin EM, Diemert DJ, McArthur JH, Perreault JR, Miles AP, Giersing BK, et al. Phase 1
602 clinical trial of apical membrane antigen 1: an asexual blood-stage vaccine for Plasmodium
603 falciparum malaria. Infect Immun. 2005;73:3677–85. doi:10.1128/IAI.73.6.3677-3685.2005.
604 35. Polley SD, Mwangi T, Kocken CHM, Thomas AW, Dutta S, Lanar DE, et al. Human antibodies
605 to recombinant protein constructs of Plasmodium falciparum Apical Membrane Antigen 1
606 (AMA1) and their associations with protection from malaria. Vaccine. 2004;23:718–28.
607 36. Stowers AW, Kennedy MC, Keegan BP, Saul A, Long CA, Miller LH. Vaccination of
608 monkeys with recombinant Plasmodium falciparum apical membrane antigen 1 confers protection
609 against blood-stage malaria. Infect Immun. 2002;70:6961–7. doi:10.1128/iai.70.12.6961-
610 6967.2002.
611 37. Moons KGM, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic
612 research: what, why, and how? Bmj. 2009;338:b375.
613 38. Harrell FE, Lee KL, Mark DB. Tutorial in biostatistics multivariable prognostic models: issues
614 in developing models, evaluating assumptions and adequacy, and measuring and reducing errors.
615 Stat Med. 1996;15:361–87.

39. Laupacis A, Sekar N, Stiell IG. Clinical prediction rules. A review and suggested modifications of methodological standards. Jama. 1997;277:488–94. http://www.ncbi.nlm.nih.gov/pubmed/9020274.
40. Peduzzi P, Concato J, Feinstein AR, Holford TR. Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates. J Clin Epidemiol. 1995;48:1503–10. doi:10.1016/0895-4356(95)00048-8.
41. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol. 1996;49:1373–9. doi:10.1016/s0895-4356(96)00236-3.
42. Vittinghoff E, McCulloch CE. Relaxing the rule of ten events per variable in logistic and Cox regression. Am J Epidemiol. 2007;165:710–8. doi:10.1093/aje/kwk052.
43. Paaijmans KP, Blanford S, Bell AS, Blanford JI, Read AF, Thomas MB. Influence of climate on malaria transmission depends on daily temperature variation. Proc Natl Acad Sci U S A. 2010;107:15135–9. doi:10.1073/pnas.1006422107.
44. Gosoniu L, Vounatsou P, Sogoba N, Maire N, Smith T. Mapping malaria risk in West Africa using a Bayesian nonparametric non-stationary model. Comput Stat Data Anal. 2009;53:3358–71.
45. Devi NP, Jauhari RK. Climatic variables and malaria incidence in Dehradun, Uttaranchal, India. J Vector Borne Dis. 2006;43:21–8. http://www.ncbi.nlm.nih.gov/pubmed/16642782.
46. Arab A, Jackson MC, Kongoli C. Modelling the effects of weather and climate on malaria distributions in West Africa. Malar J. 2014;13:126. doi:10.1186/1475-2875-13-126.
47. Magowan C, Wollish W, Anderson L, Leech J. Cytoadherence by Plasmodium falciparum-infected erythrocytes is correlated with the expression of a family of variable proteins on infected erythrocytes. J Exp Med. 1988;168:1307–20.
48. Gardner JP, Pinches RA, Roberts DJ, Newbold CI. Variant antigens and endothelial receptor adhesion in Plasmodium falciparum. Proc Natl Acad Sci U S A. 1996;93:3503–8. doi:10.1073/pnas.93.8.3503.
49. Cherif MK, Sanou GS, Maiga B, Israelsson E, Ouédraogo AL, Bougouma EC, et al. Fc$γ$RIIa Polymorphism and Anti-Malaria-Specific IgG and IgG Subclass Responses in Populations Differing in Susceptibility to Malaria in Burkina Faso. Scand J Immunol. 2012;75:606–13.
50. Stanisic DI, Fowkes FJ, Koinari M, Javati S, Lin E, Kiniboro B, et al. Acquisition of antibodies against Plasmodium falciparum merozoites and malaria immunity in young children and the influence of age, force of infection, and magnitude of response. Infect Immun. 2015;83:646–60. doi:10.1128/IAI.02398-14.
51. Struthers CA, Kalbfleisch JD. Misspecified proportional hazard models. Biometrika. 1986;73:363–9.

# Figures

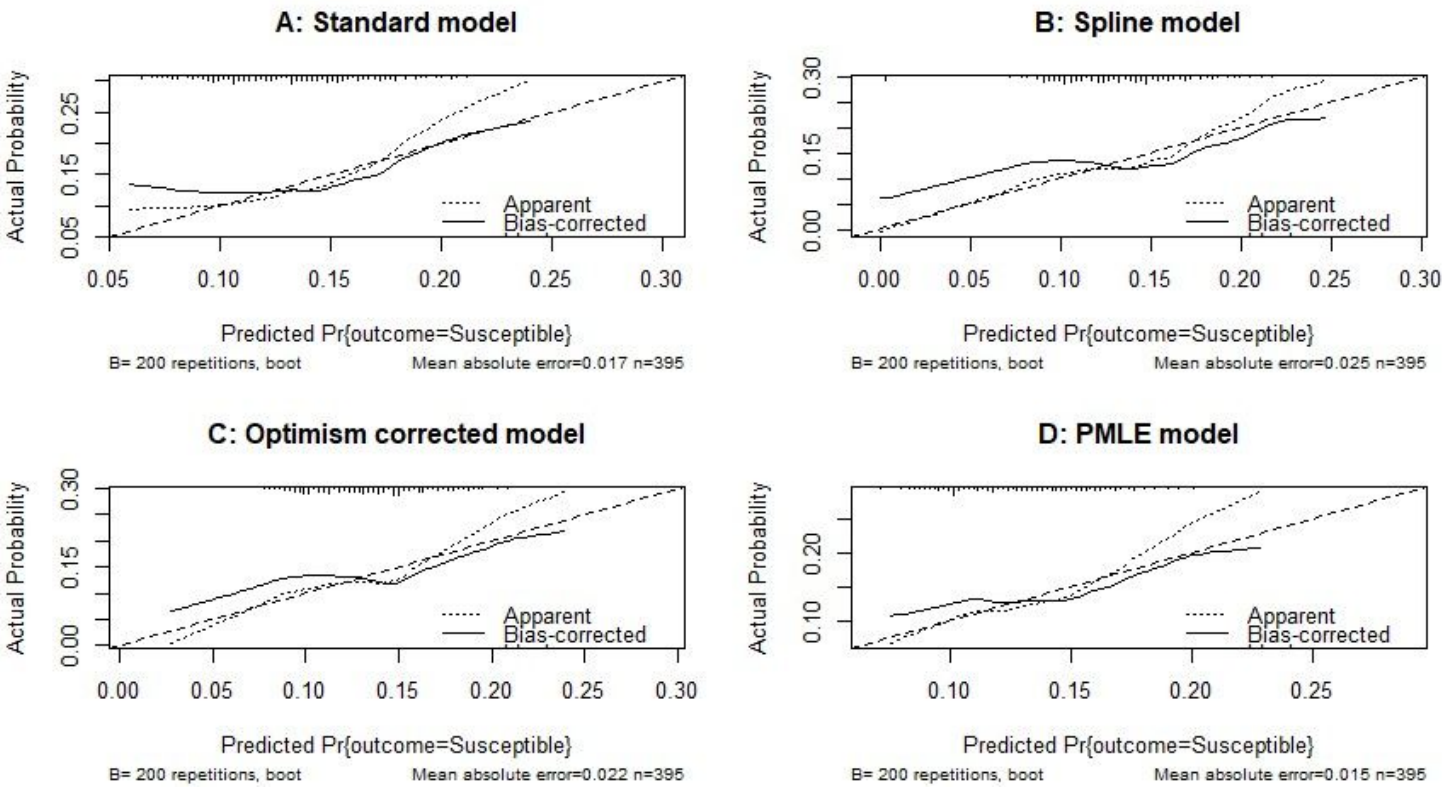



**Figure 1**

Calibration plots comparing the four baseline models

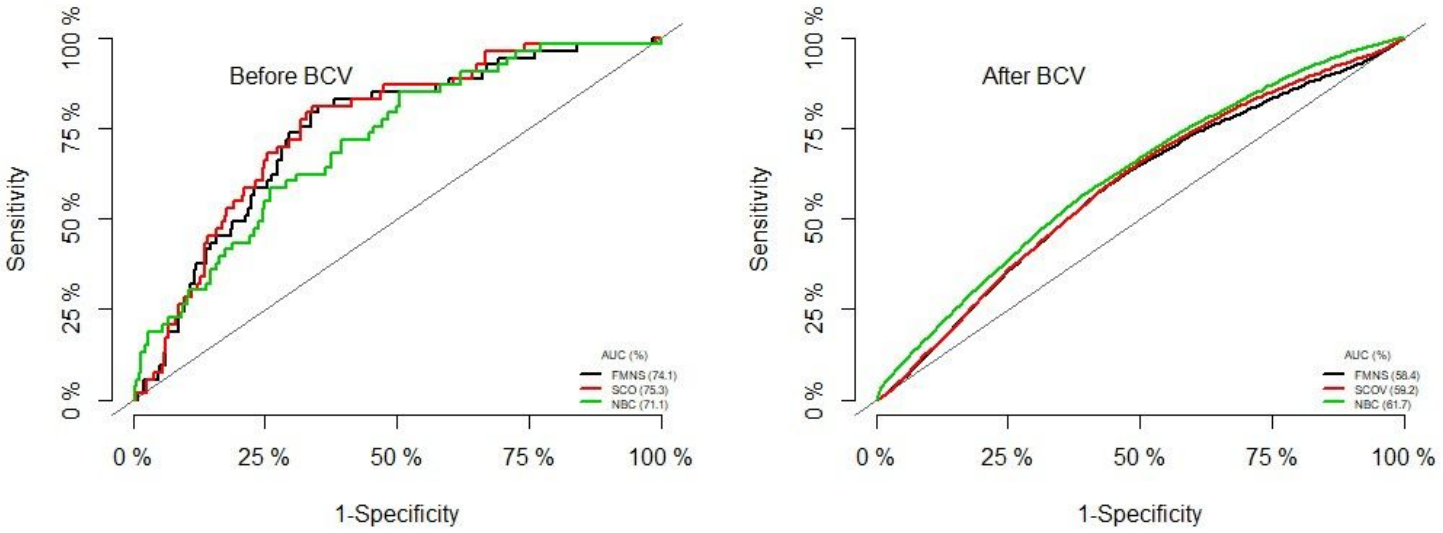
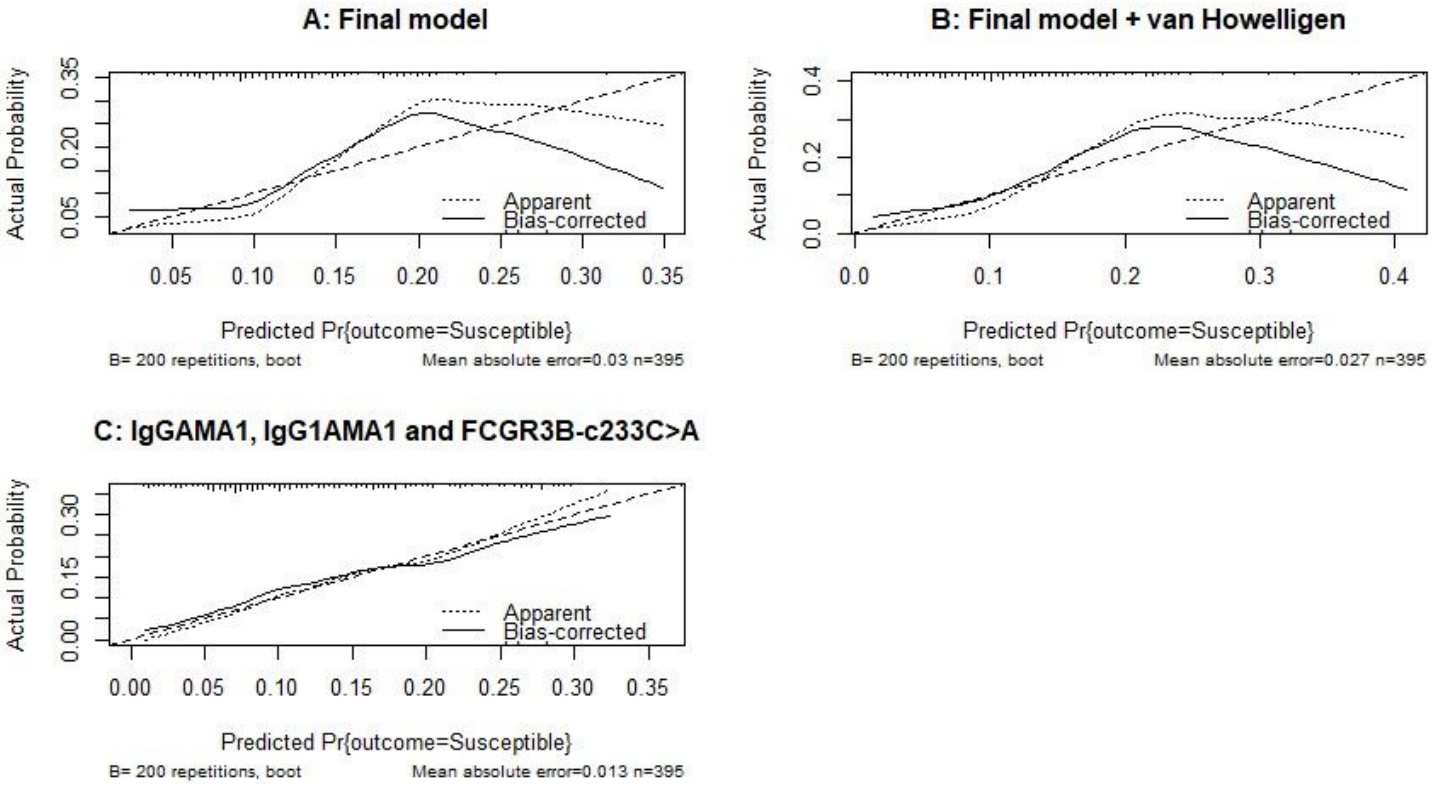


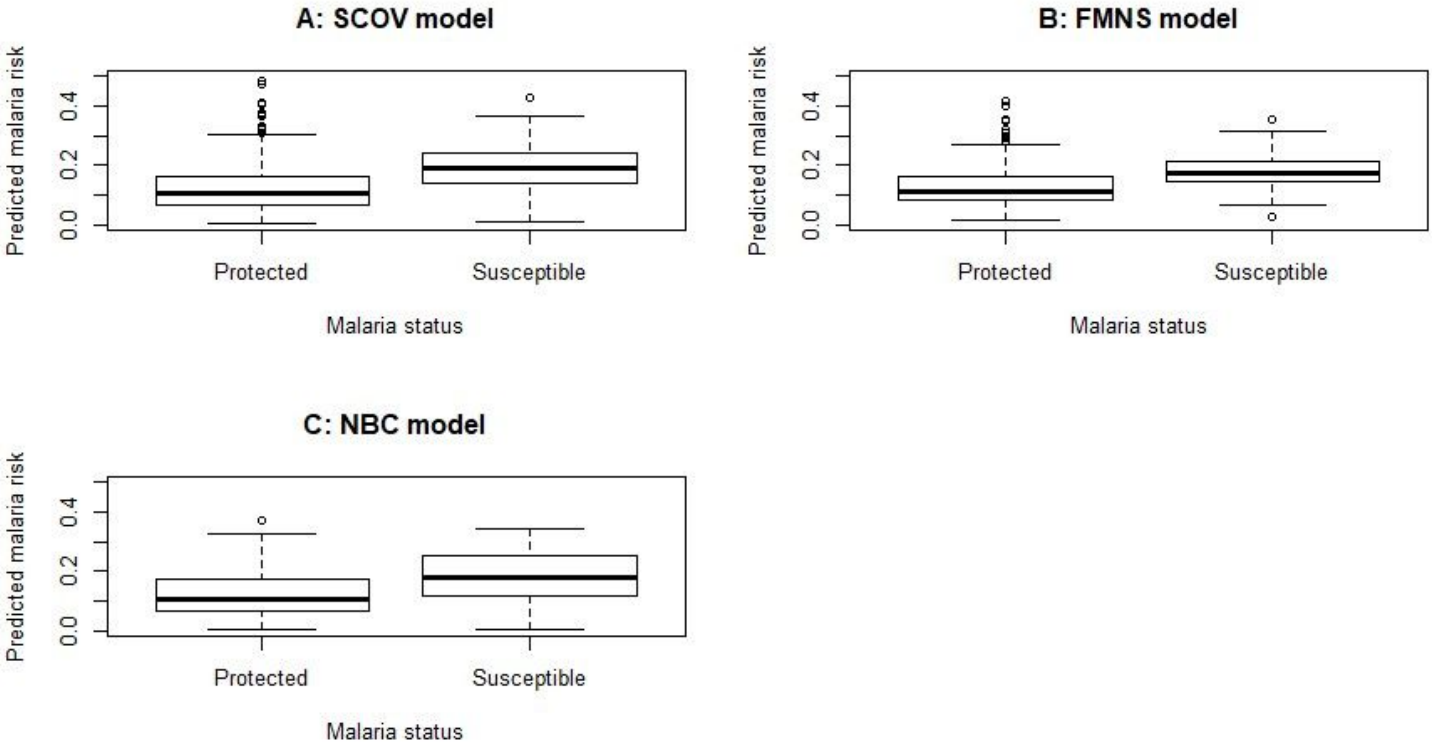

**Figure 2**

Discrimination and calibration ability of the three final models




## Figure 3

Calibration plots comparing the three final selected models



## Figure 4

Evaluating the discrimination ability of the three final selected models. Abbreviations: SCOV: slope corrected optimism model, NBC: no baseline covariate model, that is model with only IgGAMA1, IgG1AMA1, and c.233C > A genotype, FMNS: Parameter estimates via penalized maximum likelihood estimation.