**supplement material**

**The Multilevel Bayesian ZINB Model**

The ZINB model is a mixture of a zero-distribution, from which only zero values are observed, and a negative binomial (NB) distribution from which all of the nonzero and a few of the zero values are observed. Therefore, there are two sources of zero counts: a few coming from the first distribution are *structural zeros*, others coming from the NB are *sampling zeros*. Some children may never experience caries (structural zeros), while others may have recorded zero dmft in the sample time frame, and a non-zero in some future time frame (sampling zeros). The overall probability of zero counts is the combined probability of zeros from two groups; however, which group they come from is not known. Under zero-inflated distributions, dmft counts , *i*=1,2,…,*n* can be expressed by ~0 with probability  and ~NB with probability , with  the mean and *r* being the over-dispersion parameter. For independent and identically distributed responses, the ZINB model can be written as:

Prob (1)

Prob 

where the mixture proportion  denotes the probability of a structural zero. Model (1) can be extended to allow for covariates and hierarchically structured responses. With two hierarchical levels, the response is , with  and . The mixture proportion  and the mean  may depend upon covariates through appropriate link functions, which take the form of the logit and the log, respectively. Because the covariates that influence  and  are not necessarily the same, we can model the inflation and the NB components as:

logit (2.1)

log (2.2)

where  and **β** are the unknown fixed effect coefficients associated with the given covariate vectors ’s and  and  are random effect parameters associated to ’s covariates affecting heterogeneity among subjects.

Under a full Bayesian approach, prior distributions are assumed for all parameters. The overdispersion parameter *r* is assumed to follow a gamma(10-2,10-2). The fixed effects  and **β** are assumed to follow vague independent Normal distributions with zero mean and precision=10-3. For the random effects  and , the multivariate normal prior distribution with unknown mean **γ** and covariance matrix Σ is assumed. A non-informative multivariate normal prior is specified for the population mean **γ**, whilst the inverse covariance matrix **T** = Σ-1 is assumed to follow a Wishart distribution. To represent vague prior knowledge, the degrees of freedom for this distribution are chosen to be the rank of **T**. The scale matrix **R** for the Wishart distribution is specified to represent our prior guess at the order of magnitude of Σ, with 5×10-3 for the off-diagonal coefficients and 10-2 for the diagonal elements.

The simplest two-levels model, which includes only the random intercepts, is called variance components model. The random effects  and  are assumed to follow independent Normal distributions with known hyperparameters (the parameters of the prior distributions). Vague normal and gamma distributions are chosen as priors, respectively, for the hyper-mean and the hyper-variance.

WINBUGS CODES

**ZINB Two-level Variance Components Model (Table 5)**

model {

*######* *likelihood######*

for (i in 1:n) {

dmft[i]~dnegbin(mu[i],r)

mu[i] <- r/( r+lambda[i]\*(1-u[i]) )

u[i]~dbern(p[i])

logit(p[i])<-alpha[city[i]] +alphafix[1]\*gender[i]+alphafix[2]\*mother\_education[i]+alphafix[3]\*father\_education[i]+alphafix[4]\*sweetbeverage[i]+alphafix[5]\*fluoro-intake[i]+alphafix[6]\*tap[i]+alphafix[7]\*mineral[i]+alphafix[8]\*often[i]+alphafix[9]\*DI[i]+alphafix[10]\*ln\_POP[i]

log(lambda [i])<- beta[city[i]] +betafix[1]\*gender[i]+betafix[2]\*mother\_education[i]+betafix[3]\*father\_education[i]+betafix[4]\*sweetbeverage[i]+betafix[5]\*fluoro-intake[i]+betafix[6]\*tap[i]+betafix[7]\*mineral[i]+betafix[8]\*often[i]+betafix[9]\*DI[i]+betafix[10]\*ln\_POP[i]

zdp[i]<-p[i]+ (1-p[i])\*pow((r/(r+lambda[i])),r)

}

mzdp<-mean(zdp[])

*######* Priors for fixed effects*######*

r~ dgamma(0.01,0.01)

for (k in 1 : 10) {

alphafix[k] ~ dnorm(0.0, 0.001)

betafix[k] ~ dnorm(0.0, 0.001)

}

for (j in 1 : M) {

alpha[j] ~ dnorm(alpha.c,alpha.tau)

beta[j] ~ dnorm(beta.c,beta.tau)

}

*######*Hyper-priors*######*

alpha.c ~ dnorm(0.0,0.001)

alpha.tau ~ dgamma(0.001,0.001)

beta.c ~ dnorm(0.0,0.001)

beta.tau ~ dgamma(0.001,0.001)

*######* End of the model *######*

}

**ZINB random slopes Model (Figure 1)**

model

{

*######* *likelihood######*

for (i in 1:n) {

dmft[i]~dnegbin(mu[i],r)

mu[i] <- r/( r+lambda[i]\*(1-u[i]) )

u[i]~dbern(p[i])

logit(p[i])<-alpha[city[i],1]+alpha[city[i],2]\*DI[i]+alpha[city[i],3]\*ln\_POP[i] +alphafix[1]\*gender[i]+alphafix[2]\*mother\_education[i]+alphafix[3]\*father\_education[i]+alphafix[4]\*sweetbeverage[i]+alphafix[5]\*fluoro-intake[i]+alphafix[6]\*tap[i]+alphafix[7]\*mineral[i]+alphafix[8]\*often[i]

log(lambda [i])<- beta[city[i],1]+beta[city[i],2]\*DI[i]+beta[city[i],3]\*ln\_POP[i] +betafix[1]\*gender[i]+betafix[2]\*mother\_education[i]+betafix[3]\*father\_education[i]+betafix[4]\*sweetbeverage[i]+betafix[5]\*fluoro-intake[i]+betafix[6]\*tap[i]+betafix[7]\*mineral[i]+betafix[8]\*often[i]

zdp[i]<-p[i]+ (1-p[i])\*pow((r/(r+lambda[i])),r)

}

mzdp<-mean(zdp[])

*###### Covariance matrix for the inflation and the NB components ######*

cova[1:3,1:3]<-inverse(T[,])

cova1[1:3,1:3]<-inverse(T1[,])

*######* Priors for fixed effects*######*

r~ dgamma(0.01,0.01)

for (k in 1 : 8) { alphafix[k] ~ dnorm(0.0, 0.001)

betafix[k] ~ dnorm(0.0, 0.001)

}

*######* Priors for random effects*######*

for (j in 1 : M) {

alpha[j, 1:3 ] ~ dmnorm(gamma[1:3 ], T[1:3 ,1:3 ])

beta[j, 1:3 ] ~ dmnorm(gamma[1:3 ], T1[1:3 ,1:3 ])

}

*######*Hyper-priors*######*

gamma[1:3] ~ dmnorm(mn[1:3 ], prec[1:3 ,1:3 ])

T[1:3 ,1:3 ] ~ dwish(R[1:3 ,1:3 ], 3)

T1[1:3 ,1:3 ] ~ dwish(R1[1:3 ,1:3 ], 3)

*######* End of the model *######*

}