Supplementary Material –Equations and Example Code

One Year - Individual Level (SAS)

Let Y_{hijk} be the number of malaria episodes for subjects k within village j within treatment i within year h and let λ_{hijk} be its expectation.

$log(\lambda_{ijk}) = \alpha + \beta_1 X_i + \beta_2 X_{ijk} + u_{ij}$, where $X_i = 0$ (control) or $X_i = 1$ (Ivermectin), and $X_{ijk} = 0$ (female) or $X_{ijk} = 1$ (male) and:

$$\begin{aligned} \alpha &= \text{Intercept} \\ \beta_1 &= \text{Log of the rate ratio treatment vs control} \\ \beta_2 &= \text{Log of the rate ratio males vs females} \\ u_{ij} &= \text{Denotes the random village effect} \end{aligned}$$

- **IL1**

```
* Model 1 - PL no sample size correction ;
proc glimmix data=WORK.DF pconv= 1e-5 ;
class Village_code TRT(ref = "0") ;
Model Episodes = TRT / s dist=Poisson;
Random  INT / subject= village_code;
```

- **IL2**

```
* Model 2 - Laplace;
proc glimmix data=work.DF method=LAPLACE pconv= 1e-5  ;
class Village_code TRT(ref = "0") ;
Model Episodes= TRT / s dist=Poisson;
Random  INT / subject= village_code;
run;
```

- **IL3**

```
* Model 3 - Kenward Roger;
proc glimmix data=work.DF pconv= 1e-5 ;
class Village_code TRT(ref = "0") ;
Model Episodes = TRT/ s dist=Poisson ddfm=kenwardroger;
Random  INT / subject= village_code;
```

- **IL4**

```
* Model 4 - Adaptive Gaussian Quadrature;
proc glimmix data=work.DF method=quad(qpoints = 10) pconv= 1e-5
;
class Village_code TRT(ref = "0") ;
Model Episodes = TRT/ s dist=Poisson ;
Random  INT / subject= village_code;
```

- **IL5**

```
* Model 5 - Model SE
proc glimmix data=work.DF pconv= 1e-5;
class Village_code TRT(ref = "0") ;
Model Episodes = TRT/ s dist=Poisson ;
Random  residual / subject = village_code type=CS;
```

- **IL6**

```
* Model 6 - Empirical SE
proc glimmix data=work.DF empirical pconv= 1e-5;
class Village_code TRT(ref = "0") ;
Model Episodes = TRT/ s dist=Poisson ;
Random  residual / subject = village_code type=CS;
```

- **IL7**

```
* Model 7 - MBN;
proc glimmix data=work.DF empirical = mbn pconv= 1e-5;
class Village_code TRT(ref = "0") ;
Model Episodes = TRT/ s dist=Poisson ;
Random  residual / subject = village_code type=CS;
```

- **IL8**

```
* Model 8 - FG;
proc glimmix data=work.DF empirical = firoeeq pconv= 1e-5;
class Village_code TRT(ref = "0") ;
Model Episodes = TRT/ s dist=Poisson ;
Random  residual / subject = village_code type=CS;
```

One year - Cluster Level (R)

Note: Generate cluster means ("mean"), the total number of episodes ("episodes"), and the number of subjects per cluster ("n") first. DF refers to the imported data frame.

$$\begin{aligned} \overline{R}_{ij} &= \frac{\sum Episodes_k}{m_j} \\ Episodes_k &= \text{Number of episodes for subject k} \\ m_j &= \text{Number of subjects in cluster j} \end{aligned}$$

$log(\overline{R}_{ij}) = \alpha + \beta_1 X_{ij}$, where $X_i = 0$ (control) or $X_i = 1$ (Ivermectin) and:

$$\begin{aligned} \alpha &= \text{Intercept} \\ \beta_1 &= \text{Log of the rate ratio treatment vs control} \end{aligned}$$

- **CL1**

Unweighted t-test

```
glm(mean ~ TRT, data = DF,family=gaussian(link="log")
```

- **CL2**

Variance Weighted

Note: Same model described above but weights have been generated using between and within cluster variation calculated with the ICC package in R[1]; weights calculated as previously described [16].

```
glm(mean ~ TRT, data = DF, weights = w,
family=gaussian(link="log"))
```

- **CL3**

Size Weighted Gaussian

Note: Same model as above but utilizes an offset where the outcome is the sum of episodes in a given cluster and the offset, m, accounts for the number of subjects in that given cluster.

```
glm(episodes ~ TRT + offset(log(m)),family=gaussian(link="log"),
data = DF)
```

- **CL4**

Adjusted Residual

Note: a GLM has been fit to individual level data and residuals are extracted (deviance by default in R)

```
Fit GLM without treatment effect:
glm(Episodes_Year1 ~ Male, data = DF, family = poisson)
```

Note: Extract residuals and summarize cluster level means, fit this model:

```
glm(mean.resids ~ TRT , data = DF)
```

2 Year Parallel – Individual Level Models (SAS)

Time as Fixed Effect, Village as Random Effect

$log(\lambda_{ijk}) = \alpha + \beta_1 X_i + \beta_2 X_{ijk} + \beta_3 X_{hijk} + u_{ij}$, where $X_i = 0$ (control) or $X_i = 1$ (Ivermectin), $X_{ijk} = 0$ (female) or $X_{ijk} = 1$ (male) and $X_{hijk} = 1$ (Year 1) or $X_{hijk} = 2$ (Year 2) and:

α = Intercept
β_1 = Log of the rate ratio treatment vs control
β_2 = Log of the rate ratio males vs females
β_3 = Log of the rate ratio Year 1 vs Year 2
u_{ij} = Denotes the random village effect

- **IL9**

```
* Model 1 - QL no sample size correction ;
proc glimmix data=WORK.import pconv= 1e-5 ;
class Village_code Treatment(ref = "0") YEAR Male(ref = "0");
Model Episodes = Treatment  Year Male/ s dist=Poisson;
Random  INT / subject= Village_code;
```

- **IL10**

```
* Model 2 - Laplace;
proc glimmix data=WORK.import method=LAPLACE pconv= 1e-5  ;
class Village_code Treatment(ref = "0") YEAR Male(ref = "0");
Model Episodes = Treatment  Year Male/ s dist=Poisson;
Random  INT / subject= Village_code;
```

- **IL11**

```
* Model 3 - Kenward Roger;
proc glimmix data=WORK.import pconv= 1e-5 ;
class Village_code Treatment(ref = "0") YEAR Male(ref = "0");
Model Episodes = Treatment  Year Male/ s dist=Poisson ddfm=kr;
Random  INT / subject= Village_code;
```

- **IL12**

```
* Model 4 - Quad;
proc glimmix data=WORK.import method=quad(qpoints = 10) pconv=
1e-5 ;
class Village_code Treatment(ref = "0") YEAR Male(ref = "0");
Model Episodes = Treatment  Year Male/ s dist=Poisson;
Random  INT / subject= Village_code;
```

- **IL13**

```
* Model 5 - Model SE
```

```
proc glimmix data=WORK.import pconv= 1e-5;
class Village_code Treatment(ref = "0") YEAR Male(ref = "0");
Model Episodes = Treatment  Year Male/ s dist=Poisson;
Random  residual / subject= Village_code type=CS;
```

- **IL14**

```
* Model 6 - Empircal SE;
proc glimmix data=WORK.import empirical pconv= 1e-5;
class Village_code Treatment(ref = "0") YEAR Male(ref = "0");
Model Episodes = Treatment  Year Male/ s dist=Poisson;
Random  residual / subject= Village_code type=CS;
```

- **IL15**

```
* Model 7 - MBN;
proc glimmix data=WORK.import empirical = mbn pconv= 1e-5;
class Village_code Treatment(ref = "0") YEAR Male(ref = "0");
Model Episodes = Treatment  Year Male/ s dist=Poisson;
Random  residual / subject= Village_code type=CS;
```

- **IL16**

```
* Model 8 - FG;
proc glimmix data=WORK.import empirical = firoeeq pconv= 1e-5;
class Village_code Treatment(ref = "0") YEAR Male(ref = "0");
Model Episodes = Treatment  Year Male/ s dist=Poisson;
Random  residual / subject= Village_code type=CS;
```

Village as Random Effect, Cluster Period as Random Effect

$log(\lambda_{ijk}) = \alpha + \beta_1 X_i + \beta_2 X_{ijk} + u_{ij} + v_{hij}$, where $X_i = 0$ (control) or $X_i = 1$ (Ivermectin), $X_{ijk} = 0$ (female) or $X_{ijk} = 1$ (male) and:

$$
\begin{aligned}
\alpha &= \text{Intercept} \\
\beta_1 &= \text{Log of the rate ratio treatment vs control} \\
\beta_2 &= \text{Log of the rate ratio males vs females} \\
u_{ij} &= \text{Denotes the random village effect} \\
v_{hij} &= \text{Denotes the random village-period effect}
\end{aligned}
$$

- **IL17, IL18 (Maximum Likelihood mixed-effects models only)**

Same as proc statements as above, but with the following change:

```
class Village_code Treatment(ref = "0") YEAR;
Model Episodes = Treatment Male/ s dist=Poisson;
Random  INT / subject= village_code;
Random  INT / subject= YEAR(village_code);
```

2 Year Parallel Cluster Level Models (R)

$$\begin{aligned} \overline{R}_{ij} &= \frac{\sum Episodes_k}{m_j} \\ Episodes_k &= \text{Number of episodes for subject k} \\ m_j &= \text{Number of subjects in cluster j} \end{aligned}$$

$log(\overline{R}_{ij}) = \alpha + \beta_1 X_{ij}$, where $X_i = 0$ (control) or $X_i = 1$ (Ivermectin) and:

$$\begin{aligned} \alpha &= \text{Intercept} \\ \beta_1 &= \text{Log of the rate ratio treatment vs control} \end{aligned}$$

- **CL5**

Unweighted t-test – Summarize over both years to get 14 means

```
glm(Mean ~ TRT, data = DF)
```

- **CL6**

Adjusted Residual

Note: Same model as above, but year is a covariate and the GLM is fit on individual data without the treatment effect.

Fit GLM without treatment effect:

```
glm(Episodes ~ Male + Year, data = ., family = poisson)
```

Extract residuals and summarize cluster level means, fit this model:

```
glm(mean.resids ~ TRT
```

- **CL7**

Treatment + Time – Summarize cluster periods to get 28 means

Note: Same as unweighted t-test, but each cluster period has its own mean

```
glm(Mean ~ TRT + Year, data = DF, family=gaussian(link="log"))
```

- **CL8**

Weighted Treatment + Time – Summarize cluster periods to get 28 means

Note: Weights calculated by inverse weighting size as described previously [19]

```
glm(Mean ~ TRT + Year, data = DF, weights = w,
family=gaussian(link="log")), data = DF)
```

2 Year Cross-over – Individual Level Models (SAS)

Same as above, but with the following change: where "Treatment" refers to a variable indicating if ivermectin was received or not for that particular year.

Village as Fixed Effect, Cluster Period as Random Effect

- **IL19 – 26**

$log(\lambda_{ijk}) = \alpha + \beta_1 X_{hi} + \beta_2 X_{ijk} + \beta_3 X_j + v_{hij}$, where $X_{hi} = 0$ (control) or $X_i = 1$ (Ivermectin), $X_{ijk} = 0$ (female) or $X_{ijk} = 1$ (male) and $X_j = 1 - 14$ (Village 1 - 14)and:

$$\begin{aligned} \alpha &= \text{Intercept} \\ \beta_1 &= \text{Log of the rate ratio treatment vs control} \\ \beta_2 &= \text{Log of the rate ratio males vs females} \\ \beta_3 &= \text{Log of the rate ratio for each cluster versus cluster reference} \\ v_{hij} &= \text{Denotes the random village-period effect} \end{aligned}$$

```
class Village_code Treatment(ref = "0") YEAR Male(ref = "0");
Model Episodes = Treatment  Village_code Male/ s dist=Poisson;
Random  INT / subject= YEAR(village_code);
```

Village as Random Effect, Cluster Period as Random Effect

- **IL27 and 28**

$log(\lambda_{ijk}) = \alpha + \beta_1 X_{hi} + \beta_2 X_{ijk} + u_{ij} + v_{hij}$, where $X_{hi} = 0$ (control) or $X_i = 1$ (Ivermectin), $X_{ijk} = 0$ (female) or $X_{ijk} = 1$ (male) and:

$$\begin{aligned} \alpha &= \text{Intercept} \\ \beta_1 &= \text{Log of the rate ratio treatment vs control} \\ \beta_2 &= \text{Log of the rate ratio males vs females} \\ u_{ij} &= \text{Denotes the random village effect} \\ v_{hij} &= \text{Denotes the random village-period effect} \end{aligned}$$

```
class Village_code Treatment(ref = "0") YEAR;
Model Episodes = Treatment Male/ s dist=Poisson;
Random  INT / subject= village_code;
Random  INT / subject= YEAR(village_code);
```

2 Year Cross-over – Cluster Level Models (R)

- **CL9**

Treatment + Time + Cluster – Summarize cluster periods to get 28 means

```
glm(Mean ~ TRT + Year + as.factor(Village_code), data = DF,
family=gaussian(link="log"))
```

Note: same as previous models but a fixed effect for village is applied.

Simulate Data

The R code below shows how to simulate multiple years of data while maintaining the same within village random effect across years. The function was used create a 14-cluster data set but allow for different village random effect variances; here, a village random effect variance of 0.10 was used. See the simstudy R package:

> Keith Goldfeld. simstudy: Simulation of Study Data. 2019. https://CRAN.R-project.org/package=simstudy (28 July 2020, date last accessed)

```
library(simstudy)

set.seed(2007)

# Control Rate = 1.088
# Treatment Rate = 0.619
# Rate Ratio
RR <- 0.619/1.088
RR # 0.5689338

## [1] 0.5689338

# Determine the effect for the poisson model
# Control: 1.088 = exp(x)
log(1.088) # 0.08434115

## [1] 0.08434115

# Treatment:  0.619 = exp(0.08434115 + X)
log(0.619) - 0.08434115

## [1] -0.5639912

# -0.5639912 #

# Check
exp(0.08434115 + -0.5639912) #0.619

## [1] 0.619

exp(0.08434115) # 1.088

## [1] 1.088

glmerFUN <- function(VillageRE) {
  Form = "0.0843 + -0.5639912*TRT  + 0.13*Male + VillageRE"
  Form2 = "0.0843 + -0.5639912*CRXO_TRT_Year2 + 0.13*Male + VillageRE"
  Form3 = "0.0843 + -0.5639912*TRT +  -0.1053605 + 0.13*Male  + VillageRE"
  Form4 = "0.0843 + -0.5639912*CRXO_TRT_Year2 + -0.1053605 + 0.13*Male + Vill
ageRE"
```

```
  Form5 = "0.0843 + -0*TRT  + 0.13*Male + VillageRE"
  Form6 = "0.0843 + -0*TRT + -0.1053605 + 0.13*Male +VillageRE"
  Form7 = "0.0843 + -0*CRXO_TRT_Year2 + -0.1053605 + 0.13*Male + VillageRE"

  gen.village <- defData(varname = "VillageRE", dist = "normal", formula = 0,
variance = VillageRE, id = "Village_code")
  gen.village <- defData(gen.village, varname = "nSubjects", formula = "rep(c
(70,88,127,112,98,81,123,98,118,37,95,83,109,78), each = 1)")
  gen.village <- defData(gen.village, varname = "TRT", formula = "rep(c(1,0),
each = 7)")
  gen.village <- defData(gen.village, varname = "CRXO_TRT_Year2", formula = "
rep(c(0,1), each = 7)")
  dtVillage <- genData(14, gen.village)

  gen.Subject <- defDataAdd(varname = "Male", dist = "binary", formula = 0.5)
  gen.Subject <- defDataAdd(gen.Subject, varname = "Episodes_Year1", dist = "
poisson",  formula = Form , link = "log")
  gen.Subject <- defDataAdd(gen.Subject, varname = "Episodes_Year2_10", dist
= "poisson",  formula = Form3 , link = "log")
  gen.Subject <- defDataAdd(gen.Subject, varname = "Episodes_Year2_CRXO_10",
dist = "poisson",  formula = Form4 , link = "log")
  gen.Subject <- defDataAdd(gen.Subject, varname = "Episodes_Year1_Null", dis
t = "poisson",  formula = Form5 , link = "log")
  gen.Subject <- defDataAdd(gen.Subject, varname = "Episodes_Year2_10_Null",
dist = "poisson",  formula = Form6 , link = "log")
  gen.Subject <- defDataAdd(gen.Subject, varname = "Episodes_Year2_CRXO_10_Nu
ll", dist = "poisson",  formula = Form7 , link = "log")


  dtSubject <- genCluster(dtVillage, cLevelVar = "Village_code", numIndsVar =
"nSubjects",  level1ID = "Subject")

  dtSubject <- addColumns(gen.Subject, dtSubject)

}

SIM2 <- do.call("rbind.data.frame", replicate(1000, glmerFUN(VillageRE = 0.10
), simplify = FALSE))

rep <- 1000
SIM2$i <- NA
SIM2$i[SIM2$Subject == 1] <- rep(c(1:rep), each = 1)
SIM2Final <- SIM2 %>% fill(i)
```

References

1. Wolak ME, Fairbairn DJ and Paulsen YR. Guidelines for estimating repeatability. *Methods in Ecology and Evolution* 2012; 3: 129-137.

Previously cited in main text:

16. Leyrat C, Morgan KE, Leurent B, et al. Cluster randomized trials with a small number of clusters: which analyses should be used? *International journal of epidemiology* 2018; 47: 321-331.

19. Turner RM, White IR, Croudace T, et al. Analysis of cluster randomized cross-over trial data: a comparison of methods. *Stat Med* 2007; 26: 274-289.