

Technical FAQs

Last Modified on 17/04/2024 3:36 pm BST

How does MBG work in the context of trachoma elimination?

MBG uses regression models to predict the prevalence of trachoma (TF or TT). A regression model describes the relationship between a target (in this case, TF or TT prevalence), and one or more independent variables. These variables could include things such as age, gender, proximity to healthcare infrastructure, population density, temperature, and many other factors relevant to trachoma epidemiology. MBG also allows for the inclusion of spatial effects, which takes into account the observation that trachoma prevalence is often spatially correlated (i.e., trachoma prevalence is more similar when the geographical locations are closer). When estimating trachoma prevalence, MBG can also use existing survey data from nearby EUs to improve the estimate precision for the region of interest. The MBG output includes a point prevalence for the region of interest, and a Probability of being Below the elimination Threshold (PBT).

Who chooses the covariates to be included in the model?

Current research conducted by RTI International and Lancaster University is focused on investigating the most useful covariates that consistently demonstrate a significant relationship with trachoma prevalence. From a starting point of over 60 covariates (environmental, social, demographic), researchers are aiming to reduce this to a standard pool of ~20, which will be included in all starting models when using MBG. The models themselves will then be used to determine which of these 20 covariates best explain the variation in the data for each region of interest, and these covariates will then be taken forward into the final models used for estimating trachoma prevalence.

Where are the covariate data sourced from?

The covariate data are sourced from a number of places, and these may change and be updated over time. Current sources include WorldPop, Malaria Atlas Project, DHS, CHIRPS, etc.

Are older data given less importance in the models relative to more recent data?

Current models do not weight the data based on when the data were collected, so data collected many years ago will be given equal importance to data collected within the last year. In an ideal scenario, it would be possible to explicitly acknowledge time in the model, however this would increase the complexity of the models to an extent that it might increase the uncertainty of the output, potentially negating the benefits of adding in time as an additional variable.

Is it possible to "borrow" data from a long distance away, or does MBG work better when districts are contiguous?

The extent to which we're able to borrow information from other EUs depends on the data and the strength of spatial correlation, and is decided on a case-by-case basis. For example, in some cases correlations between trachoma prevalence spans a long geographic range, indicating that the same factors that influence trachoma prevalence in one EU are the same factors that influence the trachoma prevalence in an EU many miles away. In this case, it would be possible to use data from EUs from a wide geographic range. However, if the correlation between trachoma prevalence was shown to be much more localised, it would be more beneficial to use data from contiguous EUs rather than ones further away.

Does MBG use data that come from different survey types?

We do not combine data that come from areas in different stages of intervention - so for example, we would never combine baseline data (pre-MDA) with any other survey type (e.g. impact or surveillance survey). We do however on occasion combine data from impact and surveillance surveys, as long as they are in the same stage of intervention (post-MDA), and only when the data available are sparse. Ideally, all data would come from one survey type, and we would use the most recent data available.

Can data over 5 years old be used?

Technically, as long as data were generated using either GTMP or Tropical Data methods, they can be used regardless of when they were collected. Currently the most recent data available for an area are preferentially selected for use in MBG, as these are likely to be the most informative of the current situation. The use of older data can be made as long as the area covered is the same as the newer data, as this allows variation over time to be captured through the MBG model. In cases where the older data cover a different area from the newer data, the decision on the use of the old data depends on several factors such as differences in the intervention history, environment and socio-demographic traits of the populations from the two areas. In cases where such differences are deemed to be small enough, the older data can be used in the analysis. If not, the older data should not be combined with the newer data.

Has ground-truthing been done already or is there a plan to do it to validate the outcome of MBG?

Ground-truthing is difficult using field data because surveys take a sample of the total EU population to provide a prevalence estimate. To know the true prevalence, we would need to examine every individual in the EU. An alternative to ground-truthing that is currently being investigated is using a simulation model to generate datasets with known true prevalence, and see if the model estimates the prevalence precisely.

Is it possible to run MBG when there is very little trachoma, i.e. minimal TT?

It could be possible to run a model with small amounts of "positive" data, however with fewer data there is a higher chance of more uncertainty in the model, which would be demonstrated by a PBT close to 50%.
