

Feeding rate consistency

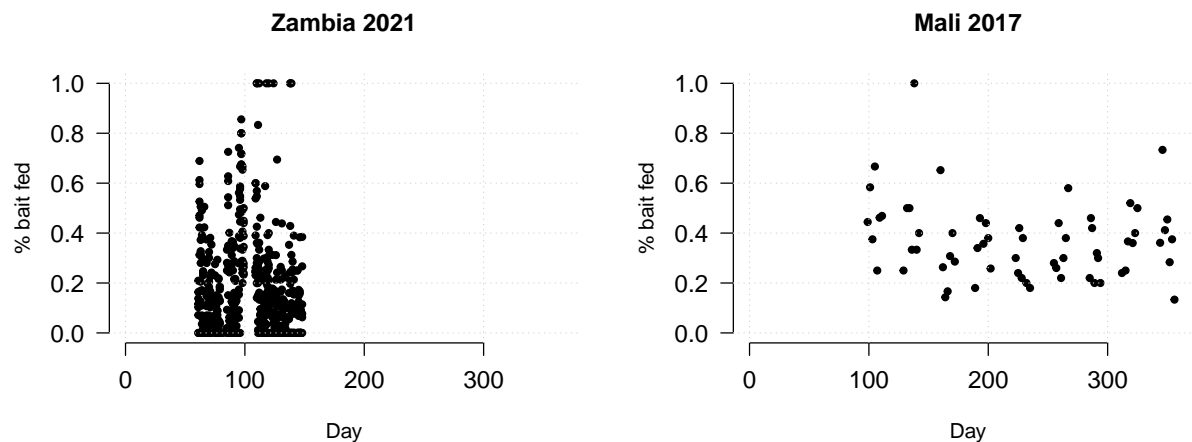
Nima Moghaddas

2023-05-29

```
library(malariasimulation)
library(lme4)
library(dplyr)
```

We are interested to know whether feeding rates change over time, which would provide reason for including a time varying feeding rate in the final simulations. One way to do this is to fit a generalised linear mixed effects model with a random effect on cluster/village and a random effect on date. If the random effect for date is greater than cluster, this provides evidence that there is consistent temporal variation in feeding rates over and above the variation between clusters. Here is the cluster level feeding rate data for Zambia and Mali.

```
par(las=1, mfrow=c(1,2))
# The Zambia trial has data for every mosquito caught so first we need to summarise across cluster and
zambia |>
  group_by(collection_date, cluster) |>
  summarise(dyed_fraction = sum(positive)/n()) -> zambia_grouped
zambia_grouped$days <- (month(zambia_grouped$collection_date)-1)*30 + day(zambia_grouped$collection_date)
plot(zambia_grouped$days,
     zambia_grouped$dyed_fraction,
     cex=1.1, pch=20, frame.plot = F, xlab = "Day", ylab = "% bait fed",
     ylim = c(0,1), cex.axis=1.2, xlim = c(1,365))
grid()
title("Zambia 2021")
# The Mali trial only has data for total stained and unstained mosquitoes by village and date
plot(mali$days,
     mali$dyed_fraction,
     cex=1.1, pch=20, frame.plot = F, xlab = "Day", ylab = "% bait fed",
     ylim = c(0,1), cex.axis=1.2, xlim = c(1,365))
grid()
title("Mali 2017")
```



As you can see, data was collected much more frequently in Zambia than Mali. In fact the Zambia data has information about every mosquito that was caught and whether it was bait fed. Whereas the Mali data only contains summary data for the total number of mosquitoes that were bait fed out of the total sample that was tested. However, the Mali data was collected over a longer period. We can fit a GLMM with random effects on cluster and day. Here are the results for Zambia.

```
fit <- glmer(
  positive ~ (1|cluster) + (1|days),
  data = zambia, family = "binomial", control = glmerControl(optimizer = "bobyqa")
)
summary(fit)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: binomial ( logit )
## Formula: positive ~ (1 | cluster) + (1 | days)
## Data: zambia
## Control: glmerControl(optimizer = "bobyqa")
##
##      AIC      BIC   logLik deviance df.resid
## 38502.9 38528.9 -19248.4 38496.9    42738
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.6014 -0.5003 -0.3449 -0.2286  6.7657
##
## Random effects:
## Groups Name Variance Std.Dev.
## days (Intercept) 0.74511 0.8632
## cluster (Intercept) 0.06173 0.2484
## Number of obs: 42741, groups: days, 65; cluster, 10
##
## Fixed effects:
## Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.7368 0.1367 -12.7 <2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The random effect variance for days is greater than the cluster variance suggesting the bait feeding rate consistently varies more over time than between clusters. Since the dataset includes other covariates, we can also include them in the model. For example here we include household as a random effect.

```
fit1 <- glmer(
  positive ~ (1|cluster) + (1|days) + (1|hh),
  data = zambia, family = "binomial", control = glmerControl(optimizer = "bobyqa")
)
summary(fit1)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: binomial ( logit )
## Formula: positive ~ (1 | cluster) + (1 | days) + (1 | hh)
## Data: zambia
## Control: glmerControl(optimizer = "bobyqa")
##
##      AIC      BIC   logLik deviance df.resid
## 33954.1 33988.8 -16973.1 33946.1    42737
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.6118 -0.3831 -0.2082 -0.1253  8.1713
##
## Random effects:
## Groups Name      Variance Std.Dev.
## hh      (Intercept) 3.08738  1.7571
## days    (Intercept) 1.23203  1.1100
## cluster (Intercept) 0.06789  0.2606
## Number of obs: 42741, groups: hh, 7654; days, 65; cluster, 10
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -2.612      0.169  -15.45  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The majority of the variance in bait feeding rate is explained by household, but time still explains more of the variance than cluster.

Here are the results for Mali.

```
mali$Village <- factor(mali$Village)
fit3 <-
  glmer(
    cbind(total_sampled, total_asb_positive) ~
      (1|Village) + (1|days),
    family = "binomial", data = mali,
    control = glmerControl(optimizer = "bobyqa"))
summary(fit3)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
```

```

## Family: binomial ( logit )
## Formula: cbind(total_sampled, total_asb_positive) ~ (1 | Village) + (1 |
##   days)
## Data: mali
## Control: glmerControl(optimizer = "bobyqa")
##
##      AIC      BIC   logLik deviance df.resid
##    307.3    313.8   -150.7    301.3      60
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.8073 -0.6908 -0.1722  0.5609  2.3049
##
## Random effects:
##  Groups   Name                Variance Std.Dev.
##  days      (Intercept)  0.016203  0.12729
##  Village   (Intercept)  0.004813  0.06937
## Number of obs: 63, groups:  days, 63; Village, 7
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.08963    0.05393   20.21  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

There is much less variance in the Mali data but still we see the variance in time is greater than the variance between villages.