Contents

[Aims 1](#_Toc127472200)

[Simulation Structure 2](#_Toc127472201)

[Matematical expressions for the Simulation 3](#_Toc127472202)

[Backgrounds 3](#_Toc127472203)

[Bärenbold paper (Bärenbold et al., 2017) 3](#_Toc127472204)

[Levecke Chapter (Levecke et al., 2015) 4](#_Toc127472205)

[Exploring Package prevalence (Devleesschauwer et al., 2022) 6](#_Toc127472206)

[Dissecting prevalence::truePrev() 6](#_Toc127472207)

[Dissecting prevalence::truePrevPools() 7](#_Toc127472208)

[For the individual samples: 8](#_Toc127472209)

[Nex step: how to add Age and infection intensity? 8](#_Toc127472210)

[We are try to use the following two expressions: 8](#_Toc127472211)

[References 9](#_Toc127472212)

# Aims

1. To estimate the STH true prevalence based on indiviudal stool samples, assuming a perfect specificity, being informed by a binary result of the Kato-Katz test, number of eggs (proxy for infection intensity) and age for each subject (optional).
   1. Variability in the population could be added explored (optional)
2. To estimate the STH true prevalence based on 10-pooled stool samples, assuming a perfect specificity, being informed by a binary result of the Kato-Katz test, infection intensity and age-average for each pool sample (optional).
   1. Variability in the population could be added explored (optional)

# Simulation Structure

odds=prev.STH/(1-prev.STH)

odd.p.s<-exp(log(odds)-log(ratio))

odd.p.e<-exp(log(odds)+log(ratio))

runif(schools, min = odd.p.s/(1+odd.p.s), max = odd.p.e/(1+odd.p.e))

However, prevalence per age can be included in the simulation DGM.

Assuming that we have a bimodal normal distribution. We only need to provide and the general prevalence

Chart

Description automatically generated

Not sure how to integrate this distribution with infection intensity. A case or a control subject should have a infection intensity related to the probability (continuous value from 0 to 1) to been a case, also with the posibility to have cyles were the STH are sexuality active or latent.

Chart, scatter chart

Description automatically generated

## Matematical expressions for the Simulation

Age for subject i follow a mixture distribution of two normal

----

before we used an unique

## Backgrounds

### Bärenbold paper (Bärenbold et al., 2017)

Sensitivity in Bärenbold model depends on the number of repeated measures.   
However here should be depending on the type of pool sample

r could a parameter for distreibution of infection acrros population. TBC.

### Levecke Chapter (Levecke et al., 2015)

#### Faecal Egg Counts (FECs) on individual stool samples

Expressed in EPG: faecal egg counts (FECs) per gram of stool

FECs presents a skewed distribution, usually modelled by negative binomial distribution or ZIP.

Hierarchical model for FECs

*, or*

*Observed FECs*

true aggregation of in a population prior to an intervention

is the amount of stool examined (in g)

is the true FEC expressed in EPG for an individual

egg counts under the microscope

as the true underlying number of eggs in grams of stool

#### Faecal Egg Counts (FECs) on multiple stool samples

Hierarchical model for FECs

, or

Observed FECs

is the amount of stool examined (in g)

is the true FEC expressed in EPG for an individual

egg counts under the microscope

as the true underlying number of eggs in grams of stool

true FEC (in EPG) in a pooled stool sample from a population prior to an intervention

# Exploring Package prevalence (Devleesschauwer et al., 2022)

### Dissecting prevalence::truePrev()

truePrev(x, n, SE = 1, SP = 1, prior = c(1, 1), nchains = 2, burnin = 10000, update = 10000, verbose = FALSE)

(no informative, large variance)

(shape rate parametrization, NI prior)

(no informative, large variance)

AP: Aparent prevalence

TP: true prevalence

x: The apparent number of positive samples

n: The sample size

SE, SP: The prior distribution for sensitivity (SE) and specificity SP); see 'Details' below for specification of these distributions

prior: The parameters of the prior Beta distribution for true prevalence; defaults to c(1, 1)

### Dissecting prevalence::truePrevPools()

truePrevPools(x, n, SE = 1, SP = 1, prior = c(1, 1), nchains = 2, burnin = 10000, update = 10000, verbose = FALSE)

for (i in 1:N) {

x[i] ~ dbern(AP[i])

AP[i] <- SEpool[i] \* (1 - pow(1 - TP, n[i])) + (1 - SPpool[i]) \* pow(1 - TP, n[i])")

SEpool[i] <- 1 - (pow(1 - SE, n[i] \* TP) \* pow(SP, n[i] \* (1 - TP)))")

SPpool[i] <- pow(SP, n[i])

}

Priors for Se, Sp, TP

Priors:

x: The vector of indicator variables, indicating whether a pool was positive ("1") or negative ("0")

n: The vector of pool sizes

SE, SP: The prior distribution for sensitivity (SE) and specificity (SP); see 'Details' below for specification of these distributions

prior: The parameters of the prior Beta distribution for true prevalence; defaults to c(1, 1)

## For the individual samples:

### Nex step: how to add Age and infection intensity?

### We are try to use the following two expressions:

#### Bärenbold:

Maybe this expression can be used, used in paper (Bärenbold et al., 2017) for repeated measures, so

, which is interesting since sensitivity could be updated by infection intensity!!

Where

#### Also (Levecke et al., 2015) (Chapt Book):

*, or*

*Observed FECs*

#### Two possible integrations

##### Proposition a

[[1]](#footnote-1)

*, or*

*Observed FECs*

where r is measurement “perfection” .   
***For me is not clear the definition of r based on the paper.***

##### Proposition b

An alternative could be to add infection intensity in to the Age regression

**Observed:** Region, School, Eggs count

**To be check:** In the datasetDo we have Age?, definiton?:

**To be estimated:** (Prevalence), Sensitivity

**Level Age**

**Level 2 mean egg intensity between individuals, “Variability in mean egg intensity between individuals within a population due to variation in infection levels between individual”**

**Level 3, “Day-to-day variability in mean egg intensity within an individual due to, for example, heterogeneous egg excretion ~~over time~~” related to individual/pooled sampling.**

**Level 4 agrregated distribution of eggs in feces**

**“Variability in mean egg intensity between multiple KK based on the same nonhomogenized stool sample due to the aggregated distribution of eggs in feces”**

**Level 5 measurement, “Variability in count observations due to random diagnostic variation”**

Chart

Description automatically generated

Level Sensitivity depending on the Infection intensity, not sure at all were the expression is comming

where r is measurement “perfection” . r “perfection” variability between slides, if r is infinite NB tends to be a Poisson. We need to check.

Posterior distribution of conditional to

Level Prevalence

# References

Bärenbold, O., Raso, G., Coulibaly, J. T., N’Goran, E. K., Utzinger, J., & Vounatsou, P. (2017). Estimating sensitivity of the Kato-Katz technique for the diagnosis of Schistosoma mansoni and hookworm in relation to infection intensity. *PLOS Neglected Tropical Diseases*, *11*(10), e0005953. https://doi.org/10.1371/journal.pntd.0005953

Devleesschauwer, B., Torgerson, P., Charlier, J., Levecke, B., Praet, N., Roelandt, S., Smit, S., Dorny, P., Berkvens, D., & Speybroeck, N. (2022). *prevalence: Tools for Prevalence Assessment Studies* (0.4.1). https://CRAN.R-project.org/package=prevalence

Levecke, B., Anderson, R. M., Berkvens, D., Charlier, J., Devleesschauwer, B., Speybroeck, N., Vercruysse, J., & Van Aelst, S. (2015). Mathematical Inference on Helminth Egg Counts in Stool and Its Applications in Mass Drug Administration Programmes to Control Soil-Transmitted Helminthiasis in Public Health. In *Advances in Parasitology* (Vol. 87, pp. 193–247). Elsevier. https://doi.org/10.1016/bs.apar.2015.01.001

1. For simplification purposes, here I am ignoring the level for population so j is suppressed. [↑](#footnote-ref-1)