Supplementary Information for

Community structure mediates Sabin 2 polio vaccine virus transmission

Famulare, M, Wong, W, Haque, R, Platts-Mills, JA, Saha, P, Aziz, AB, Ahmed, T, Islam, MO, Uddin, MJ, Bandyopadhyay, AS, Yunus, M, Zaman, K, Taniuchi, M

Mami Taniuchi

Email: mt2f@virginia.edu

This PDF file includes:

- Supplementary text
- Figures S1 to S8
- Tables S1
- SI References

Other supplementary materials for this manuscript include the following:

- Datasets S1

Supplementary Text

Supplemental - Modeling Individuals

Individuals

Agents in our model represent individuals defined by their age, sex, marital status, and immunity (see below, ***Supplemental* – Modeling Immunity**). Female individuals are also defined by their fertility, which determines the probability of birth given marital status and age. Death occurs when

individuals reach a pre-determined death age based on calibrated mortality rates. A visual summary of model calibration fits with regards to fertility and mortality are presented in **Fig. S7.**

Fertility

Fertility is determined using a joint probability function conditioned on four factors:

$$P(Birth) = P(Birth|age, n_{children}, t_{birth}, marriage) \quad (1)$$

$$P(Birth|\, x = age) = \begin{cases} Ax^4 + Bx^3 + Cx^2 + Dx + E & if\ 50 < x < 15 \\ 0 & else \end{cases} \quad (2)$$

$$P(Birth|\, x = n_{children}) = \begin{cases} \dfrac{c}{1 + e^{-k(x - x_0)}} + y_0 & if \quad x > 1 \\ 1 & else \end{cases} \quad (3)$$

$$P(Birth|\, x = t_{birth}, age) \quad (4)$$

$$= \begin{cases} \alpha \left(\dfrac{k_{age}}{\lambda_{age}}\right)\left(\dfrac{x}{\lambda_{age}}\right)^{k_{age}-1} \left(1 - e^{(-x/\lambda_{age})^{k_{age}}}\right)^{\alpha_{age}-1} e^{(-x/\lambda_{age})^{k_{age}}} & if\ x > 1 \\ 0 & else \end{cases}$$

$$P(Birth|\, marriage) = \begin{cases} 1 & if\ married \\ 0 & else \end{cases} \quad (5)$$

Equation 2 is a polynomial equation that represents the maximum, base fertility of a married female for her age. The number of previous births ($n_{children}$) and birth interval period (t_{birth}) are modifiers that decrease fertility based on the individual's previous history. P(Birth|$n_{children}$) is a sigmoid function that penalizes having additional children and P(birth| t_{birth}) is the probability density function of an exponentiated-weibull distribution that prevents births from occurring too close to one another. The parameters of this exponentiated-weibull distribution are age-specific to incorporate the increasingly large birth intervals associated with age. We assume that only married females give birth. P(Birth|marriage) is a binary flag that prevents single or widowed females from giving birth. Parameter estimates are in the **Supplemental Dataset S1.**

Birth Interval Calibration

The birth interval function describes the expected wait-time between births and is defined by an exponentiated Weibull distribution. We calibrated this function to the data reported in the 2014 Bangladesh Demographic Health Survey. Reported birth intervals were categorized into six bins (7-17 months, 18-23 months, 24-35 months, 36-47 months, 48-59 months, and 60+ months) for four age groups (15-19, 20-29, 30-39, and 40-49 years of age). We independently fit an exponentiated Weibull function to each of the age groups to obtain four sets of age-group specific parameters by maximizing a multinomial likelihood function:

$$L(\lambda_{age}, k_{age}, \alpha_{age}) = \prod_{i=1}^{6} p_{i,age}^{x_{i,age}} \tag{6}$$

$$p_{i,age} = \int_{t_1}^{t_2} P(Birth|x = t_{birth}, t)dt \tag{7}$$

where p_i is the proportion of births occurring in the time interval [t1, t2] in months for that age group. Parameter estimates are in the **Supplemental Dataset S1.**

Marriage Calibration

To incorporate marriage into our fertility calibrations, we defined marriage as:

$$P(marriage|\, x = age) = \int_{t-1}^{t} \frac{kp_0 e^{rx}}{k + p_0 e^{rx} - 1} \quad (8)$$

where t is the current timestep and the other parameters are associated with a logistic growth curve (Verhulst equation). We calibrated our marriage function to the proportion of married individuals by age data in the 2014 Bangladesh Demographic Health Report using the a non-linear least-squares optimizer (scipy.optimize.curve_fit). Note that this equation is only used for calibration. The actual marriage rates in our full model are influenced by household and population dynamics (***Supplemental - Household demographic structure****).* Parameter estimates are in the **Supplemental Dataset S1.**

Base Fertility and Child Preference Calibration

For the remaining parameters, we calibrated our fertility function to the lifetime number of children per married female and the age-specific fertility rates reported in both the 2004 and 2014 Bangladesh Demographic Health Surveys. Reported lifetime children numbers per married female were split into eight different age groups (15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49 years of age). We calibrated our model against age groups less than 35 years of age because older age groups are more likely to be influenced by past fertility trends.

By assuming independence between each of the components of our fertility equation, our birth function (equation 1) can be rewritten as:

$$P(Birth) = P(Birth|age) * P(Birth|n_{children}) * P(Birth|t_{birth}) * P(marriage|age) \quad (9)$$

The parameters for $P(Birth|t_{birth})$, the birth interval equation (equation 4) and $P(marriage|age)$, the marriage probability function (equation 8),were independently calibrated.

For the remaining parameters associated with the base fertility rate (equation 2) and the child preference function (equation 3), we created a simplified, agent-based fertility simulator consisting of 10,000 female individuals. Simulated individuals track the number of previous births, the time since most recent birth (birth interval), age, and marital status. Individual starts at age = 15 (the minimum fertility age) and are aged using one-year timesteps until they reach age 50, which we assumed to be the maximum fertility age. At each time step, we used our modified birth function (equation 9) to determine whether an individual will give birth given their age and past birth history. We then compared our simulated estimates to the reported data using a population Monte Carlo (PMC) Approximate Bayesian Computation (ABC) to obtain posterior parameter estimates. We did not re-calibrate our birth interval parameters. Our PMC contained 3000 particles and was iterated four times. Initial priors were drawn from normal distributions whose parameters were loosely based off hand calibration attempts. After each iteration, we defined distance as the Kullback-Leibler divergence and modified our acceptance threshold such that only the top 30, ten, and five percent of the distance distribution would be accepted for the second, third, and fourth iteration. The median posterior value for each parameter were used as our calibrated point estimates in our full model. Parameter estimates are in the **Supplemental Dataset S1.**

Defining Mortality

Individuals in our model are assigned a death age at birth. Mortality rates were defined using a piecewise function:

$$P(death|\ sex, x = age) = \begin{cases} max\left(m,\ e^{(B_{child}+A_{child}x)}\right) & if\ x < 12 \\ max\left(m,\ e^{(B_{elder}+A_{elder}x)}\right) & if\ x > 30 \\ \text{m} & else \end{cases} \quad (10)$$

where m is the minimum mortality rate and calibrated to both the 2004 and 2015 WHO life tables for Bangladesh. These life tables reported the mortality rate per 10,000 individuals, split into different age groups. Individual fits were made for male and females. Linear regression models were fit to the natural log mortality rate using a non-linear least squares optimizer (scipy.optimize.curve_fit). We assume a minimum mortality rate of 0.001 (the minimum number of possible deaths per 1000 individuals) to ensure that death is possible at all ages. Point estimates for the linear regression model are in the **Supplemental Dataset S1.**

Historical Fertility and Mortality Projection

To approximate the rapid, historical decline in fertility and mortality, we instituted a two-phase burn-in lasting 180 years. This burn-in was divided into two phases: $t_{historical}$, a 140-year historical period with high fertility and mortality rates, and $t_{transition}$, a 40-year transition period where fertility and mortality rates decline to their 2014 levels. We defined the historical fertility rate as:

$$P(Birth_{historical}|age, n_{children}, t_{birth}, marriage, t\ ) = H_{fertility}(t) * P(Birth_{2014}|age, n_{children}, t_{birth}, marriage) \quad (11)$$

$$H_{fertility}(t) = \begin{cases} \frac{t}{40}\left(1 - \omega_{fertility}\right) + \omega_{fertility} & if\ t < 40 \\ \omega_C & else \end{cases} \quad (12)$$

$$\omega_{fertility}(x = age) = \omega_C(\omega_a x + \omega_\beta) \quad (13)$$

Historical fertility rates are obtained by adding an age and time-dependent multiplier, H, to the 2014 fertility rate. t is the number of years prior to 2014. Assuming only the base fertility rate (equation 2) changes through time, ω represents the per-decade increase in base fertility for each decade prior to 2014. ω_C is the maximal historical increase in fertility, which was set to three to make fertility rates increase from 2 to 6 (an approximation of the historically high fertility rates in Bangladesh) children per female. ω_a and ω_β are parameters to a standard linear function and were calibrated to the 2014:2004 base fertility ratios in the 2014 and 2004 Bangladesh Demographic Health Survey data using a non-linear least squares optimizer (scipy.curve_fit). Point estimates for ω_C = 3, $\omega_a = 0.85$, ω_β = 0.18.

Similar logic was applied to the mortality rates, except that we scaled it proportionally to the 2004 and 2014 mortality rates. This did not have a significant impact, as the mortality rates between these two time periods were relatively the same. Here, mortality was capped at their 2004 levels.

$$P(death_{2014}|age, sex, t) = H_{death}(t) * P(death_{2014}|age, sex) \quad (14)$$

$$H_{death}(t) \quad (15)$$
$$= \begin{cases} \frac{t}{40}\left(P(death_{2014}|age, sex) - P(death_{2004}|age, sex)\right) + P(death_{2004}|age, sex) & if\ t < 40 \\ \alpha_C & else \end{cases}$$

$$\alpha_C = P(death_{2004}|age, sex) \quad (16)$$

These equations were calibrated to the WHO Bangladesh life tables for 2014 and 2004 using the same procedure as above.

Supplemental - Household demographic structure

Model Structure

Traditional, rural Bangladesh households are patriarchal, stem families. Stem families are a type of family system in which one child (commonly the firstborn son) stays within the family home while other children move out to live in with their in-laws or to start households of their own. Households in our model are represented as trees and based off an anthropological framework described in (1). Each node represents either a single, unmarried individual or a marital unit (one male and female). Although other types of marital units exist in traditional Bangladeshi society (particularly polygynous unions), they were not simulated. The root node of each tree represents the founding couple. The male in the root node represents the patriarch. Newly born individuals are assigned a new node connected to its parental node. This allows households to grow organically while preserving the hierarchical relationship between household members. Nodes can have one of six statuses: single (never married), married, widow, widower, orphan, and dead. Dead terminal nodes are pruned at each timestep while dead internal nodes are preserved to maintain downstream hierarchies. We define orphans as individuals whose immediate parents have died and who are younger than the minimum eligible marriage age. Widowers, orphans, and single females are considered dependent states and preferentially kept in pre-existing households (see below—*Succession and household splitting*). Although this framework was designed with Bangladesh society in mind, it can be reasonably extended to other traditional, rural societies where large extended families are common. However, because our model equates household with families of related individuals, it does not simulate living arrangements resulting from non-familial roommate situations (ie working camps or dormitories).

Our model updates individuals, households, *baris*, and villages using discrete time steps. During each timestep, our model:

1. Accounts for births and deaths
2. Updates the age, fertility, and immunity of living individuals

3. Updates households by removing dead individuals
4. Updates *baris* by removing dead households and creating new households generated during move out or succession events
5. Updates villages by removing dead *baris* and creating married couples by moving eligible, single females to their spouse's household. Newly married partners are sourced from different *baris*.

Villages are generated from a collection of single-household *baris* using a two-phase burn-in to establish a simulated population with the rapid decline in fertility and mortality observed in Bangladesh (2). Our burn-in was divided into two phases (above, *Historical Fertility and Mortality Projection*): $t_{historical}$, a 140-year historical period with high fertility (six children per female) and mortality rates, and $t_{transition}$, a 40-year transition period where fertility and mortality rates decline to their 2014 levels (two children per female). As we are interested in transmission sourced from within Matlab over a short period of time, migration into or out of our simulated villages was not allowed for convenience.

Marriage

Newly born individuals are assigned an earliest marriage age drawn from normal distributions whose mean and standard deviations were obtained from the 2014 Bangladesh Demographic Health survey. The average earliest marriage ages are 27.3 and 19.3 for males and females respectively. Once eligible, individuals are randomly paired with a partner of the opposite sex. Eligible partners are drawn from the village but excludes *bari* members to avoid incest. During marriage, females are removed from their original household and added to their spouse's node. Only females in married nodes can give birth. Should one member of the marital unit die, the node status is updated to either widow or widower. We do not allow remarriage in our model.

Succession and household splitting

The patriarchal, stem structure of Bangladesh households influences household composition through time, particularly with regards to inheritance and succession (3). Traditional Bangladesh households are governed by a patriarch (usually the eldest male) from which all other individuals either descend from or added through marriage. Inheritance and succession favor males, and elders preferentially live with their eldest son. This hierarchy reorganizes itself upon patriarch death, with the eldest son assuming the role of patriarch; younger male siblings leave after marriage to form their own households in the bari.

Succession is the primary mechanism for household formation in our model. Our household model initiates succession when the patriarch in the root node dies. During succession, our model:

1. Identifies viable subtrees within the original household tree. Viable subtrees are identified by traversing the household tree in level order (visiting every node on a level before going to a lower level, excluding the root node) and identifying the inner-most, still-living nodes. Internal nodes identified through this process represent married couples with descendants or widows/widowers with descendants. Terminal nodes represent childless couples, unmarried individuals of eligible marriage age, childless widows/widowers, or orphans whose parents have died but are too young to marry.
    a. If the root node still contains a living member (the matriarch), it is treated as a terminal node containing a childless widow.
2. Creates new households by extracting viable subtrees from the original household. Subtrees are rooted on the inner-most, still-living nodes identified in step one. Those rooted on couples (with or without children). Those rooted on single individuals, widows or orphans are considered unviable (see step three). Once extracted, these new trees represent new households within the bari. One of these trees is randomly chosen to be the successor and inherit the original household.

3. Reassigns non-viable subtrees to the newly created successor household as immediate descendants of the root.
    a. If no viable successor household was created, non-viable subtrees are assigned to a random pre-existing household in the bari. Only if no pre-existing households exist will these non-viable subtrees be used to create new households.

Our succession framework preserves the familial relationships between individuals within subtrees but does not preserve the familial relationships between subtrees (cousin, uncle, and other distant relationships are more likely to be lost). Despite this, we can still identify whether individuals are generally related to one another as new households are placed in the original *bari*. Our framework also prevents elders, widowers, orphans from living on their own. This is motivated by the living arrangements of these individuals in traditional, rural Bangladesh society. Bangladeshi elders almost exclusively live with their descendants, with a strong preference for the eldest son (3). Widowers have historically moved back to their childhood household. Although our model does not simulate these dynamics exactly, allowing these individuals to live on their own resulted in a higher proportion of single households than expected for Bangladesh society. Only after incorporating step three were we able to reduce the proportion of simulated single households to match that of Bangladesh society. In many ways, this mirrors the difficulty of real societies to care for marginalized or otherwise non-independent members.

Supplemental: Modeling Immunity

We model immunity using the mathematical model described a previous study (4). Our model relates oral susceptibility to infection, shedding duration, and viral shed concentrations to pre-exposure immunity ($N_{ab_{pre}}$); high $N_{ab_{pre}}$reduces oral susceptibility to infection, shedding duration, and viral shed concentrations. Our model also allows for waning immunity, which depends on the peak post-infection immunity ($N_{ab_{peak}}$). Immunity is defined as the OPV-equivalent antibody titer,

an indirect measure of immunity representing the serum neutralizing antibody titers due to OPV immunization or natural wild poliovirus infection. OPV induced antibody responses are predictive of fecal shedding and susceptibility while inactivated poliovirus (IPV) induced responses are not (5–9). Equations for shedding duration after OPV challenge, poliovirus stool concentrations, oral susceptibility to infection, and waning immunity were taken from supplemental equations S1-S6 from the previous study (4) and are copied here for clarity.

Oral susceptibility

Oral susceptibility is modeled as a dose-response relationship between infection, oral poliovirus ingestion, and pre-exposure immunity.:

$$P\left(infection\middle|dose, N_{ab_{pre}}, strain\right) = 1 - \left(1 + \frac{dose}{\beta_{strain}}\right)^{-\alpha\left(N_{ab_{pre}}\right)^{-\gamma}} \quad (11)$$

where dose refers to the viral dose and $N_{ab_{pre}}$ refers to pre-exposure immunity. α and β_{strain} are standard beta-Poisson parameters and γ captures the reduction in infection probability with increasing immunity. β_{strain} is type-specific and different for Sabin 1, Sabin 2, Sabin 3, and wild poliovirus. Parameter estimates for Sabin 2 and WPV are found in the supplemental of the original paper (4).

Shedding Duration and Shedding concentration

We assumed a log-normal survival distribution for shedding duration:

$$P\left(shedding\ at\ \mathrm{t}\middle|N_{ab_{pre}}; infected\ at\ t = 0\right) \quad (12)$$
$$= \frac{1}{2}\left(1 - erf\left(\frac{ln(t)\ -\ (ln(\mu)\ -\ ln(\delta) log\_2\ (N_{ab_{pre}})}{\sqrt{2} ln(\sigma)}\right)\right)$$

where μ is the median duration in days for immunologically naiive individuals ($N_{ab} = 1$), δ describes the decrease in median duration with increasing immunity, and σ describes the shape of the distribution. Infection durations ($t_{duration}$) are assigned at the start of the infection and determined by sampling from the inverse distribution. Once infection age exceeds infection duration, individual cease to shed virus and are considered uninfected.

To model viral load over time, we assume a quasi-log-normal shedding profile:

$$concentration(t \mid N_{ab_{pre}}, age) = max(10^{2.6}, \quad (13)$$

$$(peak\ CID50/g|N_{ab_{pre}}, age) * \left(\frac{exp(\eta - \frac{v^2}{2} - \frac{(log(t) -)^2}{2\left(v + \xi log(t)\right)^2}}{t} \right))$$

$$log_{10}\left(peak\ CID50/g \middle| N_{ab_{pre}}, age\right) \quad (13a)$$

$$= \left(1 - klog_2(N_{ab})\right) log_{10}\left(peak\ CID50/g \middle| N_{ab_{pre}} = 1, age\right)$$

$$log_{10}\left(peak\ CID50/g \middle| N_{ab_{pre}} = 1, age\right) = \begin{cases} S_{max} \\ (S_{max} - S_{min}) exp\left(\frac{7 - age}{\tau}\right) + S_{min} \end{cases} \quad (13b)$$

Shedding concentrations are evaluated at time points, t, falling within the interval (0, $t_{duration}$]. Parameter estimates and further details are found in the supplemental of the original paper (4).

Immune Waning

Immune waning is modeled as a power law:

$$N_{ab}(t) = max\left(1, \mathrm{N}_{\mathrm{ab}_{\mathrm{peak}}} t^{-\lambda}\right) \quad (14)$$

where t is measured in months and $\mathrm{N}_{\mathrm{ab}_{\mathrm{peak}}}$ is the peak post-infection immunity. Parameter estimates are found in the supplemental of the original paper (4).

Immune boosting

Previously, we inferred $\mathrm{N}_{\mathrm{ab}_{\mathrm{peak}}}$ based on the shedding durations of individuals whose vaccination history was known. Conditioning our analysis to individuals with known vaccination histories allowed us to infer $\mathrm{N}_{\mathrm{ab}_{\mathrm{peak}}}$from individuals with multiple reinfection histories (*e.g.* a 3x bOPV vaccination course) without having to specify the relationship between $\mathrm{N}_{\mathrm{ab}_{\mathrm{pre}}}$ and $\mathrm{N}_{\mathrm{ab}_{\mathrm{peak}}}$. However, this approach was untenable for this study because individuals in Matlab, Bangladesh have complicated immune histories due to overlapping vaccination campaigns and because we wanted to dynamically model immune dynamics following transmission and potential reinfection. To dynamically model reinfection, we needed to quantify the boost in immunity (θ) following infection. Previous serology and viral shed studies strongly suggest θ diminishes with higher pre-exposure immunity (10, 11).

We defined $\mathrm{N}_{\mathrm{ab}_{\mathrm{peak}}}$ and θ as:

$$\mathrm{N}_{\mathrm{ab}_{\mathrm{peak}}} = N_{ab} * \theta(N_{ab}) \quad (15a)$$

$$\theta(N_{ab}) = a + b log_2(N_{ab}) \quad (15b)$$

where N_{ab} represents the pre-exposure antibody titer and θ is the boost response measured in log2 units. We calibrated θ to the post-exposure antibody ratios obtained from sera collected from 150 newborn infants monitored for poliovirus infection in 1953 (11). We define post-exposure

antibody ratios are defined as the ratio between post-exposure and pre-exposure antibody titers and a direct measurement of immune boost. We first fit an ordinary least squares model to the post-exposure antibody ratios against the pre-exposure antibody titers, which revealed a negative correlation between θ and log2 N_{ab} but with heteroskedastic variance (**Fig. S8**).

The heteroskedasticity associated with high N_{ab} could be due to biological factors, such as immune exhaustion, or an artifact due to limit of quantification issues associated with sampling methodology. To differentiate these two, we split θ into two components, θ_{bio} and $\theta_{sampling}$ where:

$$\bar{\theta}(N_{ab}) = \bar{\theta}_{bio}(N_{ab}) + \bar{\theta}_{sampling} \tag{16}$$

$$\bar{\theta}(N_{ab}) = \alpha + \beta log_2(N_{ab}) + \bar{\theta}_{sampling} \tag{16a}$$

$$Var(\theta(N_{ab})) = \gamma + \delta log_2(N_{ab}) + Var(\theta_{sampling}) \tag{17}$$

For θ_{bio}, we assumed that both mean and variance decreased linearly with N_{ab} but that the mean and variance of $\theta_{sampling}$ were constant. We evaluated the six parameters in equations 12-13 (α, β, γ, δ, $\bar{\theta}_{sampling}$, $Var(\theta_{sampling})$) using a joint log-likelihood function. Our log-likelihood function is evaluated by splitting the serum antibody responses into two categories: seroconverted responses and non-seroconverted responses. We defined seroconverted individuals as those with post-exposure to pre-exposure titer ratios of at least four.

$$logL = logL_{\text{seroconverted}}\left(\alpha, \beta, \gamma, \delta, \bar{\theta}_{sampling}, Var\left(\theta_{sampling}\right)\right) + logL_{\text{nonseroconverted}}\left(\bar{\theta}_{sampling}, Var\left(\theta_{sampling}\right)\right) \tag{18}$$

$$logL_{\text{s=seroconverted}} \tag{18a}$$

$$= \sum_{N_{ab}}^{k_s} \left[-\frac{n_{j_s}}{2} (log(2\pi) + log(Var(\theta(N_{ab}))) \right.$$

$$\left. - \sum_{i}^{n_{N_{ab}}} \frac{1}{2Var(\theta(N_{ab}))^2} \left(x_i - \bar{\theta}(N_{ab})\right)^2 \right]$$

$$logL_{\text{n=nonseroconverted}} \tag{18b}$$

$$= \frac{n_{non}}{2} (log(2\pi) + log(Var(\theta_{sampling})))$$

$$- \sum_{i}^{n_{non}} \frac{1}{2Var(\theta_{sampling})} (x_i - \bar{\theta}_{sampling})^2$$

295

296 where k_s represents the binned N_{ab} categories reported in the data, n_{Nab} is the number of data
297 points in in the N_{ab} bin, and n_{non} is the total number of datapoints in the non-seroconverted
298 dataset. Our log-likelihood is the sum of two gaussian log-likelihoods, one for the seroconverted
299 data and one for the non-seroconverted data. Our log-likelihood function assumes that changes
300 in nonseroconverted individuals are due to sampling methodology while changes in
301 seroconverted individuals is due to a combination of both sampling methodology and real biology.

302 Once mle estimates for $\bar{\theta}(N_{ab})$ and $Var(\theta(N_{ab}))$ were obtained, we defined peak post-exposure
303 immunity as:

$$N_{ab_{peak}} = N_{ab} e^T \tag{19}$$

where T is a random value drawn from a normal distribution with mean $\bar{\theta}(N_{ab})$ and variance $Var\big(\theta(N_{ab})\big)$. Parameter estimates are in the **Supplemental Dataset S1.**

Initializing population-level immunity

Infants

We assumed that pre-mOPV2 challenge immunity in infants was defined by a bOPV vaccination regiment with either 1x or 2x IPV given at ages six, ten, and 14 weeks of age. Despite not containing live Sabin 2 poliovirus, bOPV does induce a small amount of heterotypic immunity against Sabin 2. The amount of heterotypic immunity can be inferred from the shedding duration of infants challenged with mOPV2. To simulate this, we devised a reinfection model where infants are administered a vaccine-equivalent dose of Sabin 1 poliovirus (10^6 infectious viruses) at six, ten, and 14 weeks of age. We assumed bOPV-induced immunity could be simulated as monotypic Sabin-1 immunity with a reduced probability of infection. At 18 weeks of age, infants were then challenged with mOPV2. We assumed infants were immunologically naiive prior to six weeks of age and that Sabin 1 shedding durations, shedding concentrations, immune boosting, and immune waning were identical to those of Sabin 2 (4). To simulate lower heterotypic immunity, we modified the oral susceptibility equation (equation 11) by introducing multiplicative modifier, ρ, that reduces infection probability:

$$P(infection) = P(infection|dose, S1, N_{ab})\ \rho \quad (20)$$

We estimated ρ using our reinfection model by calibrating it to the Sabin 2 shedding duration of infants challenged with mOPV2 using a PMC-ABC with 1000 particles and four iterations. Distance was defined as the squared difference between simulated and empirical shedding

prevalence collected weekly for five weeks post-mOPV2 challenge. Each iteration modified its acceptance threshold such that only the top 30, ten, and five percent of the distance distribution were accepted for the second, third, and fourth iteration.

Non-Infants

We assumed that pre-mOPV2 challenge in non-infants resulted from a complex immune history due to repeated vaccination or secondary transmission exposure from multiple vaccination campaigns and wild poliovirus. To simulate this, we devised a reinfection model where individuals are administered a vaccine-equivalent dose of Sabin 2 poliovirus (10^6 infectious viruses) at time intervals randomly drawn from a gamma distribution:

$$t_{interval} \sim gamma(shape(x = age), scale = 1) \quad (21)$$

$$shape(x = age) = \beta(1 - exp(-\alpha x)) + \gamma \quad (21a)$$

The shape of this gamma distribution ensures that the time interval between infection increases with age. We fit our simulation using a PMC-ABC with 1000 particles and four iterations to the previously fit equation of immunity vs age in the household contact population of Matlab, Bangladesh (5):

$$N_{ab,x=age} = Nab\big(1 + (12x - 30)\big)^{-0.24} \quad (22)$$

Distance was defined as the squared difference between simulated immunity values and immunity values estimated using equation 22 for all integer ages between five and 100. Each iteration modified its acceptance threshold such that only the top 30, ten, and five percent of the distance distribution were accepted for the second, third, and fourth iteration. The mean posterior estimates of each parameter were used as our point-estimates. Immunity was consistent with an

age-dependent exposure rate, with children inferred to have been re-exposed more frequently than adults.

Supplemental – Calibrating Transmission

We calibrated two different transmission models: a single parameter (β_{ma}) mass action transmission model and a four parameter (β_{hh}, β_{bari}, $\beta_{village}$, $\beta_{intervillage}$) multiscale transmission model. β represents the number of contacts per shedding individual per timestep. We first obtained priors for village and intervillage transmission based on Sabin 2 shedding during the enrollment period across all routine immunization trial arms (villages receiving tOPV and bOPV). With these priors, we then calibrated our model to the shedding prevalences in bOPV villages. Shedding was observed in household cohorts defined by household/bari membership and mOPV2 challenge status.

Identifying priors for village and intervillage transmission counts

Intervillage

During enrollment, a small number of subjects in bOPV2 routine immunization villages were positive for Sabin 2 due to transmission from villages assigned to tOPV routine immunization or the community outside Matlab. While exposure could come from anywhere in Bangladesh (and beyond), we by assumed all Sabin 2 exposure in the bOPV villages was due to transmission originating from villages assigned with tOPV routine immunization. This allowed us to derive an upper-bound for $\beta_{intervillage}$.

First, we calculated the intervillage transmission rate between tOPV and bOPV villages during the enrollment period. We define the intervillage transmission rate as the number of observed Sabin 2 transmission events per observed susceptible subjects in bOPV villages per number of tOPV vaccinations in tOPV villages.

$$\lambda_{inter_k} = \frac{n_{shedding_k}}{n_{observed_k}/n_{tOPV}} \quad (23)$$

where k refers the type of individual (infant or noninfants), $n_{shedding_k}$ the number of individuals of type k shedding Sabin 2, $n_{shedding_k}$ the number of individuals of type k observed throughout the enrollment period, and n_{tOPV} is the total number of tOPV vaccinations administered.

For infants, we followed 625 infants in the bOPV villages during the enrollment period, of which six shed Sabin 2. Similarly, we followed 1137 noninfants (the household contacts of enrolled individuals), of which one shed Sabin 2.

$$\lambda_{inter_{infant}} = \frac{6}{625/600} = 1.6 * 10^{-5} \quad (24)$$

$$\lambda_{inter_{noninfant}} = \frac{1}{1137/600} = 1.5 * 10^{-6}$$

The estimated ten-fold lower rate to noninfants is consistent with the differences in immunity between infants who did not receive live Sabin 2 vaccination verses older individuals who have.

Using these rates, we then estimated intervillage transmission events after the onset of the mOPV2 campaign. For simplicity, we assumed that mOPV2 challenge provides an equivalent source of virus as tOPV vaccination in unimmunized infants (5). We also assume that all non-infant people in the population have equivalent intervillage exposure as household contacts. The expected number of intervillage transmissions is estimated as

$$N_{inter} = \eta \left[\lambda_{inter_{infant}} * n_{infants} + \lambda_{inter_{noninfant}} * n_{noninfant} \right] \quad (25)$$

$$\eta = n_{infants_{mOPV2}} + \frac{n_{noninfant_{mOPV2}}}{15}$$

Where N_{inter} is the number of intervillage transmission events in bOPV villages post-mOPV2 challenge, η is the number of infant-equivalent mOPV2 recipients, $n_{infants_{mOPV2}}$ is the number of infants challenged with mOPV2, $n_{noninfant_{mOPV2}}$ is the number of household contacts challenged with mOPV2, and $n_{infant}, n_{noninfant}$ are the number of susceptible (not challenged with mOPV2) infants and noninfants. Because household contacts were older and received tOPV as routine immunization prior to our study, they were observed to shed 15x less virus after mOPV2 challenge (5) . η normalizes the difference in shedding following mOPV2 challenge in infants who received bOPV2 and noninfants who received tOPV2 during routine immunization.

In the bOPV villages, 199 infants and 2822 noninfants were challenged with mOPV2. The total number of susceptible infants and household contacts was ~1200 and ~80000.

$$\eta = 199 + \frac{2822}{15} \approx 387$$

$$N_{inter} = 387 \left[\lambda_{inter_{infant}} * 1200 + \lambda_{inter_{hhc}} * 80000 \right] \approx 54 \quad (26)$$

Thus, to constrain intervillage transmission in our multiscale transmission model, we assume a prior for $\beta_{intervillage}$ using a normal distribution with mean N_{iv} and variance 10*N_{iv}. The ten serves to inflate variance because N_{iv} was only crudely estimated. The log-likelihood component for intervillage transmission was defined as

$$logL_{inter} = -\frac{\left(N_{inter_{simulated}} - N_{inter}\right)^2}{2 * (10 * N_{inter})}$$

398

399 *Within-Village*

400 To constrain the within-village transmission parameter, we examined enrollment data from the
401 tOPV villages. For subjects in the tOPV villages, exposure prior to the first dose of routine
402 immunization at six weeks of age (Table S1 of (5)) is most likely due transmission from older
403 infants in the village shedding Sabin 2 following tOPV routine immunization. Following this logic,
404 we observed one infant and household contact infection due to within-village transmission (Table
405 S5 of (5)) among 294 infants and 547 household contacts. As above, approximately 600 children
406 received tOPV in routine immunization, spread out across 22 villages. Thus, the average number
407 of tOPV vaccine recipients in each village was 600/22 ≈27. However, using the average tOPV
408 vaccine recipient count would overestimate the expected within-village transmission rate, due to
409 the heavy skew in village sizes.

410 The village-size weighted transmission rate was estimated as a weighted sum across all villages

$$\lambda_{village_k} = \sum_{i}^{n\_villages_{tOPV}} \frac{n_{shedding_{i,k}}}{n_{observed_{k,i}}/n_{tOPV_i}} \left(\frac{n_{individuals_i}}{n_{individuals_{total}}}\right) \quad (27)$$

411

412 Which yielded:

$$\lambda_{village_{infant}} = 1 * 10^{-4} \quad (28)$$

$$\lambda_{village_{noninfant}} = 7.5 * 10^{-5} \quad (29)$$

Similarly, the weighted total number of within village events was defined as

$$N_{village} = \eta \left[\lambda_{village_{infant}} * n_{infants} + \lambda_{village_{noninfant}} * n_{noninfant} \right] * \sum_{i}^{n\_villages_{tOPV}} \left(\frac{n_{individuals_i}}{n_{individuals_{total}}} \right) \quad (30)$$

$$\eta = n_{infants_{mOPV2}} + \frac{n_{noninfant_{mOPV2}}}{15}$$

We found that $N_{village} = 140$. As with intervillage transmission, we assumed a prior for within village transmission events as a normal distribution with mean $N_{village} = 140$ and variance $10 * N_{village}$. The log-likelihood component for within village transmission was defined as:

$$logL_{village} = -\frac{\left(N_{village_{simulated}} - N_{village}\right)^2}{2 * (10 * N_{village})}$$

Shedding proportions in household cohorts

Stool samples were collected from enrolled study participants 0-10, 14, 18, and 22 weeks post-mOPV2 challenge. Shed prevalence was calculated as the proportion of shedding individuals per household cohort. We assumed that shed prevalence followed a binomial distribution at that timepoints were independent. Under these assumptions, our combined likelihood function was defined as

$$logL = \sum_{k}^{n_{traces}} \frac{\left[\sum_{i}^{hh_{cohorts}} \left[p_{i,k}^{n_{shedding_{i,k}}} * \left(1 - p_{i,k}\right)^{\left(n_{total_{i,k}} - n_{shedding_{i,k}}\right)}\right] + logL_{village,k} + logL_{inter,k}\right]}{n_{traces}} \quad (31)$$

Where k is the simulation trace, i refers to the ith household cohort, p is the proportion of shedding in Matlab, $n_{shedding,i}$ is the number of shedding individuals in the ith household cohort from our simulation, and $n_{total,i}$ the total number of simulated individuals for the ith household cohort. Due to the stochastic nature of our model, we calculated the log-likelihood for each of the 30 simulation traces run and used the average log-likelihood (12).

For our multiscale model, we first sampled 600 parameter combinations from a four-dimensional latin hyperspace cube where each dimension corresponded to one component of β used in the multiscale model. We then evaluated the profile likelihoods for each dimension (13). For the multiscale model, the profile likelihood surfaces for β_{hh} and β_{bari} were intractable; low values for either of these parameters could be compensated by increasing transmission at any of the other levels. Because it is unlikely that transmission does not occur at either the household or *bari*, to obtain point estimates for each of these parameters, we:

1. Resampled a total of 1800 points from a 4-dimensional latin hyperspace cube. The dimension associated with β_{hh} was replaced with one examining fecal-oral dose concentrations and assumed β_{hh} = 1. This forces the model to contact one household member in each timestep. Three fecal-oral doses were examined: 1e-6g/contact, 2.5e-6g/contact, and 5e-6g/contact. Of these, 2.5e-6g/contact was the optimal choice. The range of values explored for β_{bari}, $\beta_{village}$, and $\beta_{intervillage}$ranged from [1,30], [1,9], and [1,4], respectively.
2. We then identified the maximum profile likelihood point-estimates for $\beta_{village}$ and $\beta_{intervillage}$ from the subset of points where the fecal-oral dose was 2.5e-6g/contact.

3. Finally, we examined the subset of parameter combinations where the fecal-oral dose was 2.5e-6g/contact, $\beta_{village}$ and $\beta_{intervillage}$ were equal to the maximum profile likelihood point-estimates identified in step 2 to obtain our maximum profile likelihood point-estimate for β_{bari}.

The final maximum profile likelihood point-estimates for our multiscale model was: fecal-oral dose = 2.5e-6g/contact, β_{hh}=1, β_{bari}=15, $\beta_{village}$=4, and $\beta_{intervillage}$=2.

For our mass action model, we sampled β_{ma} from a set of consecutive integer values ranging from 0-50 and evaluated the average log-likelihood at each point. To make it equivalent with the multiscale model, we assumed fecal-oral dose was 2.5e-6g/contact. We evaluated β_{ma} using two different log-likelihood functions, one with the village and inter-village priors (above equation), and one without:

$$logL = \sum_{k}^{n_{traces}} \frac{\sum_{i}^{hh_{cohorts}} \left[p_{i,k}^{n_{shedding_{i,k}}} * \left(1 - p_{i,k}\right)^{\left(n_{total_{i,k}} - n_{shedding_{i,k}}\right)} \right]}{n_{traces}} \quad (32)$$

With the village and inter-village priors, our point estimate for β_{global} = 1. Without the village and inter-village priors, our point-estimate for β_{global}=19.

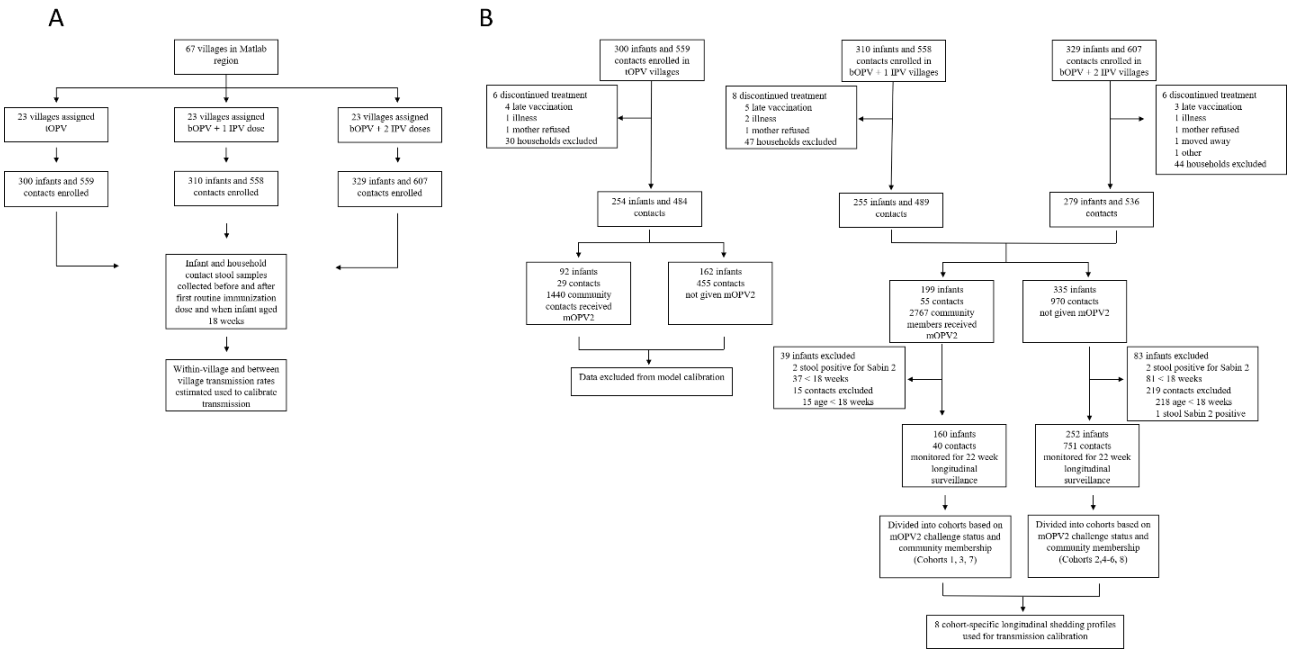

Fig. S1. mOPV2 clinical trial design. A) Routine immunization and enrollment phase and B) the mOPV2 campaign and the 22week longitudinal surveillance period. The post-mOPV2 campaign surveillance data from tOPV villages was excluded due to the larger potential of unmodeled secondary vaccine transmission from routine immunization.

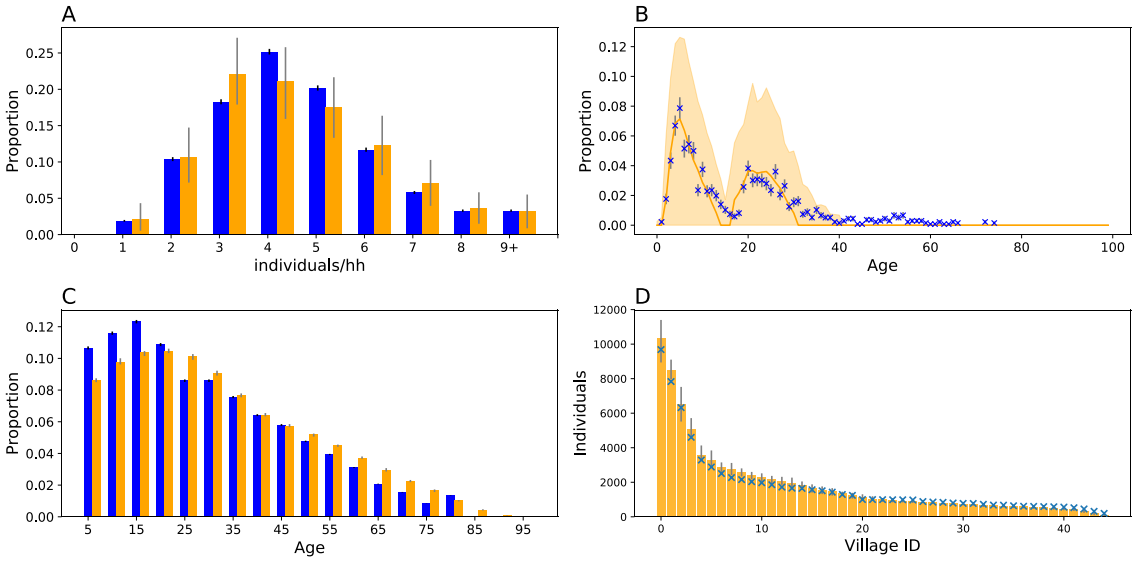

Fig. S2. Household structure. Household demography simulations (*orange*) compared against demography data from the 2014 BDHSS (blue). Error bars for simulation output represent middle 95 percentile as estimated from 20 random iterations. Error bars for the BDHSS or clinical trial data represent the 95% binomial confidence interval around the mean. A) Household size distribution. B) Housed contact age distributions. C) Age pyramid D) Village size. For D, only simulated confidence intervals are shown.

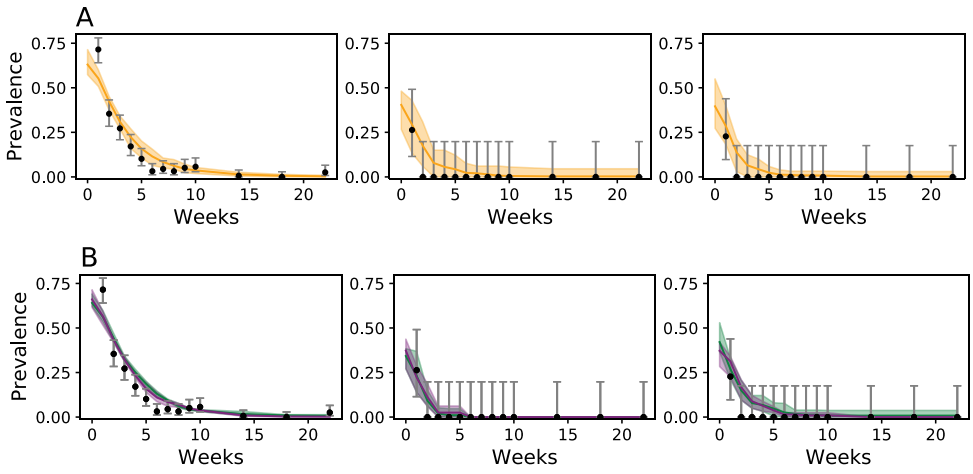

Fig. S3. Shedding profile in primary vaccine recipients. *Orange*: multiscale, *Green*: fully calibration mass action, and *Purple*: partially calibrated mass action. Cohort-specific longitudinal shedding profiles in mOPV2 vaccination recipients (left to right: cohorts one, four, seven) compared to the multiscale model (**A**) and two mass action models (**B,** green = fully-calibrated mass action, purple = partially-calibrated mass action). The solid lines are the simulated average and the shading indicates two standard deviations from the mean. Error bars for the data points represent two binomial standard errors of the mean.

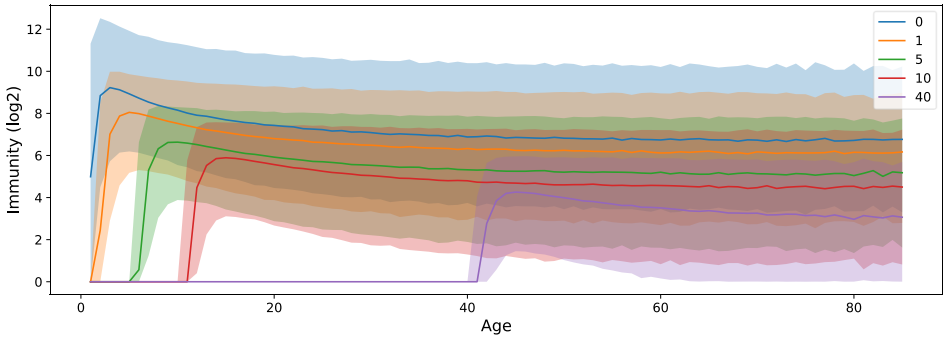

Fig. S4. Population Immunity Following Vaccination Cessation. Average immunity (log2 antibody titers) against Sabin 2 in our populations immediately after (zero years), one, five, ten, and 40 years post-vaccination cessation. Solid line indicates the population average and the shading the boundaries of the middle 95th percentile. Note the age-structured erosion of population immunity due to new births and immune waning

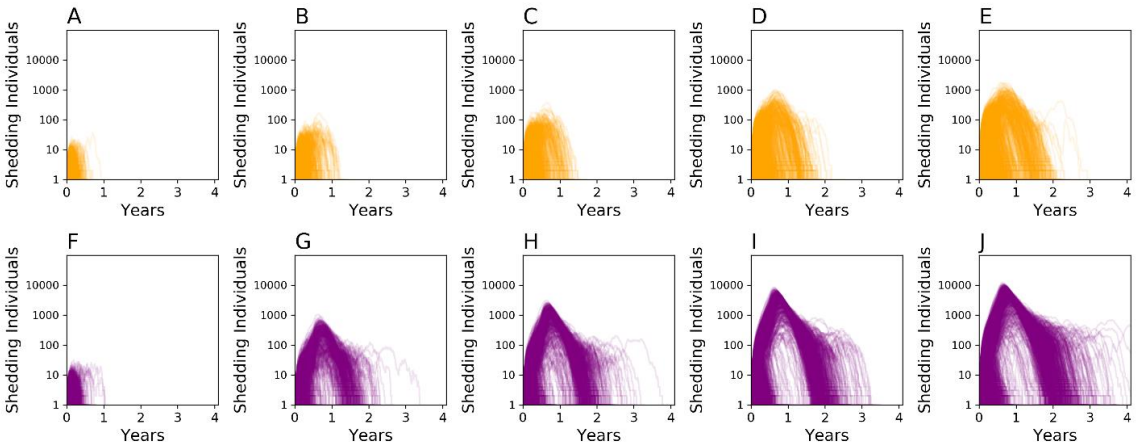

Fig. S5. Sabin 2 point importation spaghetti plots. *Top row*: Multiscale model *Bottom row*: partially-calibrated mass action model. **A/F** immediately post-vaccination cessation, **B/G** one year post-vaccination cessation, **C/H** five years post-vaccination, **D/I** ten years post-vaccination, and **E/J** 20 years post-vaccination cessation.

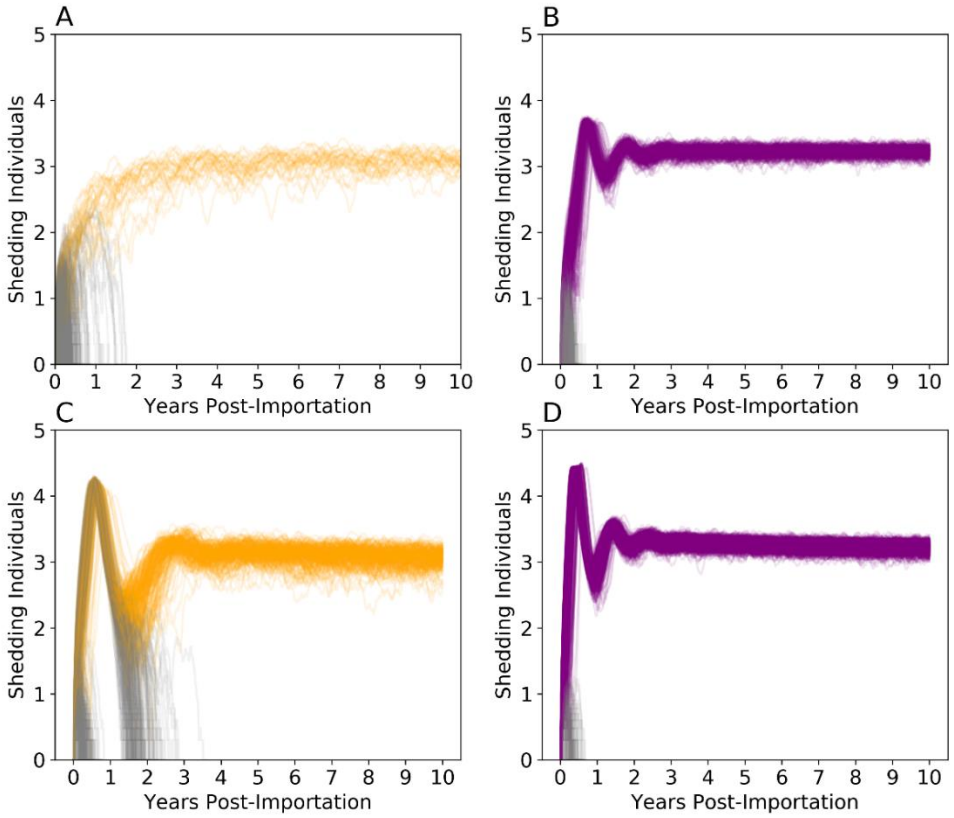

Fig. S6. WPV point importation. Simulation traces immediately after vaccination cessation using the **A**) multiscale and **B**) partially-calibrated mass action model. Simulation traces ten years post-vaccination cessation using the **C**) multiscale and **B**) partially-calibrated mass action model. Colored traces indicate those that have shedding individuals after ten years. Grey traces indicate those simulations that fadeout and cease transmission.

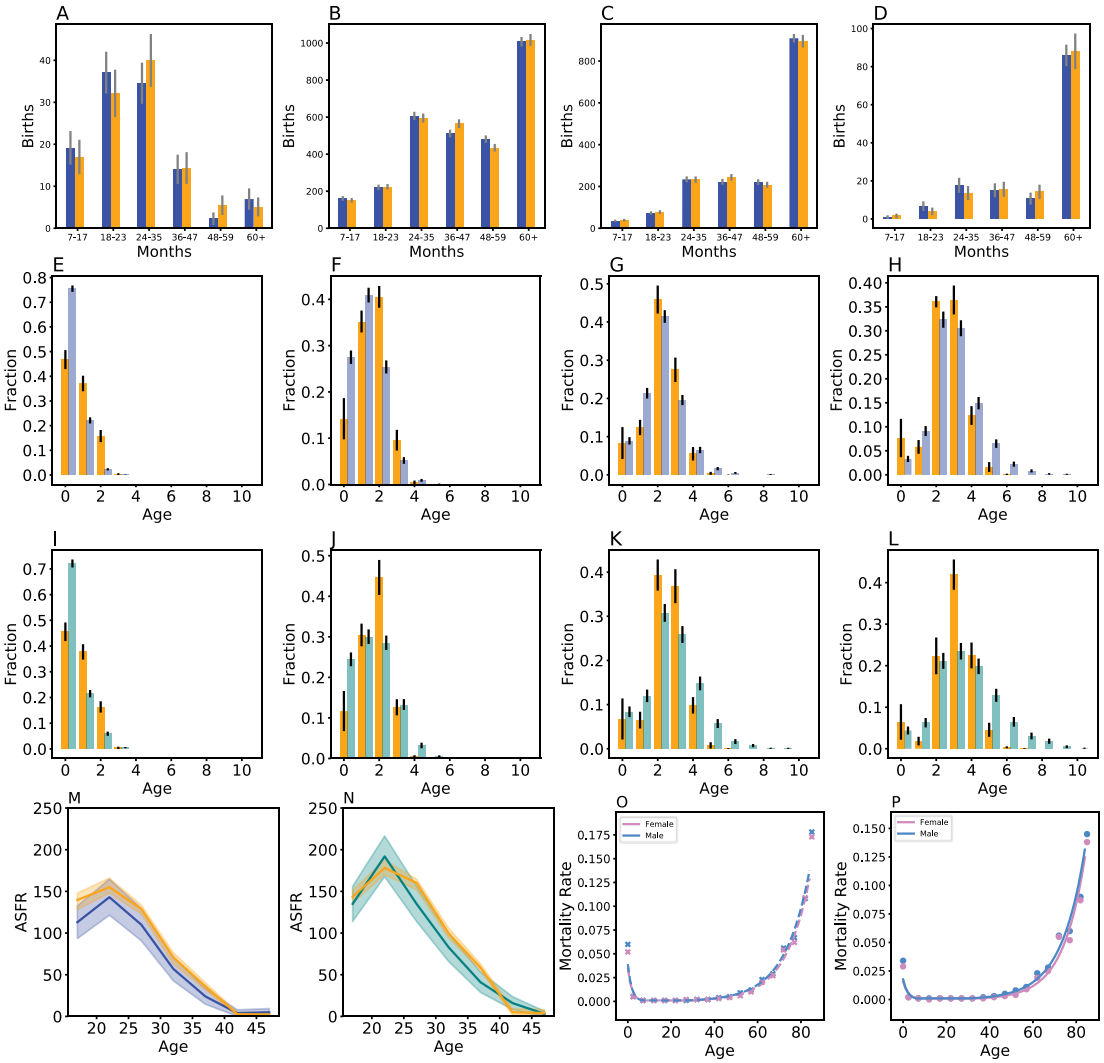

Fig. S7. Fertility and mortality fits in the demographic model. Simulated (*orange*) birth interval periods for individuals aged **A**) 15-19, **B**) 20-29, **C**) 30-39, and **D**) 40-49 compared against the 2014 Bangladesh Demographic Health Survey data (*blue*). Simulated (*orange*) number of children per married female for individuals aged **E/I**) 15-19, **F/J**) 20-24, **G/K**) 25-29, and **H/L**) 30-34 compared against the 2014 (**E-H**) and 2004 (**I-L**) Bangladesh demographic health surveys. Error bars in all bar plots indicate one binomial standard error from the mean. Simulated (*orange*) age specific fertility rates (ASFR, number of births per 1000 individuals) for 2014 (**M**) and 2004 (**N**). The solid line is the average and the shading two standard errors from the mean. Sex (male = blue, female = pink) and age specific mortality rates for 2014 (**O**) and 2004 (**P**). Data are represented by dots.

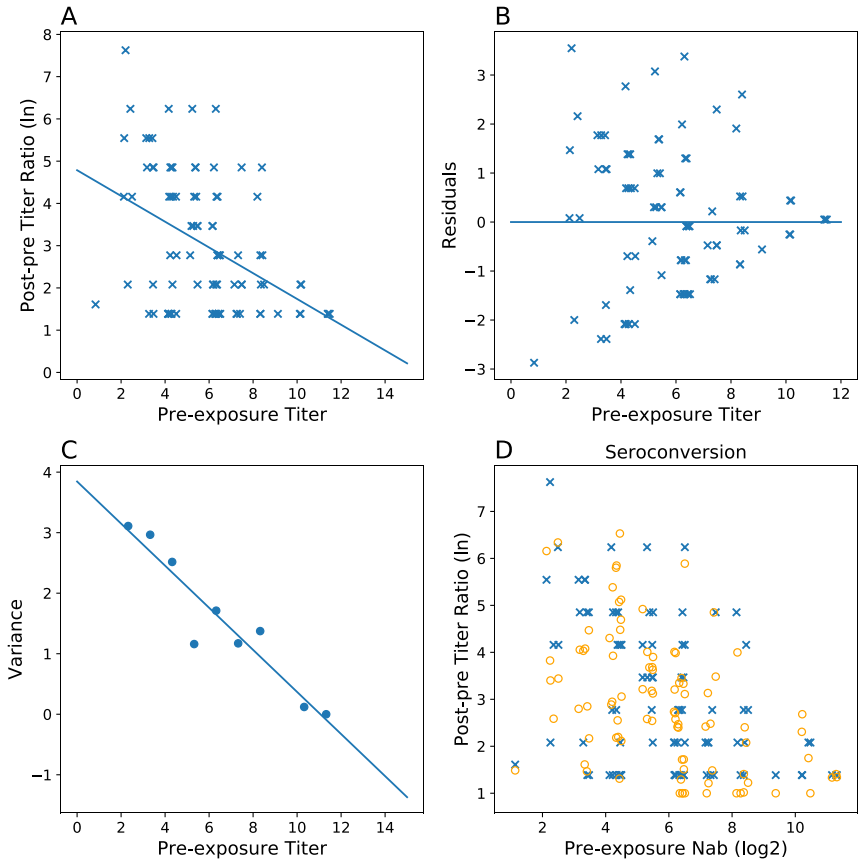

Fig. S8. Immune boosting. **A**) Ordinary least squares fit relating pre-exposure log2 antibody titers to the ratio of post-exposure to pre-exposure antibody titer observed in 1953 Louisiana.(11) Overlapping points are randomly jittered to better represent point density. Line represents the ordinary least squares fitted equation. **B**) Residual plot of our ordinary least squares fit. The fan-shaped distribution is classic signature of heteroskedasticity. **C**) Fitted ordinary least square function relating variance with pre-exposure titer. **D**) Final immune boosting model predictions (*orange*) compared against the original data (*blue*).

Table S1. **Cohort definitions.** Cohorts 1-3 are infants while cohorts 4-8 are household contacts. The “challenged with mOPV2 “column indicates whether the individuals in this cohort received Sabin 2 vaccine (+ for yes, - for no) during the mOPV2 campaign. For cohorts 4-8, the “Infant

538 status" column indicates whether the infant of the household contact received mOPV2. The "*Bari*
539 status" column indicates whether any member in the *bari* received mOPV2. The "Infection
540 Source" column indicates the type of transmission each cohort is most sensitive to. Individuals in
541 cohorts one, four, and seven received mOPV2 and shedding in these cohorts largely reflect
542 individual infection dynamics.

Cohort Number	Individual Type	Challenged with mOPV2	Infant Status	*Bari* Status	Infection Source
1	infant	+		+	Vaccination
2	infant	-		+	Household/Bari
3	infant	-		-	Village/Region
4	Household contact	+	+	+	Vaccination
5	Household contact	-	-	+	Household/Bari
6	Household contact	-	+	+	Household/Bari
7	Household contact	+	-	+	Vaccination
8	Household contact	-	-	-	Village/Region

543

544

Dataset S1 (separate file). Parameter Table

SI References

1. Skinner GW (1994) Family Systems and Demographic Processes. *Anthropological Demography: Toward a New Synthesis*, eds Kertzer DI, Fricke T (University of Chicago Press).

2. Bangladesh Demographic and Health Survey 2014 (Dhaka, Bangladesh, and Rockville, Maryland, USA).

3. Amin S (1998) Family Structure and Change in Rural Bangladesh. *Popul Stud (NY)* 52(2):201–213.

4. Famulare M, Selinger C, McCarthy KA, Eckhoff PA, Chabot-Couture G (2018) Assessing the stability of polio eradication after the withdrawal of oral polio vaccine. *PLOS Biol* 16(4):e2002468.

5. Taniuchi M, et al. (2017) Community transmission of type 2 poliovirus after cessation of trivalent oral polio vaccine in Bangladesh: an open-label cluster-randomised trial and modelling study. *Lancet Infect Dis* 17(10):1069–1079.

6. Behrend MR, Hu H, Nigmatulina KR, Eckhoff P (2014) A quantitative survey of the literature on poliovirus infection and immunity. *Int J Infect Dis* 18:4–13.

7. Alexander JPJ, Gary HEJ, Pallansch MA (1997) Duration of poliovirus excretion and its implications for acute flaccid paralysis surveillance: a review of the literature. *J Infect Dis* 175 Suppl:S176-82.

8. O'Ryan M, et al. (2015) Inactivated poliovirus vaccine given alone or in a sequential schedule with bivalent oral poliovirus vaccine in Chilean infants: a randomised, controlled, open-label, phase 4, non-inferiority study. *Lancet Infect Dis* 15(11):1273–1282.

9. Asturias EJ, et al. (2016) Humoral and intestinal immunity induced by new schedules of bivalent oral poliovirus vaccine and one or two doses of inactivated poliovirus vaccine in Latin American infants: an open-label randomised controlled trial. *Lancet (London, England)* 388(10040):158–169.

10. Jafari H, et al. (2014) Efficacy of inactivated poliovirus vaccine in India. *Science (80-)* 345(6199):922 LP – 925.

11. GELFAND HM, LEBLANC DR, FOX JP, CONWELL DP (1957) Studies on the development of natural immunity to poliomyelitis in Louisiana. II. Description and analysis of episodes of infection observed in study group households. *Am J Hyg* 65(3):367–385.

12. He D, Ionides EL, King AA (2010) Plug-and-play inference for disease dynamics: measles in large and small populations as a case study. *J R Soc Interface* 7(43):271–283.

13. L. IE, C. B, J. P, A. SR, A. KA (2017) Monte Carlo profile confidence intervals for dynamic systems. *J R Soc Interface* 14(132):20170126.