**Missing covariate data model**

Introduce a missing data model by specifying a prior distribution with common unknown parameters for both observed and unobserved values. The observed values will contribute to the estimation of the unknown parameters which in turn will inform the missing data.

We propose a negative binomial distribution for the data generating process with a logit link. This can be thought of as a generalisation of the Poisson model which can account for over or under-dispersion.

For number of lakes, number of bacteria, number of successes until stopping, probability of success, is the lake specific adjustment, are the set of regression parameters for each bacteria. *WT,* *MAT, logarea* and *logWpH* are water temperature, mean air temperature, log lake area and log water pH, respectively. The outcome data are the number of bacteria.

The missing data model, assuming ignorable missing covariate mechanism

Prior distributions are

When there are other fully observed covariates in the regression model this will not account for correlation between these. Further, we can specify a regression model to impute the missing data as a function of the other covariates.