

# Characterising effect of anaemia on mortality in severe malaria

## Contents

Background	1
Exploratory analysis	2
Predictive value of anaemia on death adjusting for confounders	2

## Background

This looks at the severe malaria legacy dataset from MORU

```
library(lme4)
```

```
## Loading required package: Matrix
```

```
# For the GAM modelling
```

```
library(mgcv)
```

```
## Loading required package: nlme
```

```
##
```

```
## Attaching package: 'nlme'
```

```
## The following object is masked from 'package:lme4':
```

```
##
```

```
##      lmList
```

```
## This is mgcv 1.8-22. For overview type 'help("mgcv-package")'.
```

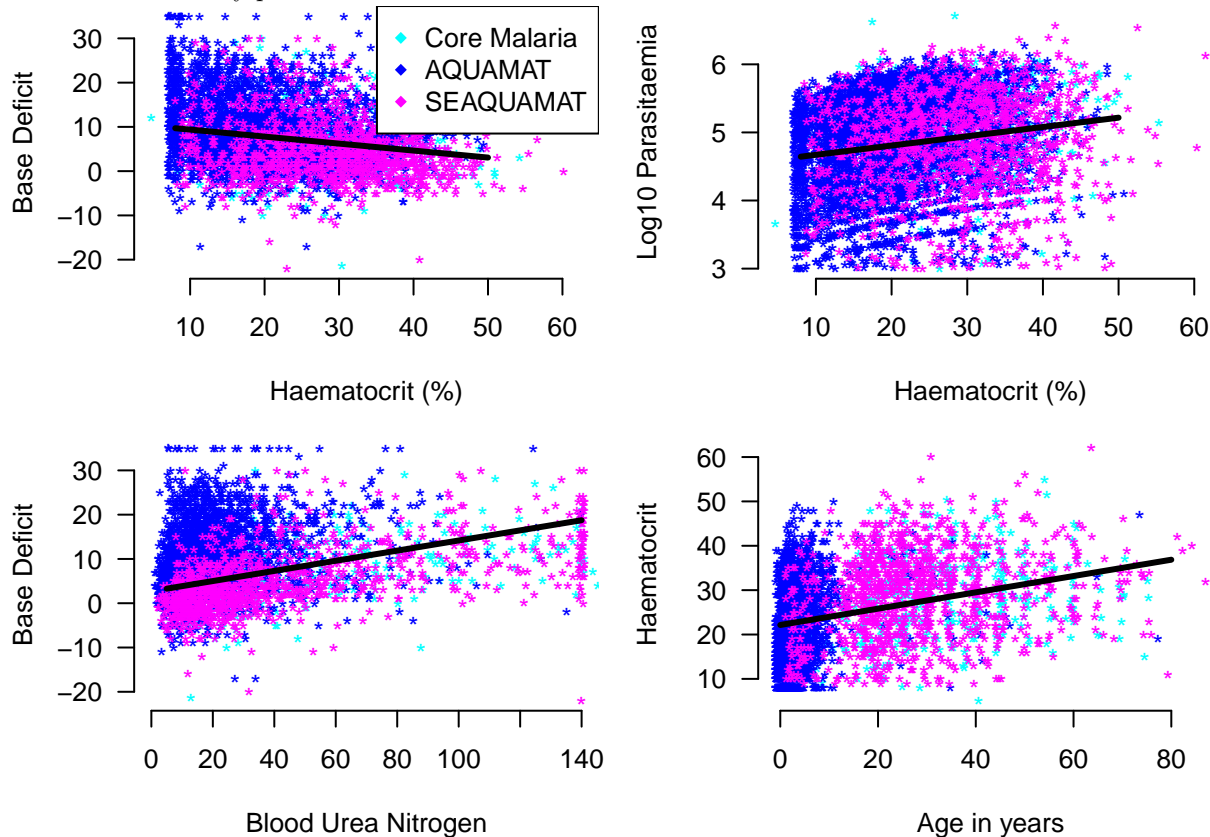
```
# For the CART modelling
```

```
library(rpart)
```

```
library(rpart.plot)
```

## Exploratory analysis

Let's look at the key predictive variables. We use a random effects term to model differences between studies.



## Predictive value of anaemia on death adjusting for confounders

Before fitting the more complex GAM models we explore the standard glm (logistic regression) models.

```
mod_full = glmer(outcome ~ HCT + LPAR + AgeInYear + BUN + BD + (1 | studyID),
  data=Leg_data, family=binomial)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, : Model is nearly unidentifiable:
## - Rescale variables?
```

```
summary(mod_full)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: binomial ( logit )
## Formula: outcome ~ HCT + LPAR + AgeInYear + BUN + BD + (1 | studyID)
## Data: Leg_data
##
##      AIC      BIC    logLik deviance df.resid
##  3214.5   3260.4  -1600.2   3200.5     5221
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -3.0755 -0.3426 -0.2344 -0.1681  9.1819
##
## Random effects:
##   Groups   Name      Variance Std.Dev.
## studyID (Intercept) 0.05531  0.2352
## Number of obs: 5228, groups:  studyID, 3
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -4.216794   0.392658 -10.739 < 2e-16 ***
## HCT          0.019182   0.005440   3.526 0.000422 ***
## LPAR         0.002291   0.066768   0.034 0.972622
## AgeInYear    0.021372   0.004786   4.465 8.00e-06 ***
## BUN          0.011738   0.001711   6.860 6.87e-12 ***
## BD           0.135063   0.006930  19.491 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr) HCT    LPAR   AgInYr BUN
## HCT          -0.241
## LPAR         -0.759 -0.171
## AgeInYear    -0.269 -0.136  0.030
## BUN          -0.115  0.074 -0.047 -0.099
## BD           -0.120  0.254 -0.140  0.070 -0.271
## convergence code: 0
## Model is nearly unidentifiable: very large eigenvalue
## - Rescale variables?
```

Now let's make counterfactual predictions of anaemia on death for the patients in the database:

```
overall_mortality = 100*mean(Leg_data$outcome)
par(las=1)
plot(NA,NA, xlim=c(4,45), ylim=c(0,50),ylab='% predicted mortality', xlab='Haematocrit (%)')
title('Counterfactual predictions of mortality')
for(HCT in 4:45){
  mydata = Leg_data
  mydata$HCT=HCT
  ys = 100*predict(mod_full, newdata = mydata, re.form=NA, type='response')

  points(HCT,mean(ys), pch=18)
  points(rep(HCT,2), quantile(ys, probs=c(0.1,0.9)), pch='-', col='red')
}
abline(h=overall_mortality, lwd=3, col='blue')
legend('topleft', col=c('black','red','blue'), lwd=3, lty=1,
      legend = c('Mean predicted mortality', '80% predicted interval','Observed mortality'))
```

## Counterfactual predictions of mortality

