Dynamic Aggregation Results

Chris Hoover et al

3/24/2021

# Results

There were 93 shehia-years in which no individuals had positive egg counts and 16 shehia-years in which the maximum individual egg count was 1 and therefore variance estimates for could not be derived. This left 447 observations of and , shown in Figure 1A along with the marginal estimate of from the fitted GEE. We estimate a median 0.012 (Bootstrapped IQR: 0.011-0.013) increase in associated with an IQR increase in , implying a negative relationship between mean egg burden and aggregation. This relationship appears to be driven by dynamics in children, as there was weak evidence for a relationship between and among adults (Fig 1B). Egg burden in adults ( 1.36) was substantially lower than in children ( 4.65), however. Shehias receiving the MDA+snail control and MDA+behavior change interventions appeared to have a stronger burden-aggregation relationship (larger , Fig 1B). An additional GEE model including main effects of intervention arm and island showed that reductions in were significantly larger in shehias receiving the MDA+snail control and MDA+behavioral change interventions ( 0.01 and 0.016, respectively).

Examination of the Bayes factors comparing the posterior probabilities of each Case shows that the mechanistic Case 3 performs better than the canonical Case 1 (mean Bayes factor = 1.97, range 0-12.78), while Case 2 (mean Bayes factor = 0.89, range 0.07-5.98), and the hybrid Case 4 (mean Bayes factor = 0.97, range 0-6.9) generally performed worse than Case 1. Superior performance of Case 3 was further confirmed by examination of the mean squared errors between the generated and observed data, with Case 3 generated data resulting in the lowest MSE across all three summary statistics (Fig 2).

Case 1 and 3 estimates of and exhibited a similar pattern to and , namely increasing aggregation (indicated by decreasing ) with decreasing measures of burden (Fig 1 and Fig 3a and 3c). However, Case 2 and Case 4 estimates showed the opposite effect of decreasing aggregation with decreasing worm burden (Fig 3b and 3d), though it should be noted that these two data-generating cases performed worse both in terms of MSE of the summary statistics and the Bayes factor.

Figure 4 shows the mating probability produced from the posterior distributions of all four data-generating cases. These estimates are also compared to analytic predictions of the mating probability for both static aggregation (Fig 4 solid lines) and dynamic aggregation (, , Fig 4 dashed and dotted lines, respectively). These analytic predictions generally align well with Case 1 and Case 3 estimates, but do frequently underestimate the mating probability.

## Figure captions

**Figure 1. Relationship between community egg burden and dispersion.**: Panel A shows the scatterplot of all estimates of $\\kappa\_{st}$ and $\\mathcal{E}\_{st}$ where each point is sized according to its weight, derived as the inveerse of the standard error of $\\alpha\_{st}$. Points are also symbolized according to their assigned intervention group–MDA only ($\\filledtriangleup$), MDA+snail control ($\\filleddiamond$), or MDA+behavioral intervention ($\\filledmedsquare$)–and the island of the shehia. The solid black line in panel A represents the marginal estimate of $\\mathbb{E}(\\kappa\_{st}|\\mathcal{E}\_{st})$ from the fitted GEE. Panel B shows estimates of $\\Delta\\kappa\_{IQR}$, the change in aggregation for an interquartile range increase in community egg burden stratified by population, treatment group, and island. The black point and error bars correspond to the unstratified marginal estimate of $\\Delta\\kappa\_{IQR}$ as in panel A. Error bars correspond to the interquartile range of estimates derived from B=5000 bootstrapped samples.

**Figure 2: Comparison of observed summary statistics to the median summary statistics generated from 1,000 simulated datasets derived from the ABC posterior distributions for every shehia-year.** Colors and right-hand facet labels indicate the data generating Case and the 1:1 line implying perfect agreement between observed and generated data is shown. Posterior predictions from the mechanistic Case 3 data-generating mechanism (row 3, purple) were found to better reproduce the observed data, particularly the adjusted prevalence measure, though all data-generating mechanisms tended to overestimate mean community egg burden and standard error in low- burden settings.

**Figure 3:**  Summary of approximate Bayesian computation estimates of community worm burden (W\_st) and aggregation (κ\_st^W) for Case 1 (Together, A), Case 2 (Separate, B), Case 3 (Mechanistic, C), and Case 4 (Hybrid, D). Dashed lines depict the marginal estimates of the fitted GEE for each Case: E(κ\_st^(W\_D ) |W\_st^D). Axes are log transformed to aid visualization.

**Figure 4: Comparison of mating probabilities derived analytically and from ABC posterior distributions.** Points depict the median mating probability across the median community worm burden from data generated for every shehia-year and Case from ABC posterior distributions. Solid lines depict analytical estimates of the mating probability assuming constant aggregation parameter, κ, equal to the median across all shehia-year observations. Dotted lines depict analytical estimates of the mating probability assuming aggregation changes as a function of burden according to the fitted egg burden GEE, and dashed lines depict analytical estimates of the mating probability assuming aggregation changes as a function of worm burden according the fitted worm burden GEE for each case (shown in Figure 3). Note that Case 2 analytical mating probability estimates are derived according to the Case 2 (distributed separately) mating probability function from [CITE Anderson], while all others are derived for the canonical Case (distributed together) mating probability function.