Supplementary material

Expanding vaccine efficacy estimation with dynamic models fitted to cross-sectional prevalence data post-licensure

Erida Gjini[a], M. Gabriela M. Gomes[b,c,d]

[a]*Instituto Gulbenkian de Ciência, Apartado 14, 2781-901 Oeiras, Portugal.*
[b]*CIBIO-InBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Universidade de Porto, Portugal.*
[c]*Instituto de Matemática e Estatística, Universidade de São Paulo, Brazil.*
[d]*Liverpool School of Tropical Medicine, Liverpool, United Kingdom*

Text S1. Coexistence endemic equilibria in the competition models prior to vaccination

- **SI-2 model**
  The equilibria for this model are calculated by setting all equations (§2.2) in the main text equal to zero, with $\rho = 0$. Besides the type 1 only and type 2 only endemic equilibria, the SI-2 model admits a coexistence equilibrium given by:

$$\begin{aligned} S^* &= \frac{1}{R_0} \\ I_1^* &= \frac{1}{R_0}\Big(\frac{1}{\sigma_1} - \frac{2}{(R_0-1)\sigma_1\sigma_2 + \sigma_1 + \sigma_2}\Big) \\ I_2^* &= \frac{1}{R_0}\Big(\frac{1}{\sigma_2} - \frac{2}{(R_0-1)\sigma_1\sigma_2 + \sigma_1 + \sigma_2}\Big) \\ I_{12}^* &= 1 - (S^* + I_1^* + I_2^*) \end{aligned} \tag{1}$$

  where $R_0 = \beta/\mu$ in the SI-2 framework. Stability of this pre-vaccine equilibrium requires:

$$\sigma_1 > \frac{\sigma_2}{(R_0-1)\sigma_2 + 1}; \quad \sigma_2 > \frac{\sigma_1}{(R_0-1)\sigma_1 + 1}.$$

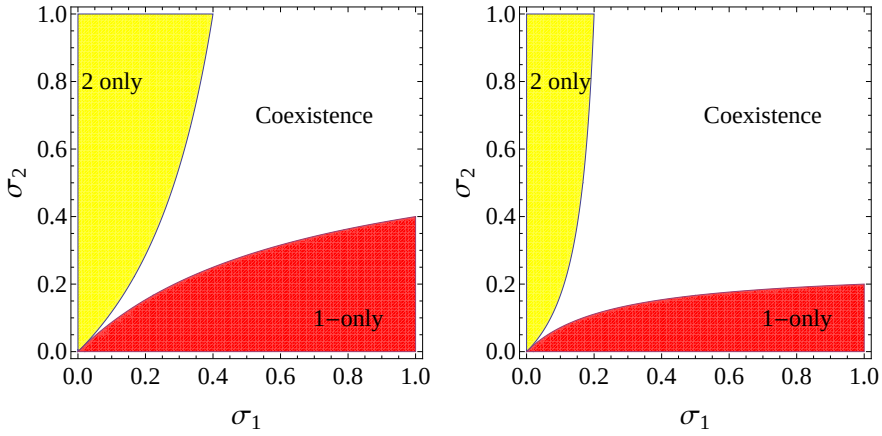
  For an illustration of these criteria, depending on σ_1 and σ_2 see Figure S1.

- **SIS-2 model**
  We find the equilibria pre-vaccine by setting equations (§2.3) in the main text equal to zero (assuming $\rho = 0$), and then simplify the expressions using $R_0 = \beta/(\gamma + \mu)$. The endemic coexistence equilibrium is given by:

$$\begin{aligned} S^* &= \frac{1}{R_0} \\ I_1^* &= I_2 = \frac{R_0 - 1}{R_0[2 + (\kappa\sigma + \sigma)(R_0 - 1)]} \\ I_{11}^* &= I_{22}^* = \frac{\kappa\sigma(R_0-1)^2}{2R_0[2 + (\kappa\sigma + \sigma)(R_0 - 1)]} \\ I_{12}^* &= 1 - (S^* + I_1^* + I_2^* + I_{11}^* + I_{22}^*), \end{aligned} \tag{2}$$

  where the indices specify carriage: whether single or double carrier of pathogen types of group 1, 2 or both. Asymptotic analysis of the system reveals that stability of this equilibrium requires $\kappa < 1$.

Supplementary figures



(a) Low transmission intensity (b) Higher transmission intensity

Figure S1: **Stability criteria on the competition coefficients σ_1 and σ_2 in the SI-2 model that guarantee stable coexistence between the two pathogen types at the endemic equilibrium.** The regions are computed by checking the eigenvalues of the Jacobian matrix evaluated at the endemic equilibria. In our vaccination models, we consider parameter regimes of stable coexistence prior to vaccine introduction (white region), where the vaccine targets type 1. Parameter values: a) $\beta = 0.05, \mu = 0.02$ ($R_0 = 2.5$); b) $\beta = 0.1, \mu = 0.02$ ($R_0 = 5$).

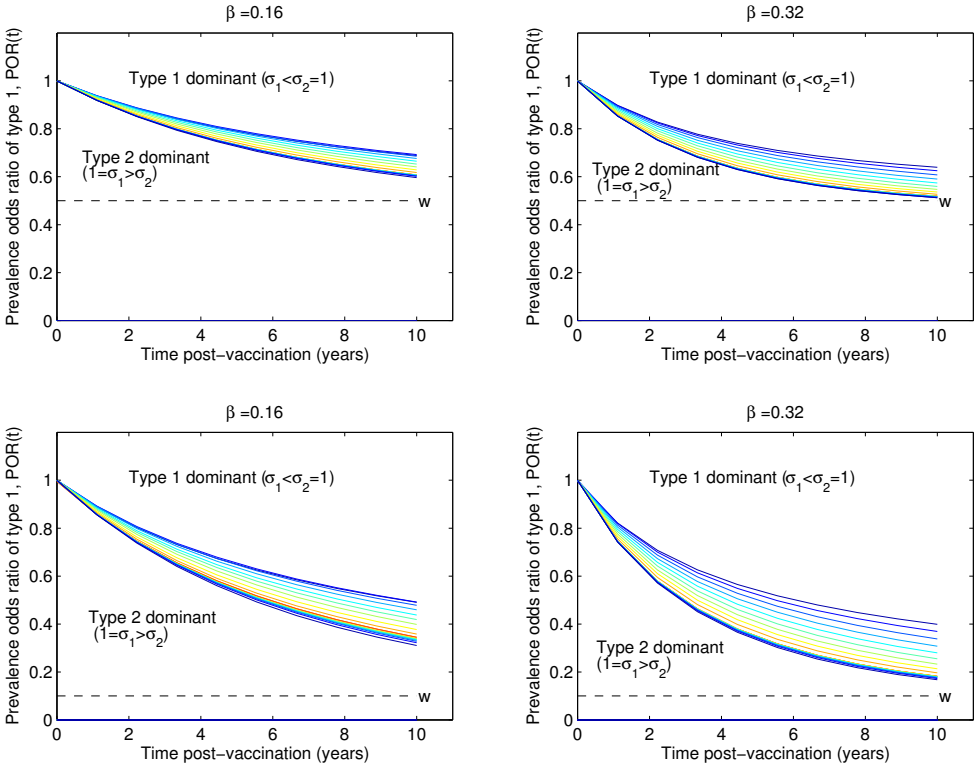


Figure S2: **Prevalence-odds-ratio (POR) of target type in vaccinated and non-vaccinated hosts vs. true relative risk in the SI-2 model.** Type 1 can be dominant prior to vaccination ($0.2 \leq \sigma_1/\sigma_2 \leq 1$), or alternatively, type 2 can be dominant ($0.2 \leq \sigma_2/\sigma_1 \leq 1$). The coloured lines from blue to red correspond to increasing values of the competition ratio, σ_1/σ_2 and σ_2/σ_1 from 0.2 to 1. Other parameter values: $\mu = 0.0167, \rho = 0.5$. Initial conditions at endemic equilibrium. Time is in units of years. The low transmission cases ($\beta = 0.032$), correspond to $R_0 = 1.9$. While the high transmission cases ($\beta = 0.32$) correspond to $R_0 = 19$. POR(t) is closer to w than $PR(t)$.

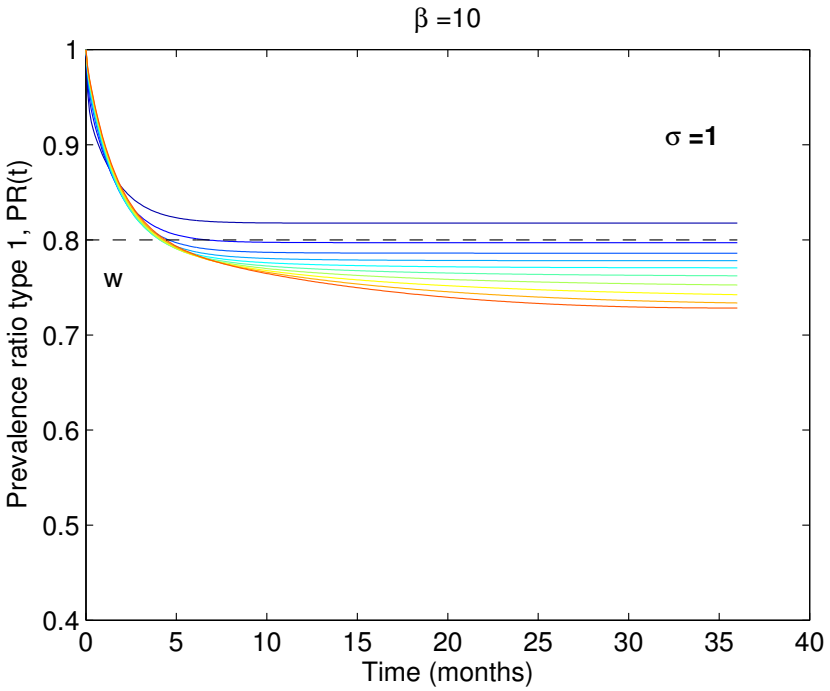


Figure S3: **Prevalence ratio can under-estimate relative risk for high β (SIS-2).** Parameter values: $\sigma = 1, \gamma = 0.6, \mu = 0.02, \rho = 0.5$ and $\beta = 10$, κ is varied between 0.1 and 0.9 (colored lines from blue to red) to reflect different scenarios of within-group competition. The true risk ratio is $w = 0.8$, which implies $VE = 0.2$. The worst discrepancy between PR and w is observed for κ close to 1, which leads to a negative deviation, indicating an over-estimation of VE if PR were to be used for this purpose.

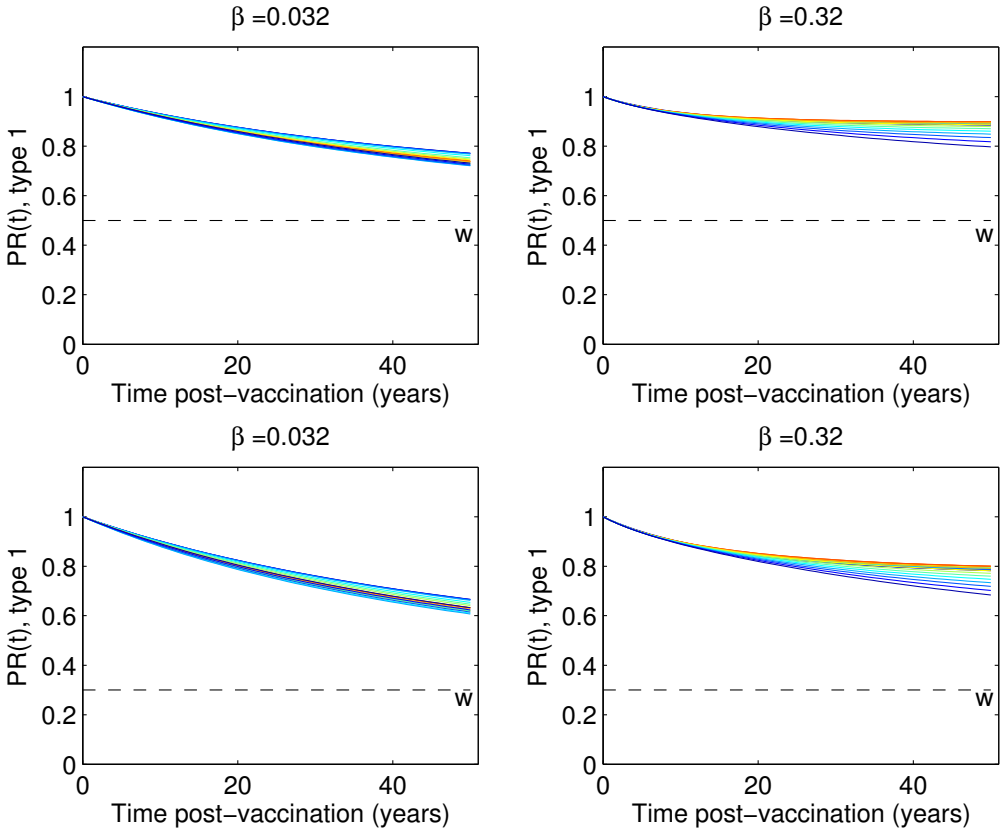


Figure S4: **Prevalence ratio in the SI-2 model (counting for mixed carriage I_{12}).** Type 1 can be dominant prior to vaccination ($0.2 \leq \sigma_1/\sigma_2 \leq 1$), or alternatively, type 2 can be dominant ($0.2 \leq \sigma_2/\sigma_1 \leq 1$). The coloured lines from blue to red correspond to increasing values of the competition ratio, σ_1/σ_2 and σ_2/σ_1 from 0.2 to 1. Parameters as in Figure 3 in the main text. The sensitivity to competition asymmetries decreases but the discrepancy with w increases overall.

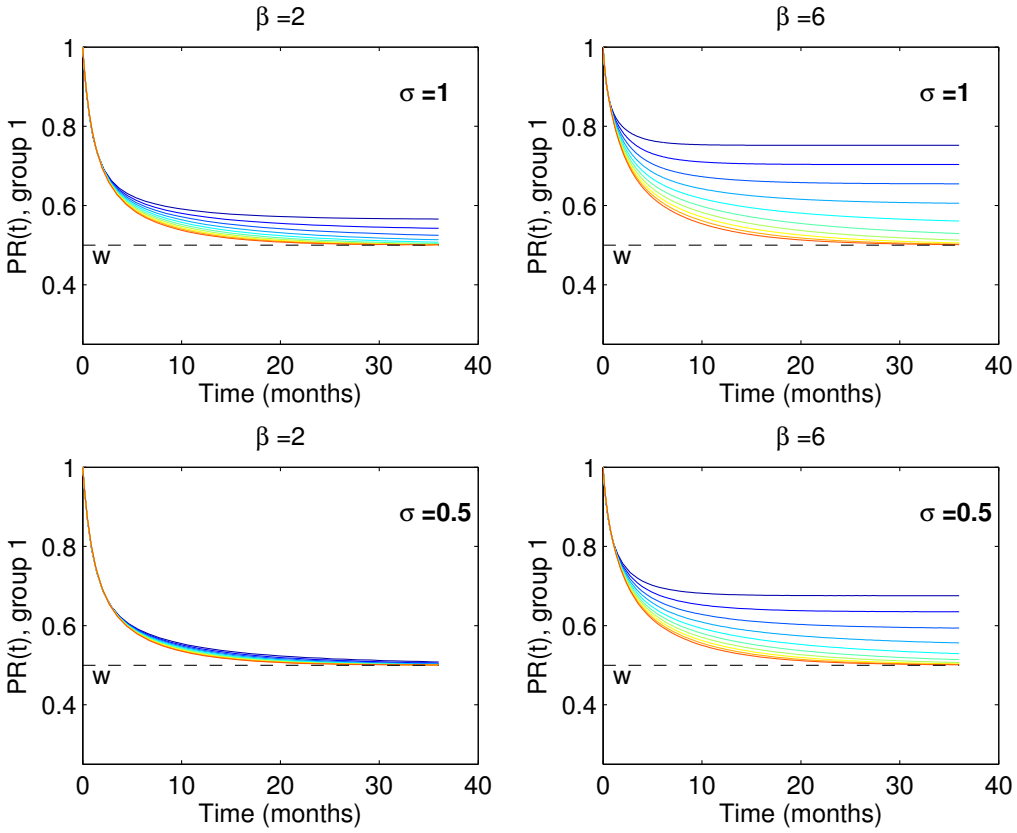


Figure S5: **Prevalence ratio in the SIS-2 model (counting for mixed carriage I_{12}).** Analogous to figure 4 in main text. We can see the sensitivity of $PR(t)$ to competition hierarchies (within/between group, represented by κ), reflected in the blue to red lines, increases, especially for large transmission intensity. This indicates that mixed multiple carriage holds important information about indirect vaccine effects, but may be unsuitable to include in the analyses aimed at extracting true relative risk.

Text S2. The SIR-2 model

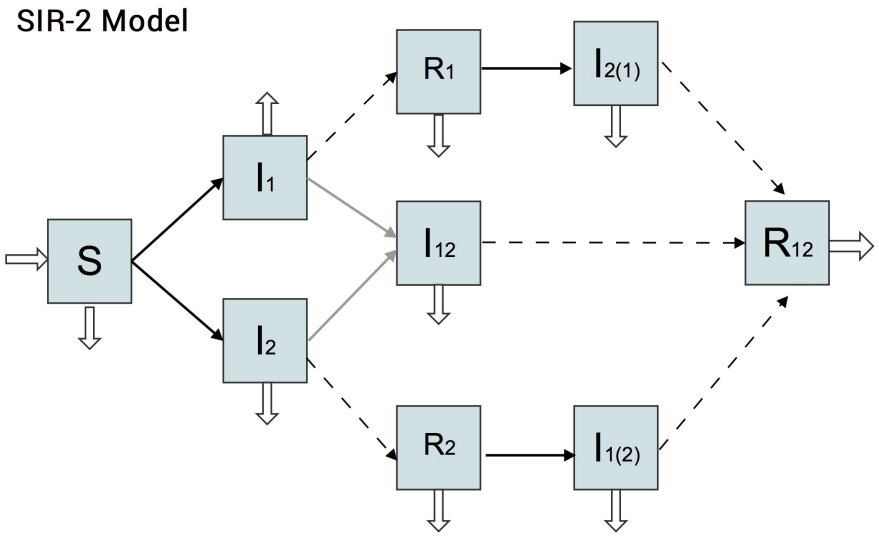
Model structure and assumptions

In this model, we change the SI-2 model, assuming hosts recover with life-long type-specific immunity, at rate γ. They enter these recovery classes: R_1, refers to those that are immune against type 1 but can acquire type 2; R_2, refers to those that are immune against type 2 but can acquire type 1, and R_{12}, describes hosts immune to both circulating types. Singly infected hosts with type 1, and immune to type 2 are denoted by $I_{1(2)}$, and viceversa. Hosts infected with type 2, but immune against type 1, are denoted by $I_{2(1)}$. Both these types of hosts contribute now to the forces of infection λ_1 and λ_2. Vaccination (status denoted by subscript 0/1) acts as before, against type 1, with partial protection, given by the factor $w(0 \leq w \leq 1)$. After hosts experience multiple infection with both types, or sequential infection by two types, they recover with full immunity to both types R_{12}. Clearance of single and double carriage is assumed to occur at equal rates, as in the earlier models. We assume no cross-immunity.

$$
\begin{cases}
\text{Non-vaccinated hosts} \\
\frac{dS^0}{dt} = \mu(1-\rho) - (\lambda_1+\lambda_2)S^0 - \mu S^0 \\
\frac{dI_1^0}{dt} = \lambda_1 S^0 - I_1^0\sigma_1\lambda_2 - (\mu+\gamma)I_1^0 \\
\frac{dI_2^0}{dt} = \lambda_2 S^0 - I_2^0\sigma_2\lambda_1 - (\mu+\gamma)I_2^0 \\
\frac{dI_{12}^0}{dt} = \sigma_1\lambda_2 I_1^0 + \sigma_2\lambda_1 I_2^0 - (\mu+\gamma)I_{12}^0 \\
\frac{dR_1^0}{dt} = \gamma I_1^0 - \lambda_2 R_1^0 - \mu R_1^0 \\
\frac{dR_2^0}{dt} = \gamma I_2^0 - \lambda_1 R_2^0 - \mu R_2^0 \\
\frac{dR_{12}^0}{dt} = \gamma(I_{12}^0 + I_{2(1)}^0 + I_{1(2)}^0) - \mu R_{12}^0 \\
\frac{dI_{2(1)}^0}{dt} = \lambda_2 R_1^0 - (\mu+\gamma)I_{2(1)}^0 \\
\frac{dI_{1(2)}^0}{dt} = \lambda_1 R_2^0 - (\mu+\gamma)I_{1(2)}^0
\end{cases}
\qquad
\begin{cases}
\text{Vaccinated hosts} \\
\frac{dS^0}{dt} = \mu\rho - (w\lambda_1+\lambda_2)S^1 - \mu S^1 \\
\frac{dI_1^1}{dt} = w\lambda_1 S^1 - I_1^1\sigma_1\lambda_2 - (\mu+\gamma)I_1^1 \\
\frac{dI_2^1}{dt} = \lambda_2 S^1 - I_2^1\sigma_2\lambda_1 - (\mu+\gamma)I_2^1 \\
\frac{dI_{12}^1}{dt} = \sigma_1\lambda_2 I_1^1 + w\sigma_2\lambda_1 I_2^1 - (\mu+\gamma)I_{12}^1 \\
\frac{dR_1^1}{dt} = \gamma I_1^1 - \lambda_2 R_1^1 - \mu R_1^1 \\
\frac{dR_2^1}{dt} = \gamma I_2^1 - w\lambda_1 R_2^1 - \mu R_2^1 \\
\frac{dR_{12}^1}{dt} = \gamma(I_{12}^1 + I_{2(1)}^1 + I_{1(2)}^1) - \mu R_{12}^1 \\
\frac{dI_{2(1)}^1}{dt} = \lambda_2 R_1^1 - (\mu+\gamma)I_{2(1)}^1 \\
\frac{dI_{1(2)}^1}{dt} = w\lambda_1 R_2^1 - (\mu+\gamma)I_{1(2)}^1
\end{cases}
$$

where

$$
\begin{aligned}
\lambda_1 &= \beta(I_1^0 + I_{12}^0/2 + I_{1(2)}^0 + I_1^1 + I_{12}^1/2 + I_{1(2)}^1) \\
\lambda_2 &= \beta(I_2^0 + I_{12}^0/2 + I_{2(1)}^0 + I_2^1 + I_{12}^1/2 + I_{2(1)}^1).
\end{aligned}
$$

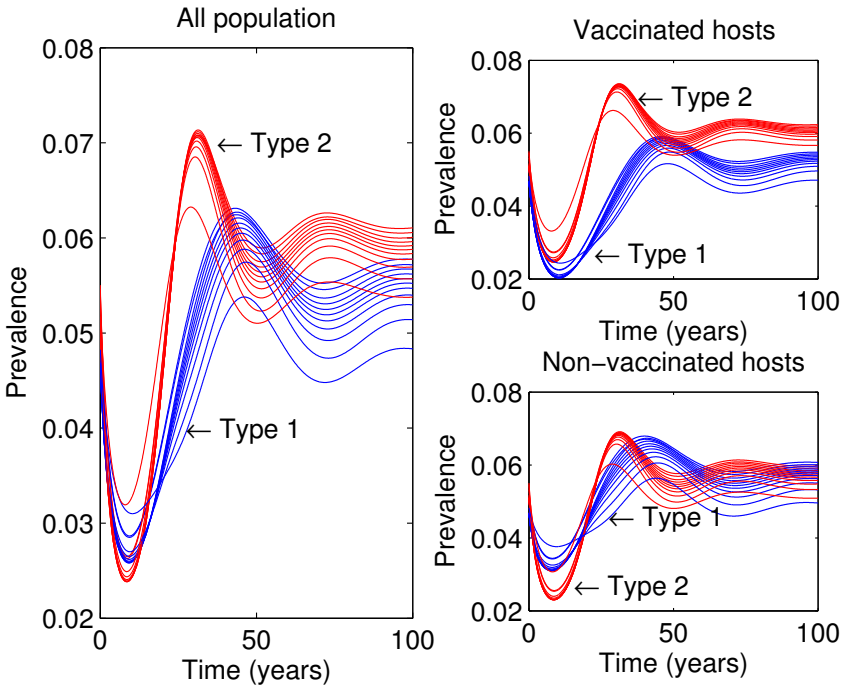


Inference under the SIR-2 model

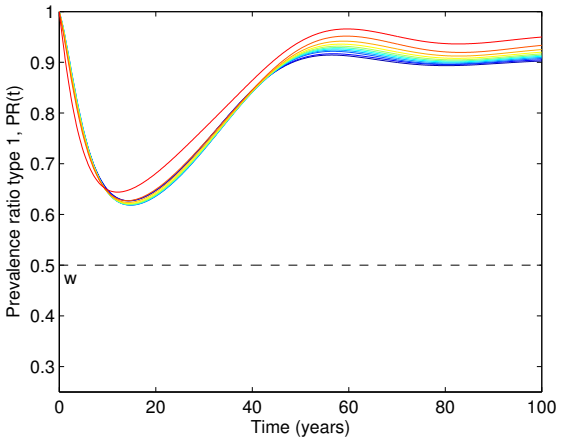
We generate hypothetical trajectories using the SIR-2 equations, for given coverage ρ, demographic parameter μ and clearance rate γ (e.g. Figure S3 a). The prevalence ratio of type 1 pathogen in vaccinated and non-vaccinated hosts still shows significant deviation from the true vaccine protection parameter w (Figure S3 b).

For model-based inference, we vary the other four parameters: β representing the transmission rate, w, describing vaccine protection against type 1 where vaccine efficacy is $VE = 1 - w$, and the direct competition parameters σ_1 and σ_2 acting at co-infection. By sampling the pathogen type prevalences at different time points post-vaccination, and fitting the dynamic model to these 'data', we can recover back the original parameters. We group observations into pathogen-free hosts: $S + R_1 + R_2 + R_{12}$, hosts carrying type 1: $I_1 + I_{1(2)}$, hosts carrying type 2: $I_2 + I_{2(1)}$, and hosts carrying both types I_{12}. We do this grouping both for vaccinated and non-vaccinated hosts. Given that all host proportions sum to 1, this results in 6 data points at each snapshot considered. If observations are perfect, the backward inference works very well: all four parameters can be estimated accurately from model trajectories, even when the prevalence proportions are extremely low (data not shown). As we realistically allow for sampling error, we superimpose a multinomial sampling process on the deterministic model proportions, and then apply nonlinear least squares optimization to the resulting synthetic *sample proportions*. In this case, larger sample sizes are needed to identify σ_1 and σ_2, because the expected pathogen prevalence (and consequently of constituent subtypes) is rather low in a SIR-2 framework. As γ decreases, the prevalence of infection increases (and there are less immune individuals), augmenting our statistical power for the inference of competition parameters, for a given sample size (Figure S4 c-d). The inference of β and $w = 1 - VE$ is however accurate, even in the presence of observation error.

Notice that here we have assumed that the serological status of sampled individuals is not known, which may not always be the case with ever advancing molecular technologies. Thus model fitting was performed only on aggregated carriage status variables. It is likely that if the serological status (naive/immune to 1/immune to 2/ immune to both) is known, we will have much more power to extract competition parameters σ_1 and σ_2 by fitting all the ODE's to the complete data, even when γ is large.

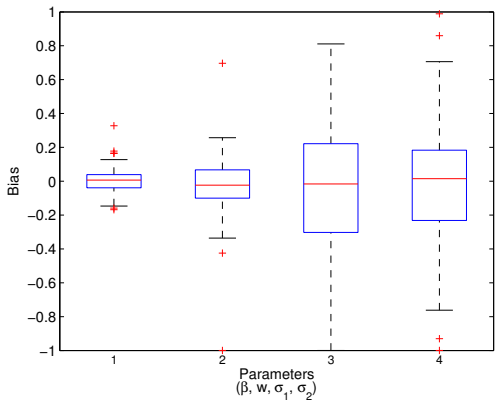


(a) Dynamics in the years post-vaccination

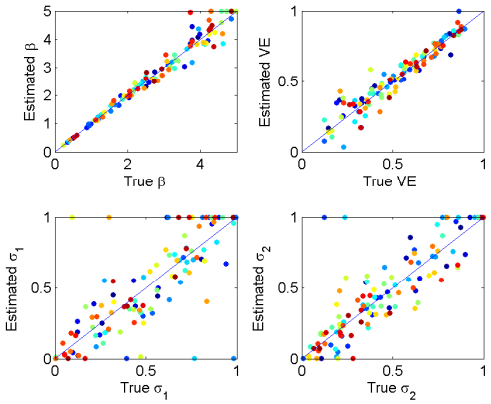


(b) Prevalence ratio and direct competition

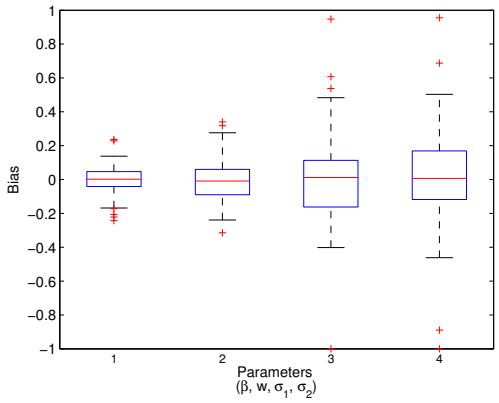
Figure S6: **SIR-2 model framework.** a) The lines show dynamics post-vaccination for $\beta = 2$, $\gamma = 0.2, \mu = 0.0167$, and different competition coefficients σ_1 (<=> $\sigma_2 = 0.25$). The endemic pre-vaccine equilibrium is assumed as initial conditions. Vaccination is implemented at time $t = 0$ with coverage $\rho = 0.5$ and vaccine efficacy $VE = 0.5$. Average life-expectancy of any individual in the population is $1/\mu$, assumed 60 years. The panels illustrate oscillatory prevalences of type 1 and type 2 in the entire population (left), and in the vaccinated and non-vaccinated hosts (right). Such oscillations are typical of an SIR framework, making the sensitivity of dynamic prevalences to σ_1 and σ_2 only moderate. b) Prevalence ratio in the years post-vaccination, does not reflect accurately the true relative risk in the SIR-2 model, and there is minimal sensitivity to direct competition parameters.



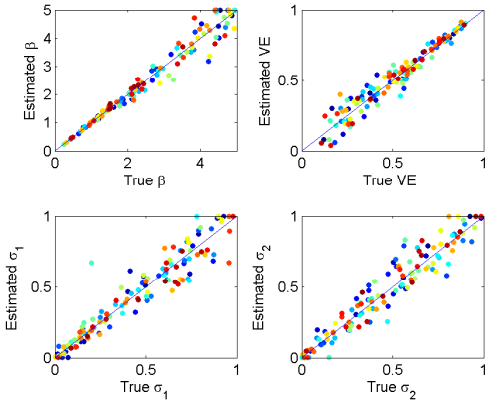
(a) Bias ($\gamma = 0.2$)



(b) Range ($\gamma = 0.2$)



(c) Bias ($\gamma = 0.1$)



(d) Range($\gamma = 0.1$)

Figure S7: **Dynamic model fitting in the SIR-2 framework tested for random parameter combinations.** a)-b) Bias and parameter range in the high clearance rate scenario, thus smaller prevalence at endemic equilibrium pre-vaccination. c)-d) Bias and parameter range in the smaller clearance rate case, thus lower prevalence of carriage at endemic equilibrium pre-vaccination. The parameters are ordered as $\beta, w, \sigma_1, \sigma_2$. We simulated the model for 150 different parameter combinations. Fixed parameter values: $\mu = 0.0167$, $\rho = 0.5$, sample size $N = 5000$. Initial conditions are always fixed at pre-vaccine endemic equilibrium (solved numerically for each parameter combination). The time points (years in this model) post-vaccination used are $t_i = 5, 10, 20, 30$, where the population is sampled multinomially for infection status (uninfected, infected with type 1, with type 1, with both). The performance of the method to estimate the competition coefficients σ_1 and σ_2 in the SIR-2 model depends on the prevalence that can be observed in a population. Lower rate of recovery γ leads to higher prevalence of infection, thus enables better estimation of σ_1, σ_2, for same N, due to less sampling error. The inference of β and vaccine efficacy remains remarkably stable.

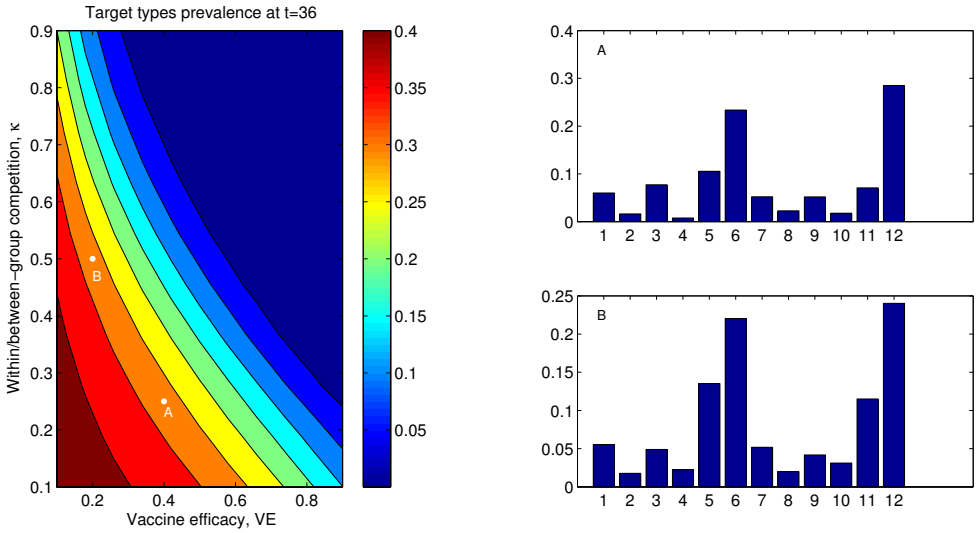


Figure S8: **Parameter correlation in the SIS-2 model.** Parameters κ and VE may appear correlated at one scale, namely, if only global prevalence of target types at a given time post-vaccination is available: Prev (type 1) $= I_1 + I_{11} + I_{12}/2$ (contour plot on the left). However, when one zooms further into each scenario (marked white dots), by using the information contained in the finer-scale epidemiological variables: $S, I_1, I_2, I_{11}, I_{22}, I_{12}$ across vaccinated (listed 1-6) and non-vaccinated individuals (listed 7-12), differing between A and B (right sub-panels), the parameters κ and VE should be suitably detangled. Fixed parameter values $\beta = 6, \mu = 0.02, \gamma = 0.6, \rho = 0.5, \sigma = 1$. Initial conditions at pre-vaccine endemic equilibrium.