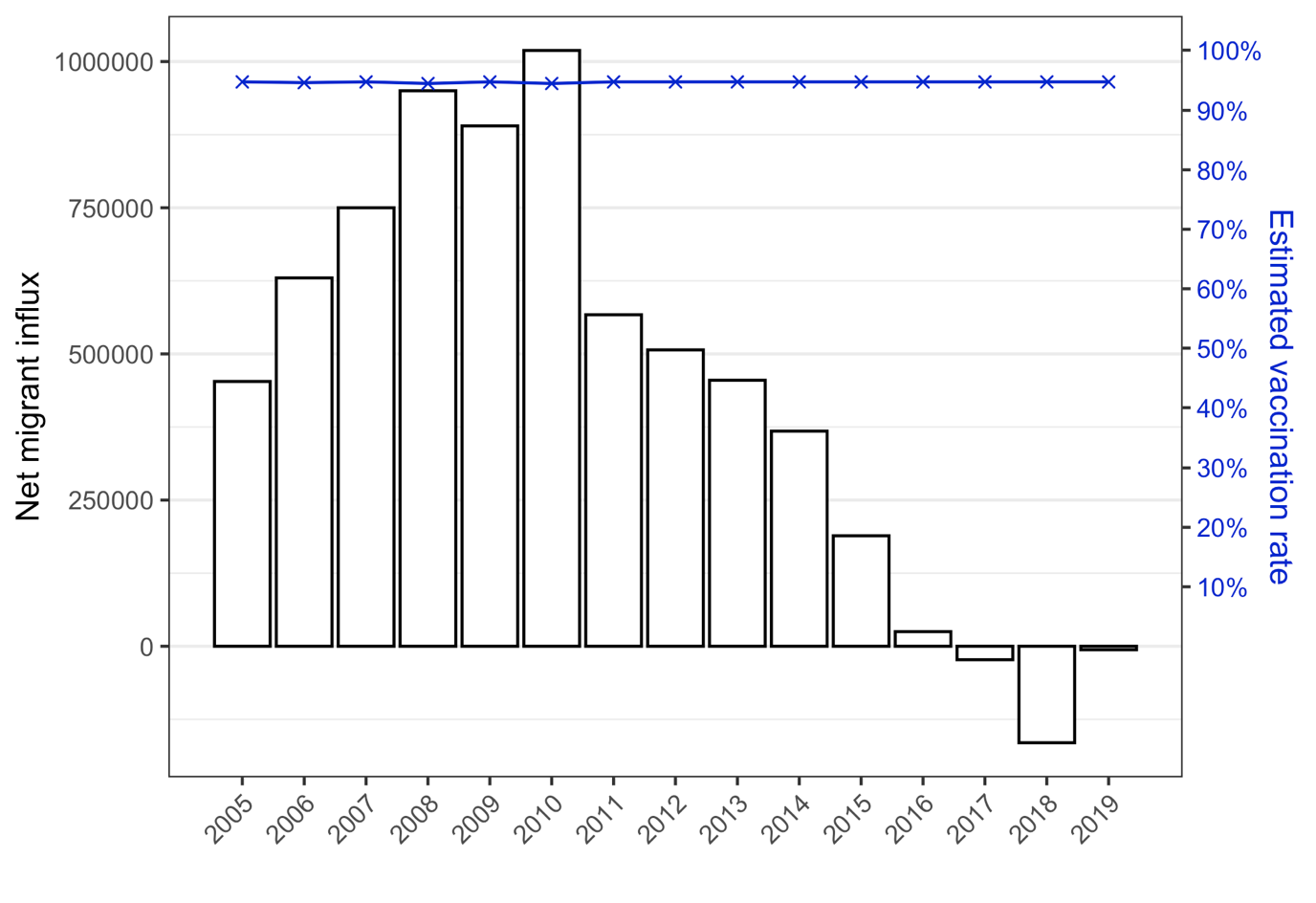
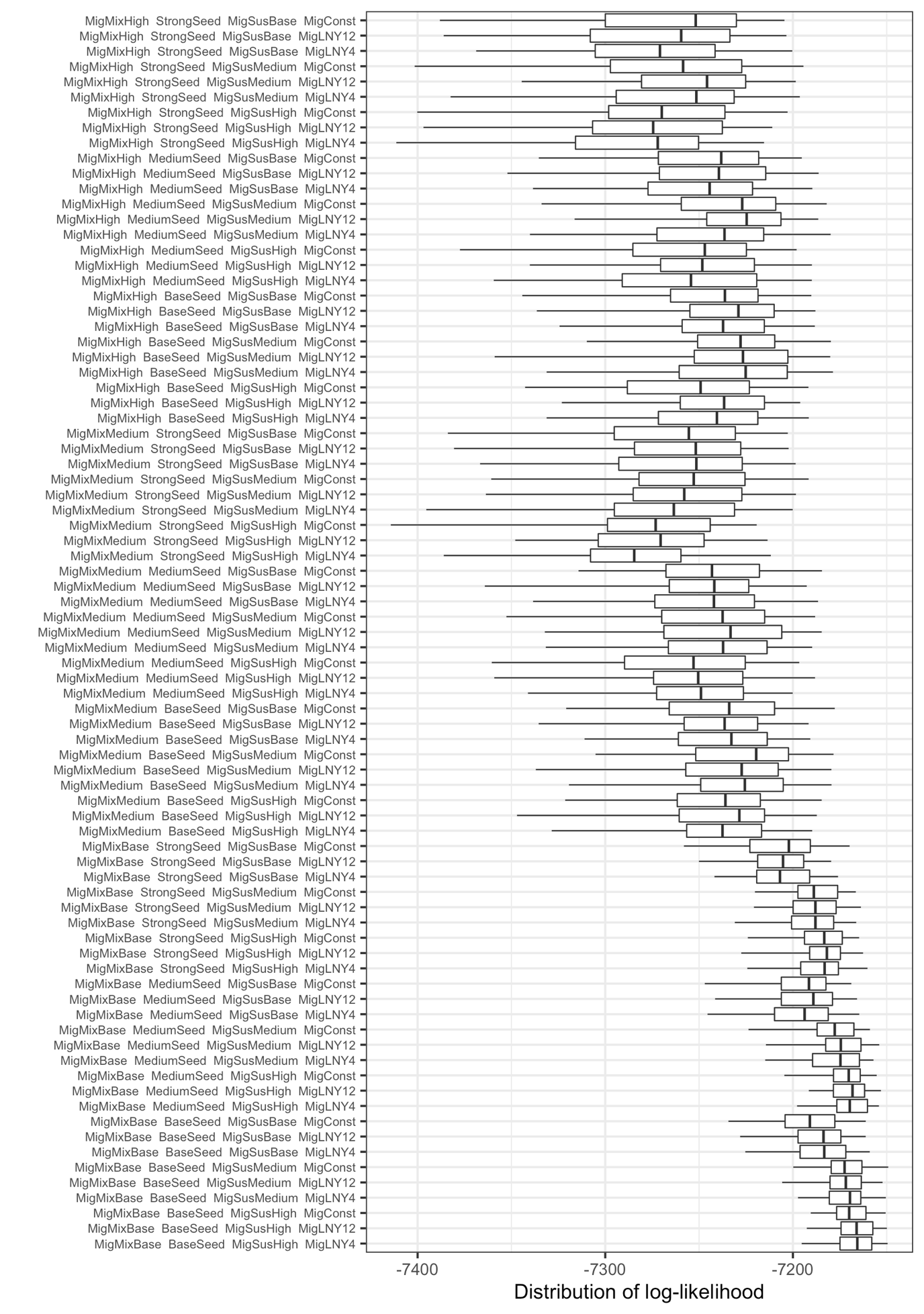


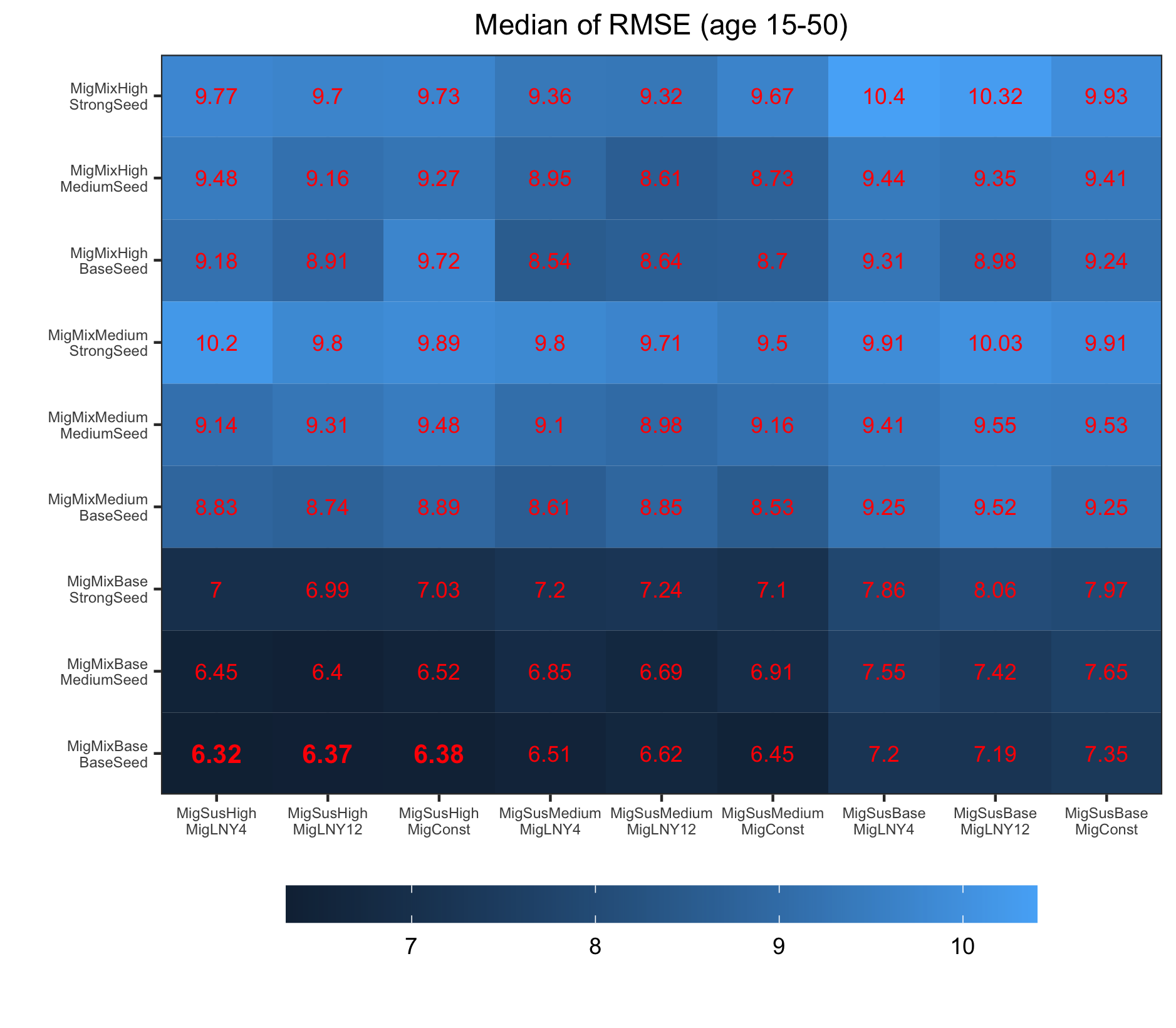
**Fig S1. Model schematic.** Modeled measles transmission follows the Susceptible-Exposed-Infectious-Removed (SEIR) dynamics. The model is age-structured, with age indicated by . For simplicity, only one age group is shown; birth, aging, and death are modeled but not shown here. Compartment M represents newborns with maternal immunity. is the force of infection; , the population size in age . is the rate of progression from the exposed to the infectious; , the recovery rate. is the rate of migration; is the fraction of migrants in compartment (). is the mean duration of maternal immunity. is an indicator function that equals 1 when and 0 otherwise. is the number of people vaccinated during catch-up or supplementary vaccination programs. represents routine vaccination (note both routine vaccination doses are administered around 1 year of age in Beijing, i.e., when individuals from group 1 are aging into group 2; and are the aging rate and routine vaccine coverage, respectively). represents seeding due to migration, and applies only if . Black arrows apply to all age groups; blue arrows apply to a subset of age groups.



**Fig S2. Trends of net migration and vaccination rate in Beijing during 2005-2019.** Due to a lack of detailed migration data, the models used net migration (histograms) as a proxy of total migration. This likely underestimated migrant influx, and more so for years after 2011 when net influx decreased to low or negative values. We estimated vaccination rate (blue line) by combining coverage of 2-dose measles vaccine and vaccine efficacy (90% and 95% for 1- and 2-doses, respectively; see details in Section 2.1 of Supplementary Materials).



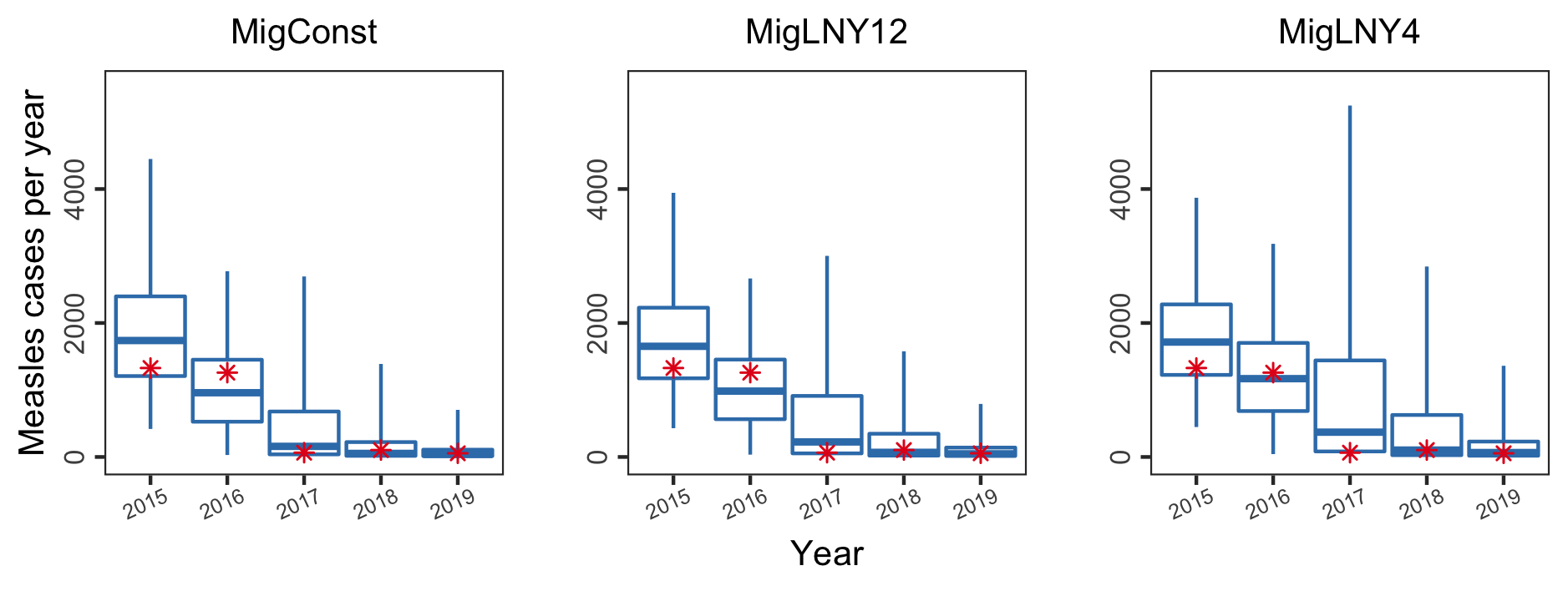
**Fig S3. Comparing migrant-related hypotheses based on distributions of log-likelihood.** We formulated four sets of migrant-related hypotheses (increased migrant mixing, seeding intensity, higher migrant susceptibility, and timing of migrant influx; corresponding to the 1st, 2nd, 3rd, and 4th terms in y-axis), and tested all combinations of the four sets (n=81, i.e., the number of boxplots). For each model (or hypothesis combination), we conducted 100 model inference runs using a model-filter system, and show the distribution of log-likelihood using boxplots (whiskers = 95% interval, box edges = Interquartile range, and thick middle bar = median). For infant-related hypotheses, all models tested here assumed Imm180 and InfantMixBase as baseline scenarios. See Table 1 for a summary of all hypotheses.



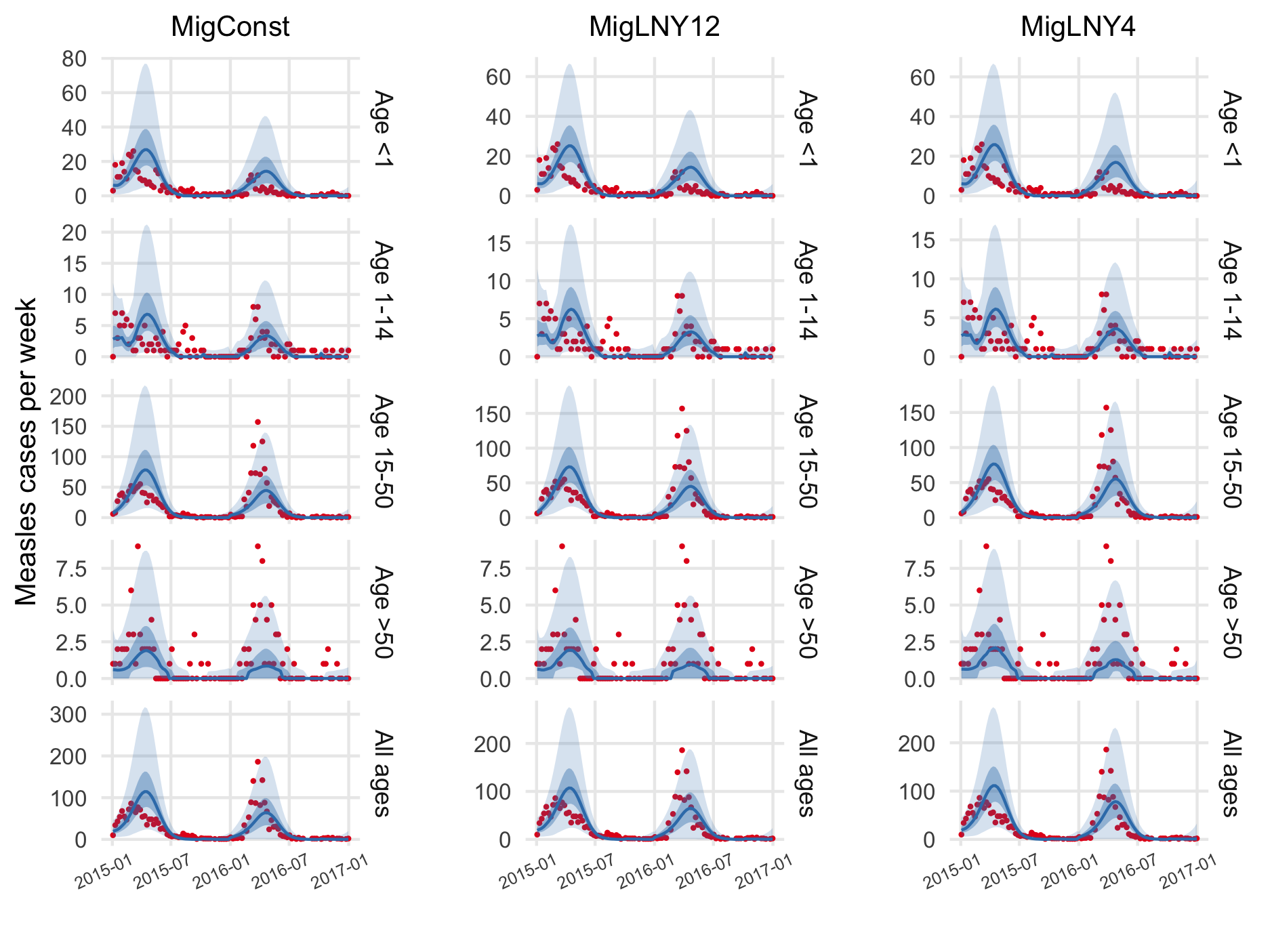
**Fig S4. Comparing migrant-related hypotheses based on medians of RMSE (age 15-50).** We formulated four sets of migrant-related hypotheses (increased migrant mixing [MigMixBase/MigMixMedium/MigMixHigh], seeding intensity [BaseSeed/MediumSeed/StrongSeed], higher migrant susceptibility [MigSusBase/MigSusMedium/MigSusHigh], and timing of migrant influx [MigConst/MigLNY12/MigLNY4]), and tested all combinations of the four sets (n=81, i.e., the number of cells in the Fig). For each model (or hypothesis combination), we conducted 100 model inference runs using a model-filter system, and show the median of the 100 RMSEs (age 15-50) in red. For infant-related hypotheses, all models tested here assumed Imm180 and InfantMixBase as baseline scenarios. See Table 1 for a summary of all hypotheses.



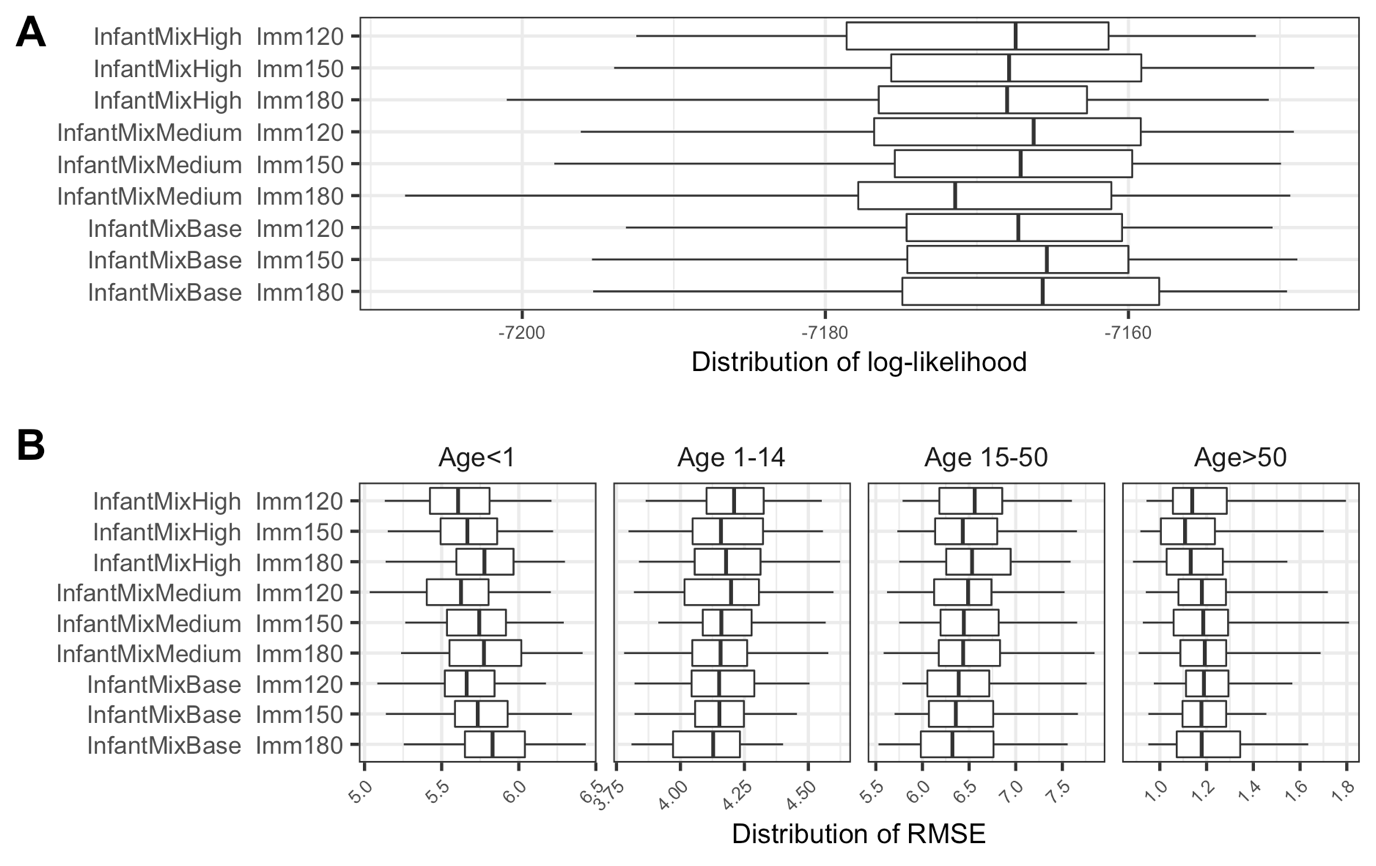
**Fig S5. Comparing migrant-related hypotheses based on distributions of RMSE.** We formulated four sets of migrant-related hypotheses (increased migrant mixing, seeding intensity, higher migrant susceptibility, and timing of migrant influx; corresponding to the 1st, 2nd, 3rd, and 4th terms in y-axis), and tested all combinations of the four sets (n=81, i.e., the number of rows of boxplots). For each model (or hypothesis combination), we conducted 100 model inference runs using a model-filter system, and show the distribution of RMSE using boxplots (whiskers = 95% interval, box edges = Interquartile range, and thick middle bar = median). For infant-related hypotheses, all models tested here assumed Imm180 and InfantMixBase as baseline scenarios. See Table 1 for a summary of all hypotheses.



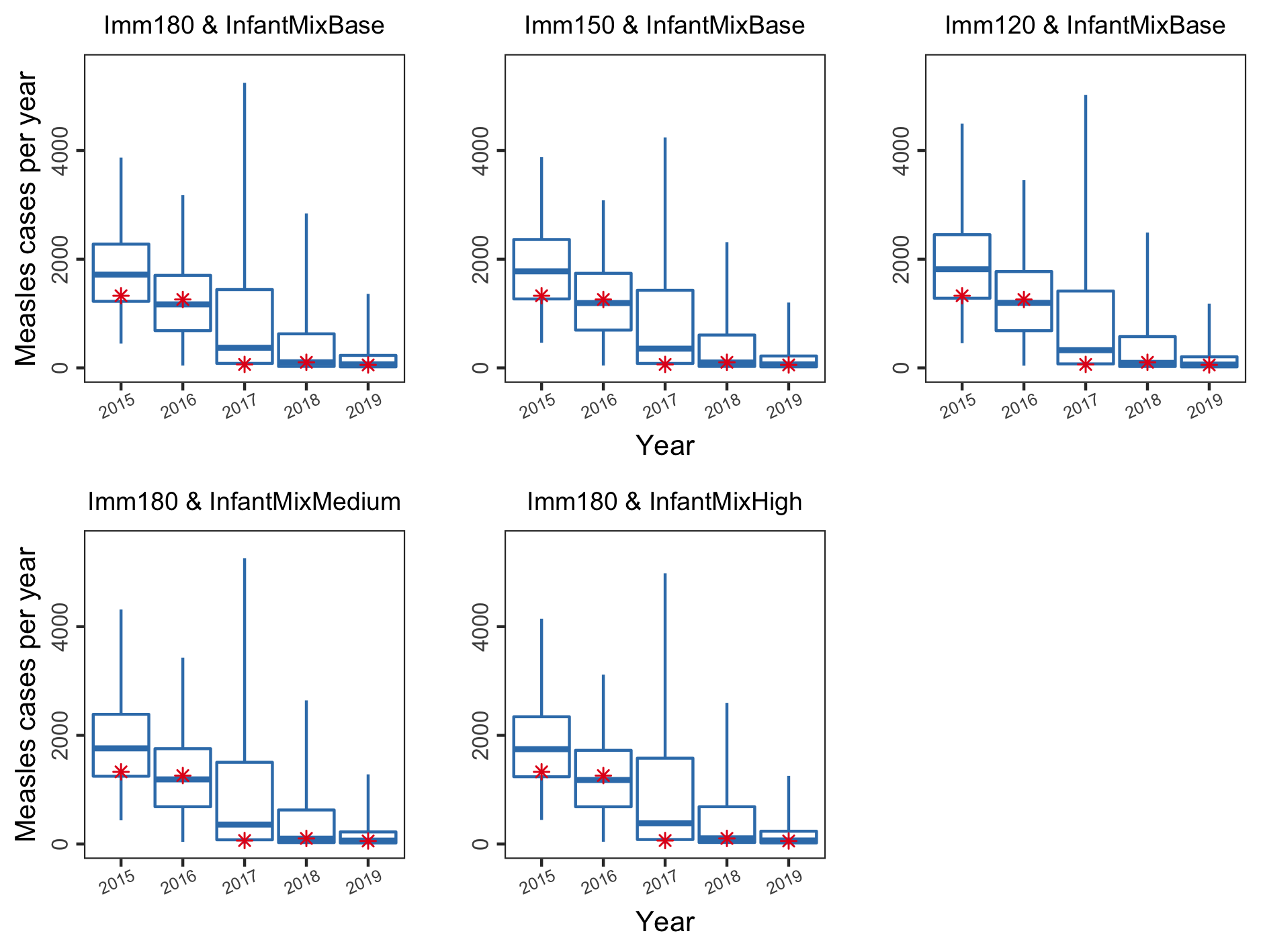
**Fig S6. Comparing migrant influx timing hypotheses based on out-of-fit prediction by year.** For the other migrant-related hypothesis settings, the three models assumed MigMixBase, BaseSeed, and MigSusHigh, because the comparison based on model fit identified this combination as the most plausible scenario. For infant-related hypotheses, the three models assumed Imm180 and InfantMixBase as baseline scenarios. See Table 1 for a summary of all hypotheses. The red asterisks represent reported measles cases for each year during 2015-2019. The boxplots represent the predicted yearly measles cases by the model replicates (8000 particles × 100 runs). The whiskers, box edges, and thick horizontal segment in the middle represent the 2.5th (or 97.5th) percentile, interquartile range, and median, respectively.



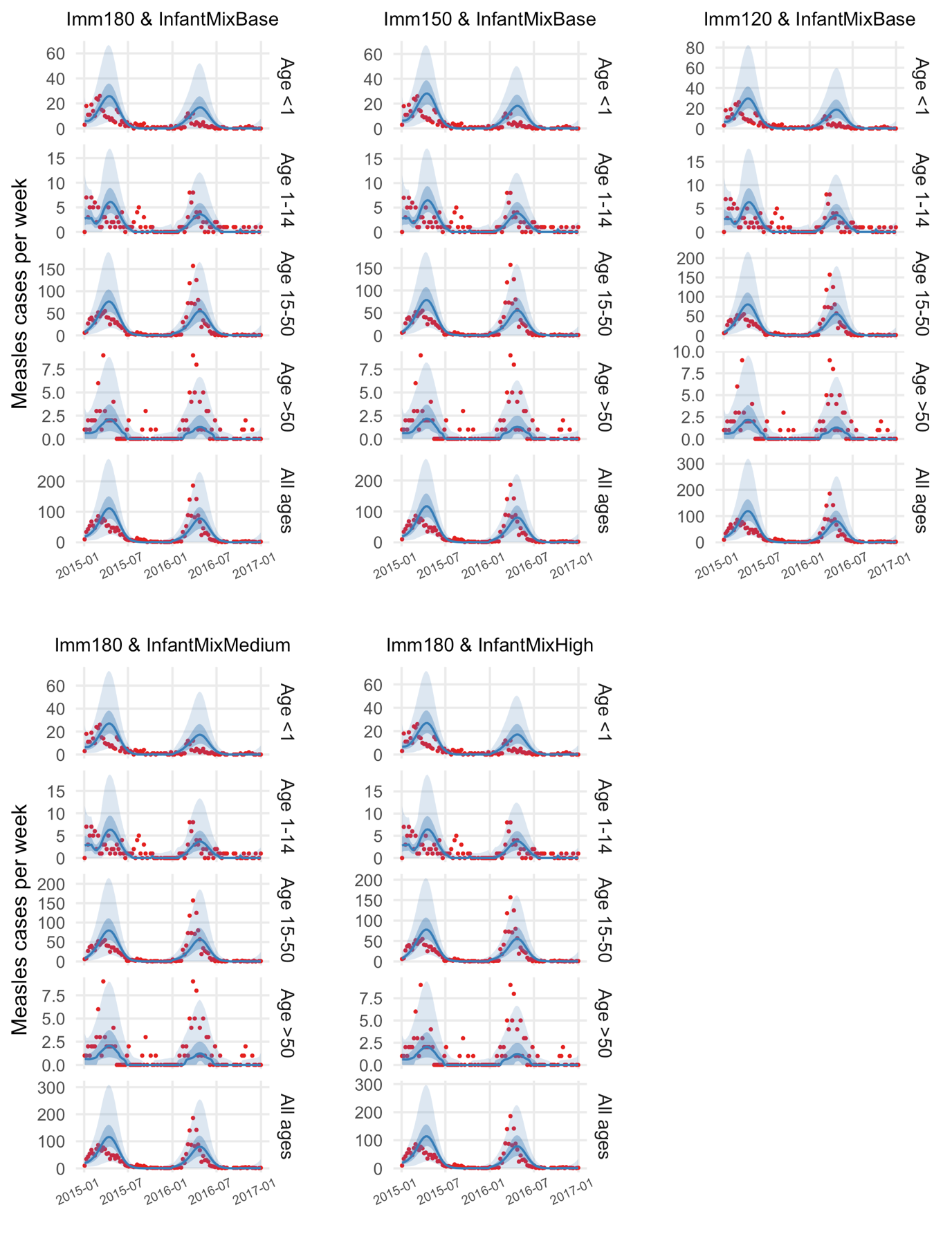
**Fig S7. Comparing migrant influx timing hypotheses based on out-of-fit prediction by age and week.** For the other migrant-related hypothesis settings, the three models assumed MigMixBase, BaseSeed, and MigSusHigh, because the comparison based on model fit identified this combination as the most plausible scenario. For infant-related hypotheses, the three models assumed Imm180 and InfantMixBase as baseline scenarios. See Table 1 for a summary of all hypotheses. The red dots represent reported measles cases for each week during 2015-2016; blue lines show the median of predicted cases; surrounding darker blue areas indicate interquartile range and lighter blue areas indicate 2.5 to 97.5 percentiles.



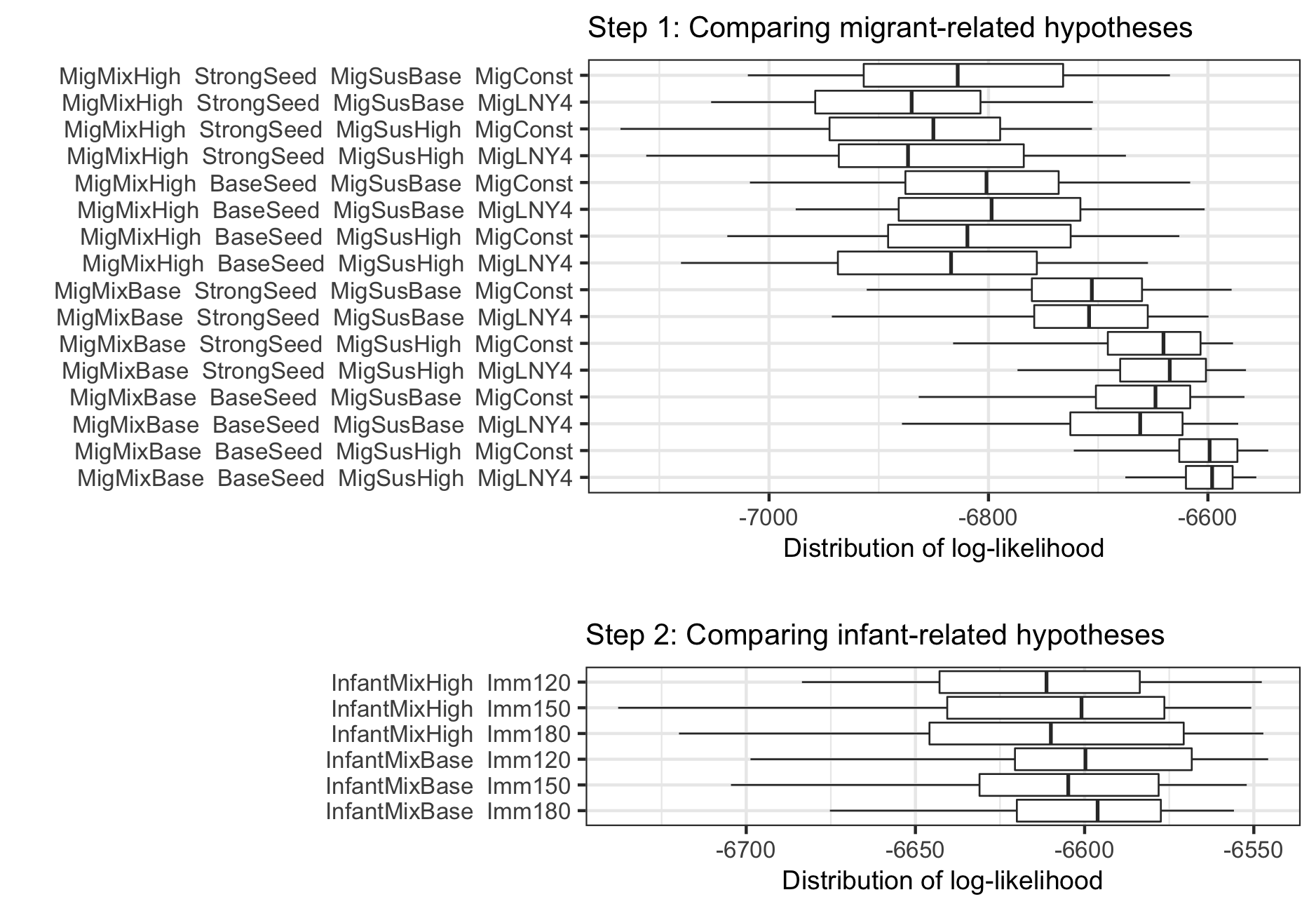
**Fig S8. Comparing infant-related hypotheses based on distributions of log-likelihood and RMSE.** We formulated two sets of infant-related hypotheses (increased infant mixing and duration of maternal immunity; corresponding to the 1st and 2nd terms in y-axis), and tested all combinations of the 2 sets (n=9, i.e., the number of rows of boxplots in each panel). For each model (or hypothesis combination), we conducted 100 model inference runs using a model-filter system, and show the distribution of log-likelihood (in **A**) and RMSE (in **B**) using boxplots (whiskers = 95% interval, box edges = Interquartile range, and thick middle bar = median). For migrant-related hypotheses, all models tested here assumed MigMixBase, BaseSeed, MigLNY4, and MigSusHigh, because the previous comparison step identified this combination as one of the most plausible scenarios. See Table 1 for a summary of all hypotheses.



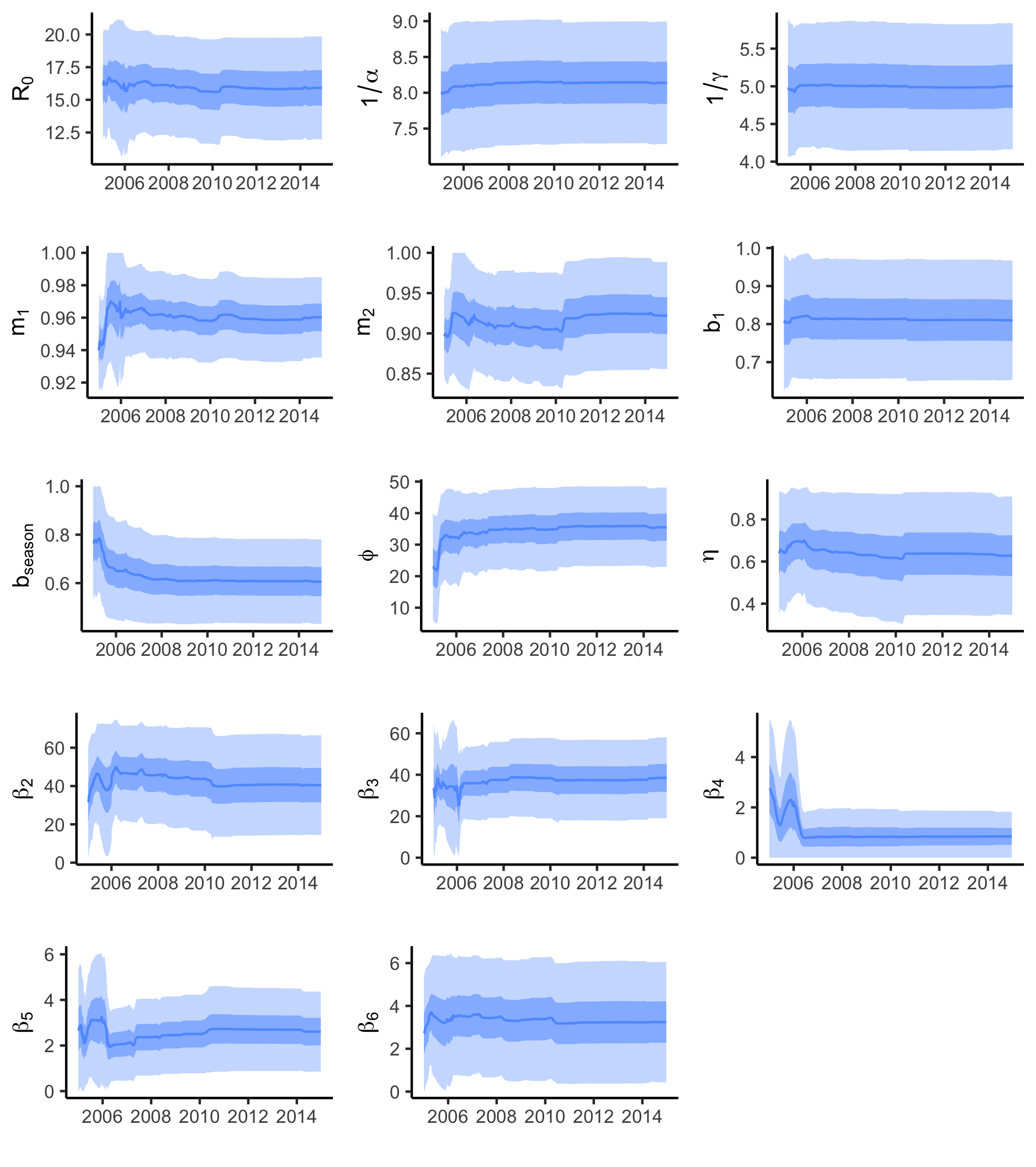
**Fig S9. Comparing infant-related hypotheses based on out-of-fit prediction by year.** For migrant-related hypotheses, all models tested here assumed MigMixBase, BaseSeed, MigLNY4, and MigSusHigh, because the previous comparison step identified this combination as one of the most plausible scenarios. See Table 1 for a summary of all hypotheses. The red asterisks represent reported measles cases for each year during 2015-2019. The boxplots represent the predicted yearly measles cases by the model replicates (8000 particles × 100 runs). The whiskers, box edges, and thick horizontal segment in the middle represent the 2.5th (or 97.5th) percentile, interquartile range, and median, respectively.



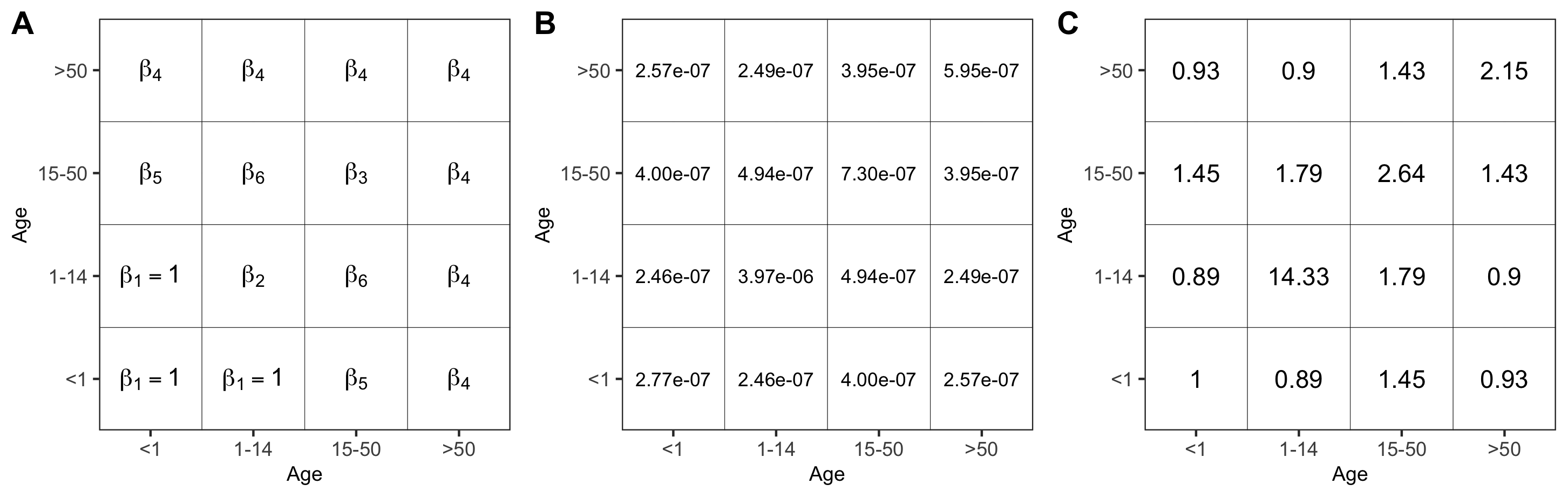
**Fig S10. Comparing infant-related hypotheses based on out-of-fit prediction by age and week.** For migrant-related hypotheses, all models tested here assumed MigMixBase, BaseSeed, MigLNY4, and MigSusHigh, because the previous comparison step identified this combination as one of the most plausible scenarios. See Table 1 for a summary of all hypotheses. The red dots represent reported measles cases for each week during 2015-2016; blue lines show the median of predicted cases; surrounding darker blue areas indicate interquartile range and lighter blue areas indicate 2.5 to 97.5 percentiles.



**Fig S11. Comparing hypotheses based on distributions of log-likelihood (sensitivity analysis with smaller observational error).** Theupper and lower panels were the same with Figs S3 and S8A, respectively, except that we used smaller observational errors in model inference and tested fewer combinations of hypotheses.



**Fig S12. Estimates of model parameters.** The model assumed MigSusHigh, BaseSeed, MigLNY4, MigMixBase, Imm180, and InfantMixBase (see Table 1 for a summary of all hypotheses). Blue lines and surrounding regions show the means and 50% and 95% credible intervals of estimates. represents basic reproductive number; , duration of the exposed state; , duration of the infectious state; and , mixing parameters; and , amplitudes of school term forcing and sinusoidal seasonal forcing, respectively; , the day of a year when the sinusoidal forcing reaches the maximum; , reporting rate in measles surveillance; (), relative transmission rates.



**Fig S13. Using contact data from the literature to inform the priors of to .** In our model, parameters (see **A**) govern mixing among age groups. Because we estimated the magnitude of to relative to , we divided all parameters by , thus setting to 1. Based on the contact data generated by Mistry et al., we derived a contact matrix (see **B**), with each element representing the contact rate between two age groups. To match the Mistry’s matrix (**B**) with our matrix (**A**), we divided the Mistry’s matrix by the element on Row 1 and Column 1, and used these scaled elements (see **C**) to inform the priors used.