SI for: Cost-effectiveness of pertussis vaccination schedule in Israel

Dean Langsam[1], Dor Kahana[1], Erez Shmueli[1,*], Dan Yamin[1,*]

[1] Department of Industrial Engineering, Tel-Aviv University, Tel-Aviv, Israel.

* Equally contributed

Population Dynamics

The model considers 27 age groups: 0-2 Months, 2-4 Months, 4-6 Months, 6-8 Months, 8-10 Months, 10-12 Months, 1-2 years, 2-3 years, 3-4 years, 4-5 years, 5-6 years, 6-7 years, 7-8 years, 8-9 years, 9-10 years, 10-11 years, 11-12 years, 12-13 years, 13-15 years, 15-18 years, 18-21 years, 21-25 years, 25-35 years, 35-45 years, 45-55 years, 55-65 years, and >65 years. Those age groups were taken from Israel's central bureau of statistics and were chosen as the finest resolution for good, significant data both on the demographics point of view and our primary case data.

Population Growth Model

We developed an aged-structured population growth model for the years 1951-2016. The model explicitly considered the variation in birth, mortality and immigration over the years as reported by the Israeli Central Bureau of Statistics (CBS). A snapshot of the year 1988 is depicted in Figure S1.

To detirmine the intial number of individuals in each health compartment in the transmission model, we fixed the age distribution of 1951, and ran the dynamic model for 30 years. Given that pertussis is highly contagious and that there was no vaccination during that time period, we

arbitrarily initalized the following paramters to $S(t=0)$=0.2, $I^s(t=0)$=0.02, $R(t=0)$=0.78, which we confirmed to have no effect on our simluations.

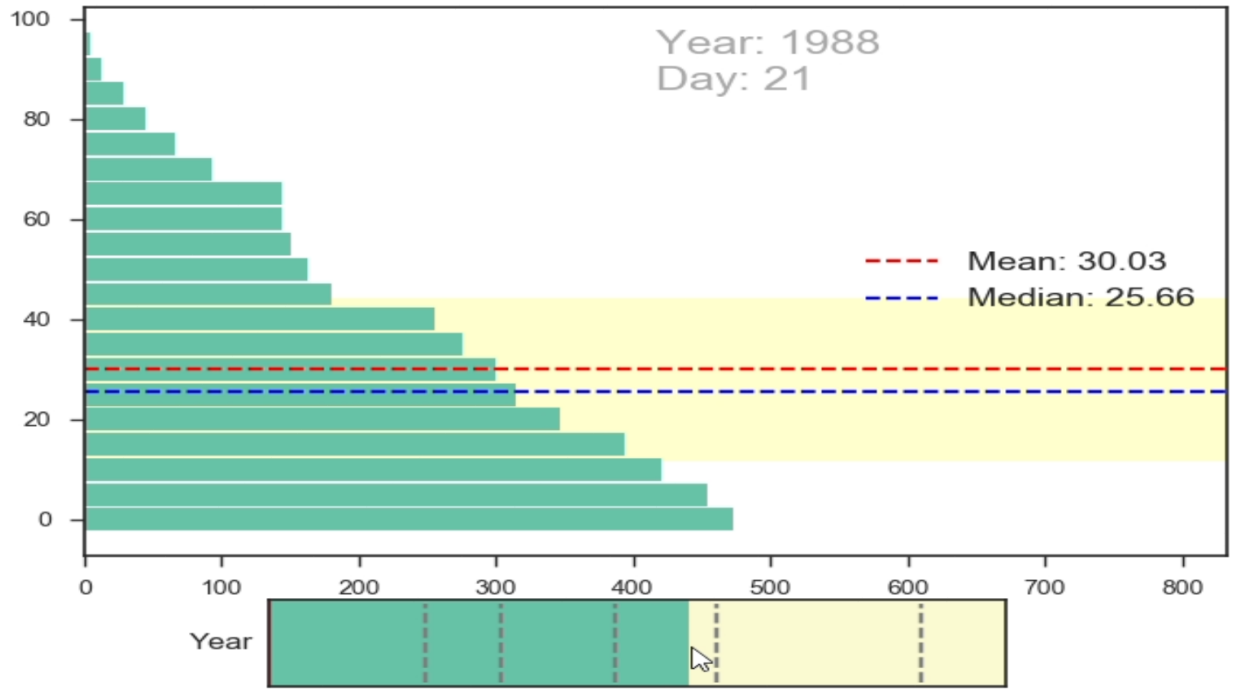

Figure S1: A snapshot of the age distribution at the year 1988, as considered by our population growth model. The information is presented as a population pyramide, where the X-axis represents the number of people (in thousands) and the Y-axis represents 5-year age groups. An interactive video showing the changes in age distribution over the years is available at [1].

Contact Mixing Patterns

Given a person of age group j and their infected contact of age group i, we evaluated the daily contact rate $C_{i,j}$, using data from a previous study [2]. In order to match the contact ages to our selected demographics and in order to have a symmetric matrix, we adjusted the matrix according to the means for reciprocal age group pairing as conducted by a previous study [3] (Figure S2).

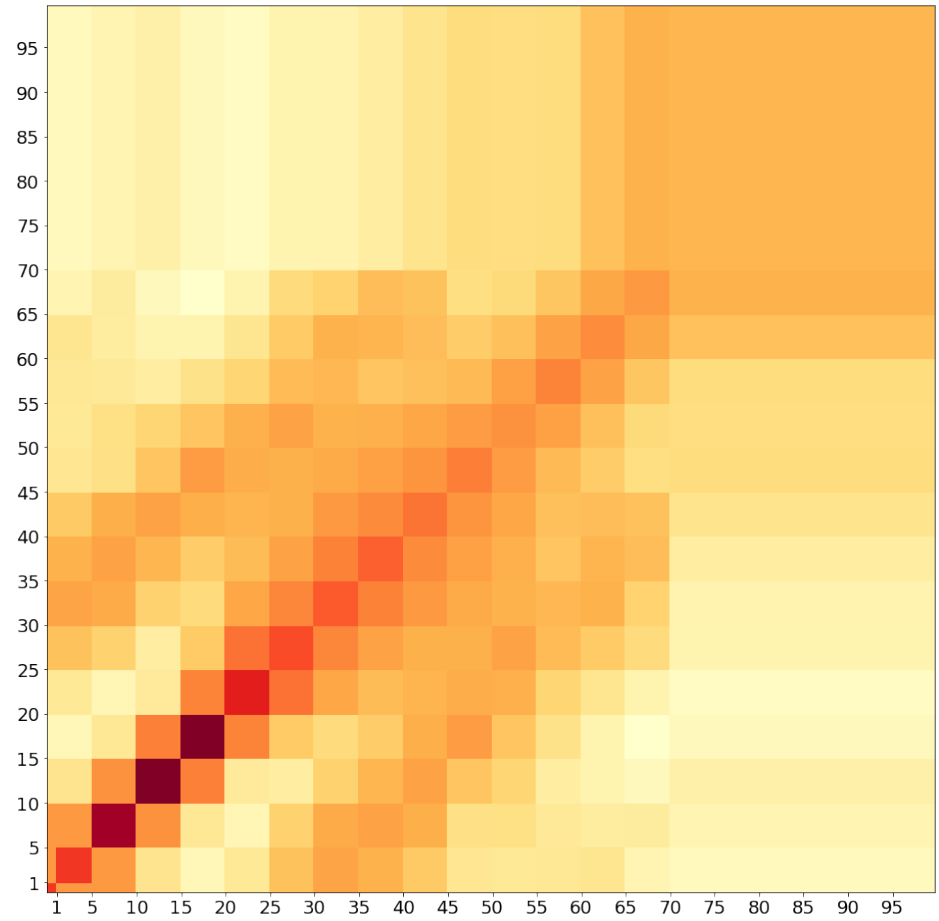

Figure S2: Contact matrix between different age groups. Darker colors stand for higher values

Parameters Estimation

Our model includes several missing parameters that were evaluated using primary data on pertussis. This primary data from the Israeli Ministry of Health (IMoH) encompass all reported pertussis cases for the entire Israeli population from January 1998 to December 2013. The dataset contains over 20,000 reported pertussis cases and includes demographic information such as age, ethnicity, and residence, as well as medical information such as date of GP and hospitalizations visits, and laboratory test results. We also used another, publically available (REF) IMoH dataset of yearly-aggregated cases of pertussis in Israel since 1951. We normalized both pertussis datasets

using census data from the Israeli Central Bureau of Statistics to calculate the number of pertussis cases per 100,000 population. Given that all records in Israel are electronically reported, we consider our data to be 100% of all reported cases. Yet, due to under reporting, misclassification, and asymptomatic infections, we assumed that only 1.5% are reported. This assumption is consistent with previous studies in the US, and inline with a previous serological study conducted in Israel [4,5].

Unifying Age Groups

In order to get a more representing fit, we needed to choose age groups that would capture population dynamics while still having a statistical significance within every group. The 27 age were summed over to create 3 new "super" age groups.

To determine which ages should be combined to each group, the cases were stratified by age. First, the cases and were fitted a general mixture model of 3 normal distribution using expectation minimization. Second, the age boundaries were chosen to reflect population dynamics, while still taking the general mixture model into account. Infants under 1 are the main driver of the disease, and 21 is the nominal age for military discharge. The age groups that were chose where 0y-1y, 1y-21y, >21y (Figure S3).

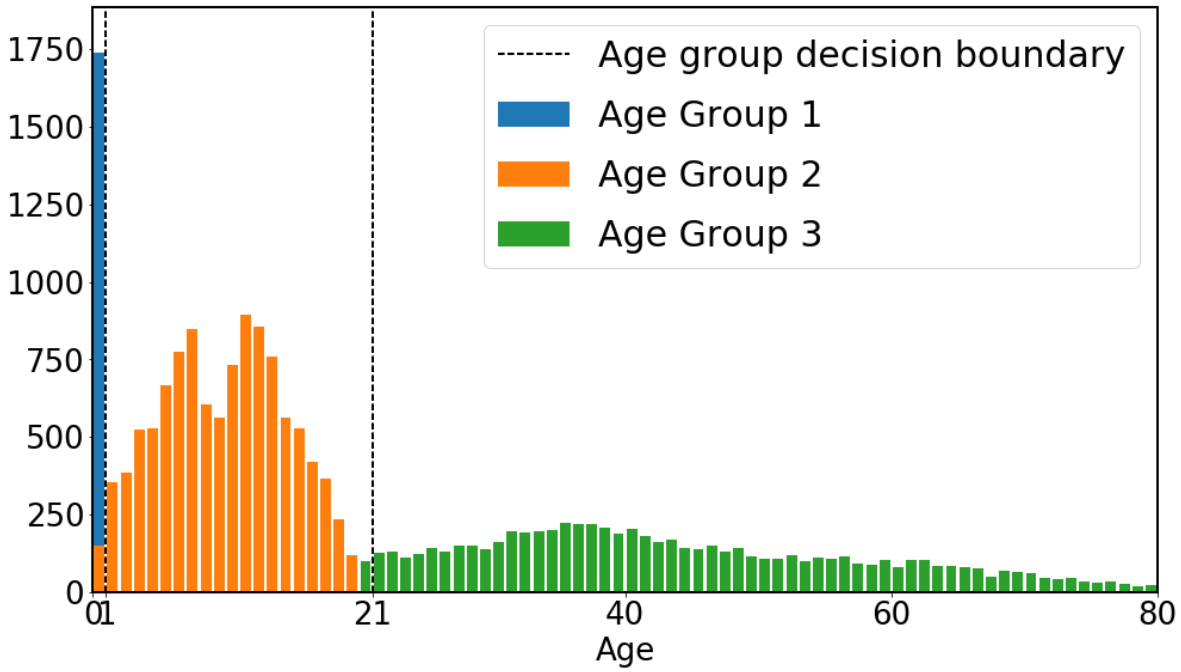

Figure S3: Case distribution by age with fitted labels from a general mixture model. Clear pattern can be seen of the three distinct age groups. Decision boundaries were chosen to reflect both the fitted distribution and the country's population dynamic.

Parameter estimation using MCMC

We estimated 4 model parameters for the force of infection and validated them by fitting the model to the age-stratified reports of our primary data representing 16 years from 1998-2014 of Israeli pertussis cases.

The age-dependent susceptibility was modelled by fitting susceptibility amplitude rates f_j for ages 0-1, 1-21, >21, multiplied by the periodical frequency Ω, periodical offset ϕ and an additive parameter ρ that represents a constant, "basic" force.

The proposed model Equations 1 has no closed-form solution, and the estimated parameters are very difficult to measure in a study. Therefore, we have to measure them through are primary population-level data.

Markov Chain Monte Carlo simulations were used to estimate these parameters (Figures S4 and S5). In that Bayesian approach, the parameters are not valued by their point estimate, but by a set of values that can be described as a probability distribution. The MCMC is in an iterative process where the parameter starts with an initial, prior probability (Table 2 main text), and then, using Bayes' Theorem, is updated to the posterior distribution.

We applied the MCMC using the Metropolis Hastings algorithm. We assumed a multi-normal proposal distribution, and updated the standard deviation to be the product of covariance between the parameters in the last 250 iterations and a scaling factor.

We assumed normal error between our model and the data, thus taking the likelihood function as the mean-squared error between 192 months over 3 age groups, in our data, and their corresponding predictions in our model. The data was multiplied by to 100 (1%) for infants and by 67 (1.5%) for other age groups, to account for reporting rates. After reaching convergence on the algorithm, we ran it for 25000 more iterations, with a constant proposal distribution.

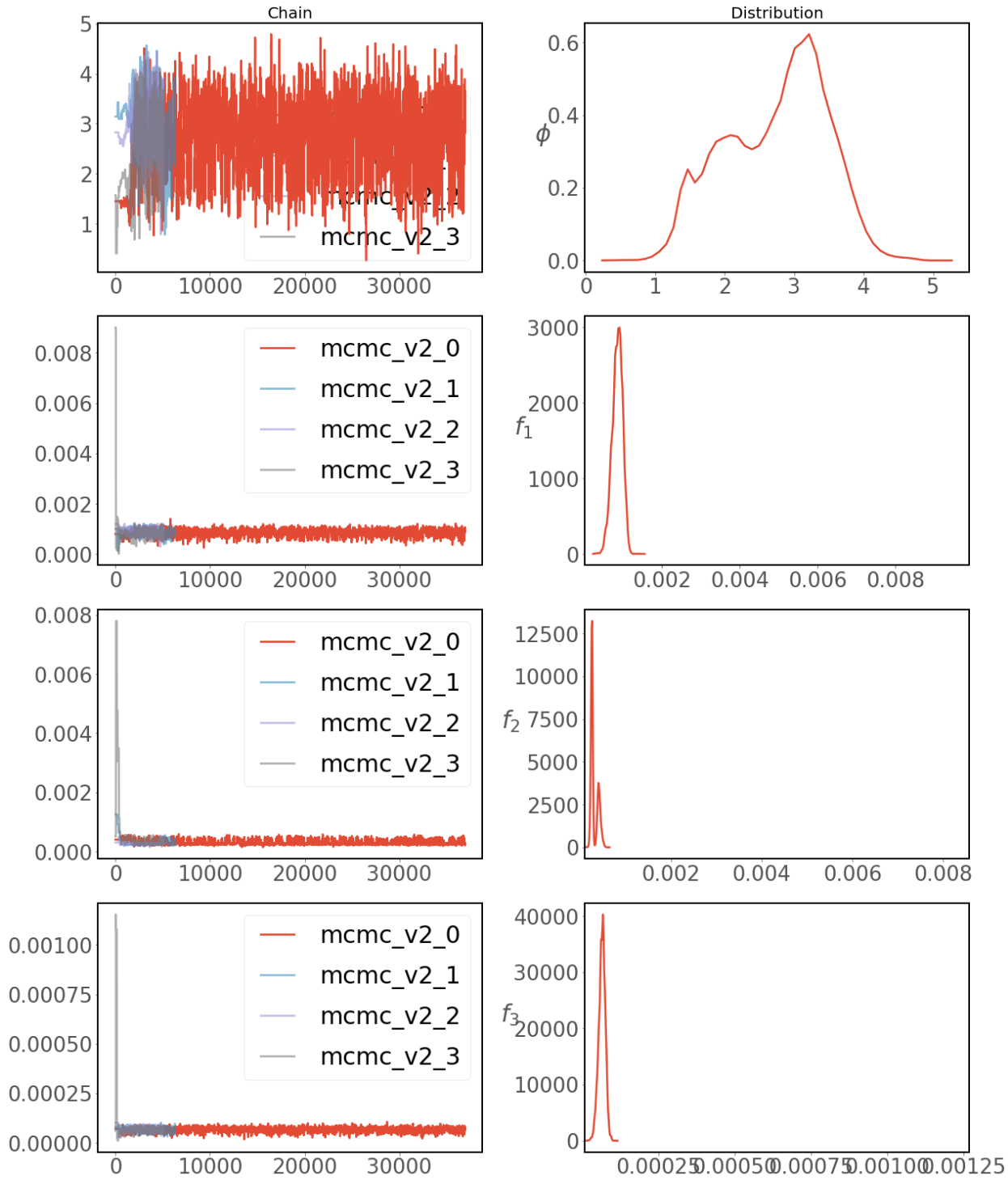

Figure S4: MCMC Chain Convergence with four chains

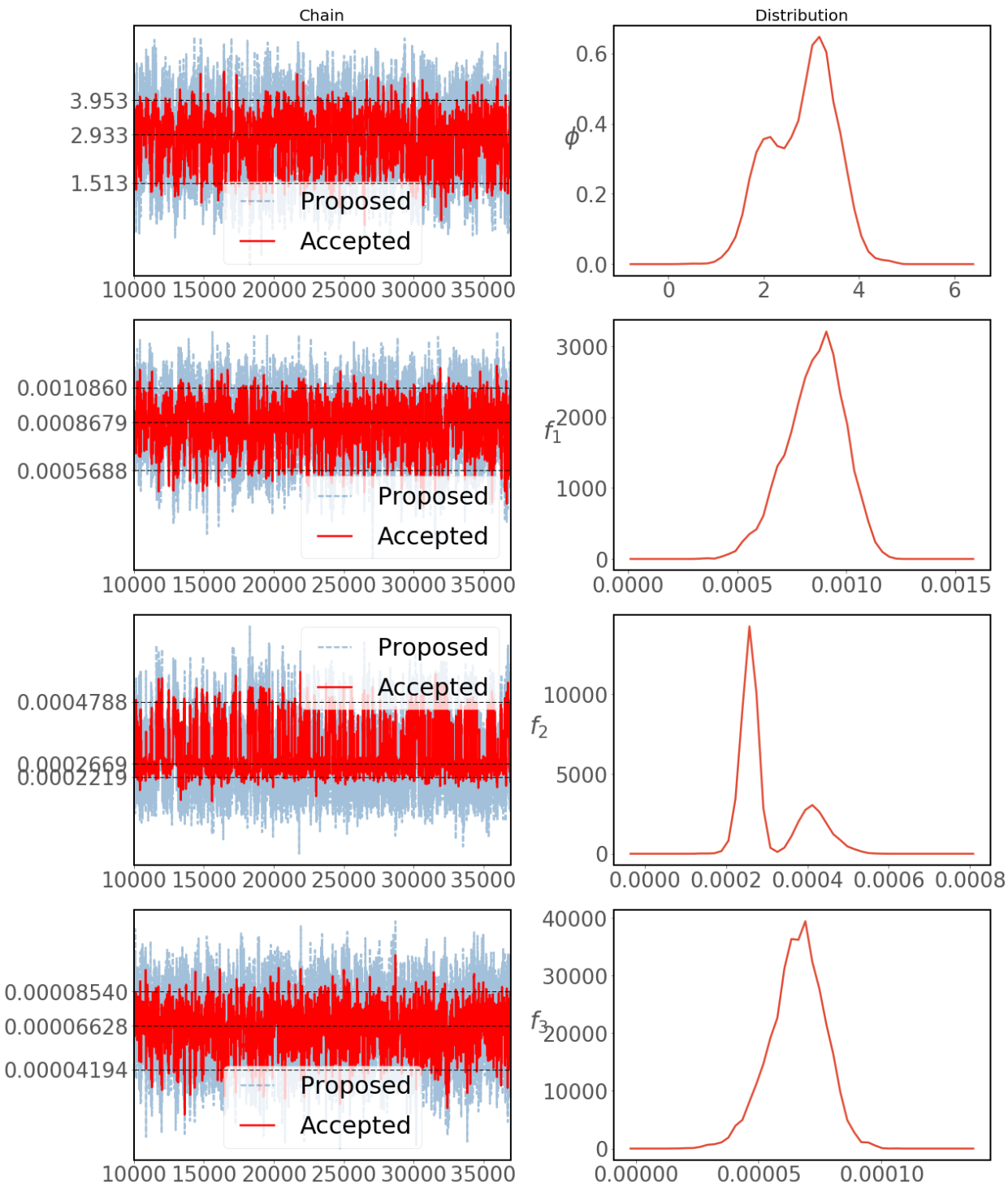

Figure S5: MCMC Chains and Distributions for Chosen chain after Gelman-Rubin convergence

Policy Determination

Effective Sample Size

Given a chain of a converged MCMC, we have chosen a subset of 2000 parameter sets (A 2000 by 4 matrix) - First we used a burning of 10,000 samples as this was the warming-up period were the chain is not reversible. Then, from the remaining 25,000 sample we have taken a thinning rate of $\max\left(\frac{1}{ess(p)}\right), p \in [\phi, f_1, f_2, f_3]$ where $ess(p) = \frac{N}{1 + 2 * \sum_{k=1}^{\inf} acf(k,p)}$ and $acf(k,p)$ is the autocorrelation of parameter p with lag k. We ended up with ~2000 samples for each MCMC scenario.

Simulation Process

From the ~2000 sample, we ran 5000 simulation iterations. At each iteration of the simulation we sample 1) one set of parameters, 2) a uniformly-generated maternal vaccine coverage 3) $p_{h_k}, k \in 1,2,3$ that represents the chance to be hospitalized given infection. p_{h_k} are taken from a beta distribution $p_{h_k} \sim Beta(a_k, b_k)$ where a_k is the number of hospitalized cases between 2004 and 2014, while b_k are the non-hospitalized cases. We ran a simulated future between the years 2014 and 2026 for each of the 30 policies. We compare each policy pairwise with a base policy such that for every tested policy we have 1) aggregated cases for the 12-year period, 2) aggregated hospitalizations, 3) aggregated cases of children under 1, and 4) aggregated hospitalizations of

children under 1. Those are simulated results based on the same parent distribution, thus assuring a result is significant without the need to account for multiple comparisons.

Parameter	Distribution
Maternal Coverage	$Uniform(0,1)$
p_{h_1}	$Beta(500,1147)$
p_{h_2}	$Beta(201,9560)$
p_{h_3}	$Beta(6703,129)$

Table 2: Draw distribution for simulation-level parameters.

References

[1] Yamin D. LEMA 2019. http://danyamin.com/.

[2] Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. PLoS Med 2008;5:e74. doi:10.1371/journal.pmed.0050074.

[3] Medlock J, Galvani AP. Optimizing influenza vaccine distribution. Science 2009;325:1705–8. doi:10.1126/science.1175570.

[4] Rendi-Wagner P, Tobias J, Moerman L, Goren S, Bassal R, Green M, et al. The seroepidemiology of Bordetella pertussis in Israel--Estimate of incidence of infection. Vaccine 2010;28:3285–90. doi:10.1016/j.vaccine.2010.02.104.

[5] Atkins KE, Fitzpatrick MC, Galvani AP, Townsend JP. Cost-Effectiveness of Pertussis Vaccination During Pregnancy in the United States. Am J Epidemiol 2016;183:1159–70. doi:10.1093/aje/kwv347.