Online Supplement

S1. Modification of vaccination effectiveness by vaccination history

There are plausible mechanisms causing interference of previous seasons' vaccination (and/or natural infection) with the immune response to the current season's vaccination: If, e.g., the two vaccines or, more generally, the cause of the previous immune response against influenza antigens and the current vaccine, were antigenically similar, then persisting antibodies may inactivate vaccine antigen, thus interfering with the vaccine response ("antigenic distance hypothesis", [1]). This mechanism would result in effect modification, i.e. the modification of VE by vaccination history. The following model could then be used instead of (5)

$$\text{logit}\left(\Pr(\text{case}_i = 1|v, h_2, \ldots, h_K)\right) = \beta_0 + \beta_v v + \sum_{m=2}^{K} \alpha_m \nu_m + \sum_{m=2}^{R} \gamma_m d_m, \tag{S1.1}$$

where R is the number of vaccination history subsets that are associated with distinct VEs, and

$$d_m = \begin{cases} 1 \text{ if belonging to subset } m, \forall m \in (1, \ldots, R) \\ 0 \text{ otherwise.} \end{cases}$$

Model (S1.1) gives rise to the following R history-specific VE estimates:

$$\widehat{\phi}_m = \begin{cases} 1 - \exp \widehat{\beta}_v & \text{if } d_1 = 1 \\ 1 - \exp(\widehat{\beta}_v + \widehat{\gamma}_m) & \text{otherwise} \end{cases} \tag{S1.2}$$

S2. Simulation model

Model description

We simulated test-negative design (TND) studies of seasonal influenza VE. The source population of 10^6 consisted of subjects with five different levels of health behavior $\xi \in \Xi$, where $\Xi = (1, \ldots, 5)$. The assumed prevalences q_ξ for different levels of ξ, the probabilities to get vaccinated in any given season p_ξ and the probabilities to seek medical care for an acute respiratory illness that met the case definition, w_ξ are listed in Table S2.1. We simulated six influenza seasons that gave rise to $2^6 = 64$ distinct full vaccination

ξ	q_ξ	w_ξ	p_ξ
1	0.3	0	0.2
2	0.2	0.2	0.2
3	0.2	0.75	0.55
4	0.2	0.9	0.65
5	0.1	0.95	0.8

Table S2.1: Assumed values of parameters determining the effect of health behavior ξ on vaccination uptake and health care-seeking behavior in simulation studies: Prevalence q_ξ, vaccination uptake w_ξ and probability of health care-seeking when case definition was met (ARI, ILI, etc.).

histories (including current season). The per-season vaccination probabilities, that remained constant for each subject, and their distribution in the population, determined the distribution of vaccination histories in the population. For example, each of the $\binom{6}{2}$ trajectories of two vaccinated and four unvaccinated seasons previously and including the current season was calculated as the sum of the probabilities contributed by each ξ-type, weighted by the ξ-types prevalence, e.g.

$$\Pr(\overline{v} = (0,1,0,0,0), v = 1) = \sum_{\xi \in \Xi} q_\xi \; p_\xi^2 (1 - p_\xi)^4, \tag{S2.1}$$

Generally, the probability of k vaccinated seasons out of six is

$$\Pr(n_v = k) = \binom{6}{k} \sum_{\xi \in \Xi} q_\xi \; p_\xi^k (1 - p_\xi)^{6-k}, \tag{S2.2}$$

where n_v is the number of vaccinations received in all the previous seasons. We further assumed only one influenza virus entity to circulate and assumed that the influenza risk in a particular season was only a function of a subject's immunity, acquired either by vaccination or natural infection.

All individuals were assumed to be immunologically naïve—and thus unvaccinated—before the start of vaccination prior to the first simulated season. All subjects infected with the influenza virus remained immune to influenza for the remainder of the season. Those with natural immunity in a given season, say k, either from natural infection in that season or carry-over had a probability ω of carrying their natural immunity over to the following season, $k+1$. Similarly, we assumed that subjects with vaccine-derived immunity would carry over their immunity to the following season with probability ρ. Carry-over of infection-derived immunity was assumed to be at least as likely as vaccine-derived immunity, i.e., $\omega \geq \rho$. Therefore, the subdiagonals in all Figures—except for Figure 1, which depicts a DAG—are left blank.

We assumed that—except for the sensitivity analysis of Online Supplement S4—seasonal risk of influenza infection in individuals susceptible at the beginning of a season was constant at 0.2 and VE (ϕ) was 60% over all seasons, while the seasonal risk for non-influenza respiratory infection was 0.3. Of those with influenza infection, half developed a syndrome making them eligible for study inclusion, while that proportion was only 30% in those with non-influenza respiratory infection, independent of their ξ-type, but the probability of seeking care with given symptoms was only determined by w_ξ.

VE represents the probability that a vaccinated, previously susceptible, individual is fully protected from influenza. The complement, $1-\phi$, is the probability that a vaccinated individual will remain fully susceptible despite vaccination. Vaccination-derived immunity was only acquired by subjects without infection-derived immunity and *vice versa*.

There were no measured or unmeasured confounders or covariates other than vaccination history, current vaccination status and health behavior, ξ. Influenza infection was determined with perfect accuracy. When seeking care, all subjects were enrolled into the study and tested for influenza. Study recruitment—i.e., the limitaton of the numbers of subjects enrolled by enrollment targets—was ignored except in the analysis of Type I errors in Online Supplement S3, because our focus was bias and not precision. Unless study enrollment, given a health care visit for acute respiratory illness, depends on either $\overline{v}$, v, or ξ this choice does not affect the average results.

Simulation with vaccination status and history assessed with perfect accuracy

The simulation was implemented by first calculating the event probabilities (case event=influenza ARI and control event=non-influenza ARI). Generally, the number of influenza infections risk for season k was calculated for each full vaccination history stratum as a binomial pseudo-random number using the seasonal attack rate in susceptibles λ_I and the number susceptible at the beginning of season k, using the stratum size and the proportion susceptible, $S_k = 1 - I_k - V_k$, where I_k is the prevalence of protection from natural influenza infection at the beginning of season k and V_k is the prevalence of protection due to vaccination. Again, immunity is always assumed to be absolute and prevalence of immunity is not assumed to decrease during an influenza season. For simplicity, vaccination history indices are omitted. I_k was then iteratively calculated, for each vaccination history stratum separately, as

$$I_k = \left(S_{k+1}\lambda_I(k+1) + I_{k+1}\right)\omega, \forall k \in (4, \ldots, 0) \tag{S2.3}$$

and

$$V_k = \left(V_{k+1}\,\rho + v_k\,(1 - I_k - V_{k+1}\,\rho)\,\phi_k\right), \forall k \in (4, \ldots, 0) \tag{S2.4}$$

where $I_5 = V_5 = 0$, i.e. absence of vaccine- or infection-derived immunity at the beginning of the first season. Note that, by assumption, immunity from infection and vaccination are mutually exclusive. Prevalence of natural and vaccine-derived influenza immunity decayed from season to season at independent, constant rates. Together with the infection and vaccination events in the past seasons, these decay rates determined the level of susceptibility to influenza infection during the final season, which was of interest for VE estimation.

Thus we calculated S_0 iteratively for all 64 vaccination histories. Using these probabilities and the sizes of the respective source populations, pseudo-random numbers were generated from a binomial distribution. This was done 10,000 times (number of simulations), unless otherwise indicated. Similarly, numbers of controls were generated for each stratum (vaccination history), using λ_{nI}, independent of vaccination or infection history. These procedures resulted in two $10,000 \times 64$ matrices, one for cases and one for controls. The nth row of the respective matrix represented data from the nth simulated study.

Simulation with vaccination status and history assessed with error

To investigate the impact of misclassification of current and prior season's vaccination status we assumed that misclassification of current and prior season's vaccination status, v and $\tilde{v}$, respectively, was characterized by the same sensitivity $\sigma = 95\%$ and specificity $\zeta = 90\%$.

Each subject falls into either of the following four categories:

1. Unvaccinated both in previous and current season, n_1
2. Unvaccinated in prior season—vaccinated in current season, n_2
3. Vaccinated in prior season—unvaccinated in current season, n_3
4. Vaccinated in prior season—vaccinated in current season, n_4

let $\mathbf{n} = (n_1, \ldots, n_4)$. The observed numbers, $\mathbf{n}^{*\intercal} = (n_1^*, \ldots, n_4^*)$ will be multinomially distributed according to probabilities $\mathbf{p}^\intercal = (p_1, \ldots, p_4)$ and the total number $N = \sum_{i=1}^4 n_i$; accordingly, $\mathrm{E}(\mathbf{n}^*) = \mathbf{p}N$. The

probabilities $\mathbf{p}$ are given by the matrix product of

$$\mathcal{C} = \begin{bmatrix} \zeta^2 & (1-\sigma)\zeta & \zeta(1-\sigma) & (1-\sigma)^2 \\ (1-\zeta)\zeta & \sigma\zeta & (1-\zeta)(1-\sigma) & \sigma(1-\sigma) \\ \zeta(1-\zeta) & (1-\sigma)(1-\zeta) & \zeta\sigma & (1-\sigma)\sigma \\ (1-\zeta)^2 & \sigma(1-\zeta) & (1-\zeta)\sigma & \sigma^2 \end{bmatrix} \tag{S2.5}$$

and the normalized vector $\frac{\mathbf{n}}{N}$.

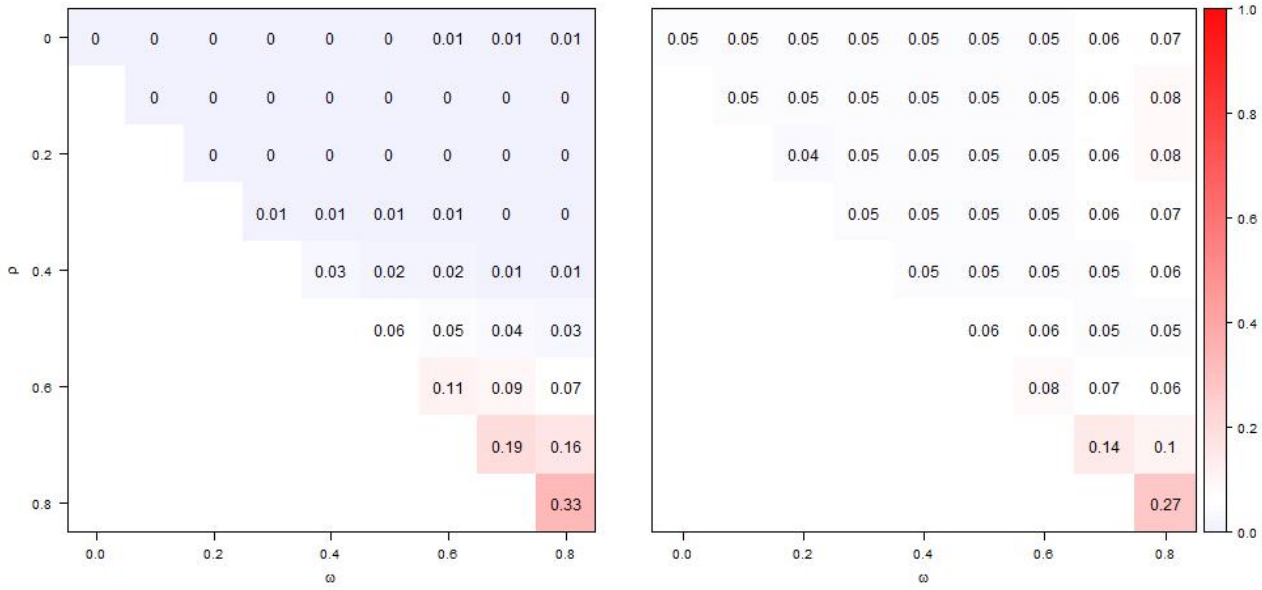
Using these parameters the 4×4 matrix $\mathcal{C}$ (S2.5) can be calculated. For each simulation, $\mathcal{C}$ was multiplied by the vector having the true numbers in each category as entries, devided by the total, to obtain $\mathbf{p}$ (see text). This was used to generate pseudo-random numbers from a multinomial distribution.

Example R code

Sample R code is available as online material.

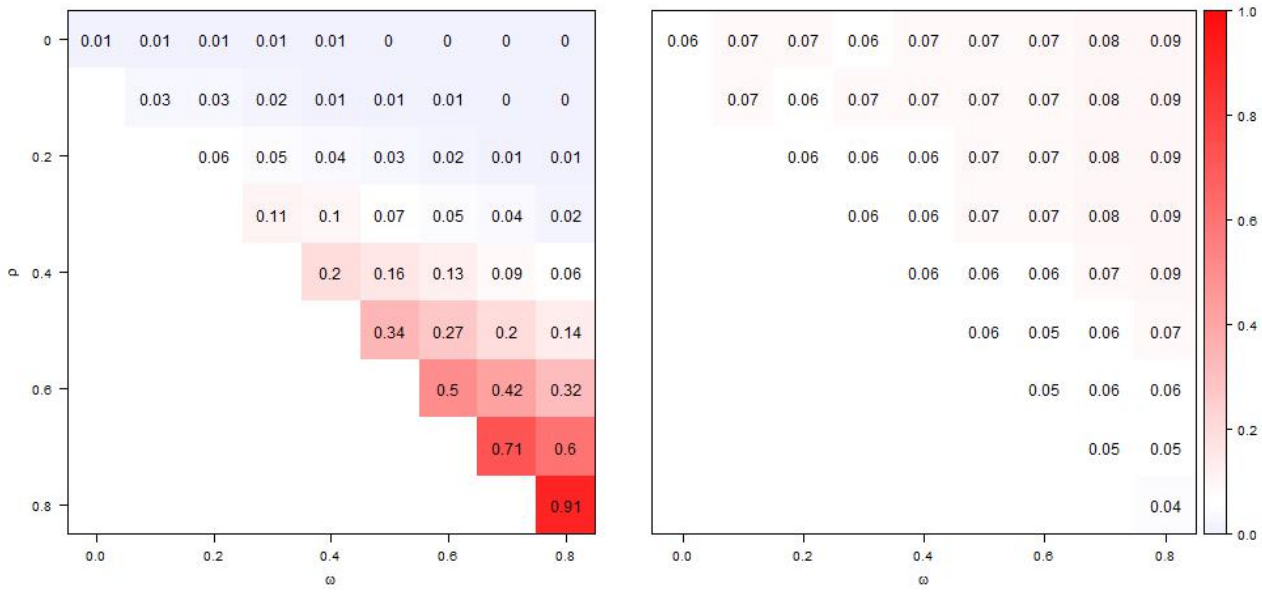
S3. Statistically significant spurious differences in VEs from models 2 and 4

For both model 2 and model 4 we estimated the probabilities that the hypothesis of modification of VE in the current season by vaccination in the prior season would be accepted, despite the absence of such an effect (Type I error). First, this was done for accurate assessment of current and prior vaccination status, then for misclassified vaccination status.



(a) (b)

Figure S3.1: Proportion of statistically significant differences between VE for those vaccinated vs. those unvaccinated in previous season (a) from model 2 and (b) from model 4, when current and prior vaccination status is accurately measured, with 1000 cases and 2000 controls; 10 000 simulations per parameter combination. Blue cells indicate probabilities of rejection of the null hypothesis of no effect modification (Type I error) lower than $\alpha = 0.05$, while shades of red indicate Type-I error probabilities $> \alpha$.



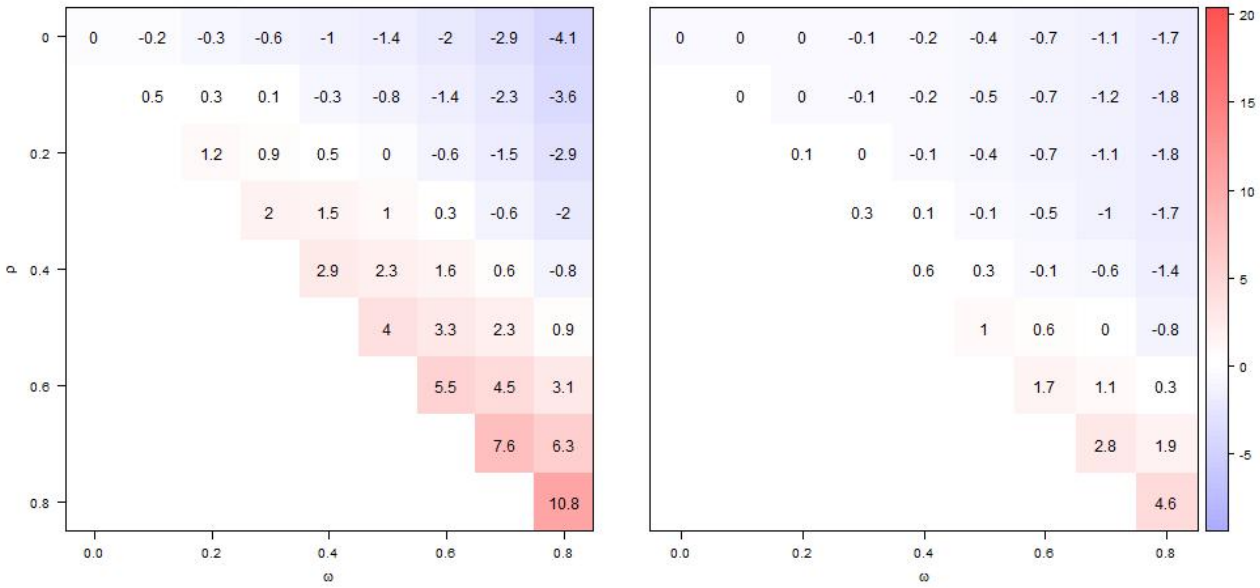
(a) (b)

Figure S3.2: Proportion of statistically significant differences between VE for those vaccinated vs. those unvaccinated in previous season (a) from model 2 and (b) from model 4, when current and prior vaccination status is misclassified ($\sigma = 90\%, \zeta = 95\%$), with 1000 cases and 2000 controls; 10 000 simulations per parameter combination. Blue cells indicate probabilities of rejection of the null hypothesis of no effect modification (Type I error) lower than $\alpha = 0.05$, while shades of red indicate Type-I error probabilities $> \alpha$.

S4. Simulations using variable VE and seasonal influenza risk

We investigated the effect of seasonally variable VE and influenza risk. First we assumed that VE was 20, 60, 30, 40, 30 and 60% in seasons 1 through 6, respectively, and the corresponding seasonal influenza risk was 20, 10, 15, 18, 15 and 10% in those susceptible at the beginning of the season (Figure S4.1). We then assumed that VE in the current season was 20% instead and influenza risk in that season was 30% (Figure S4.2).

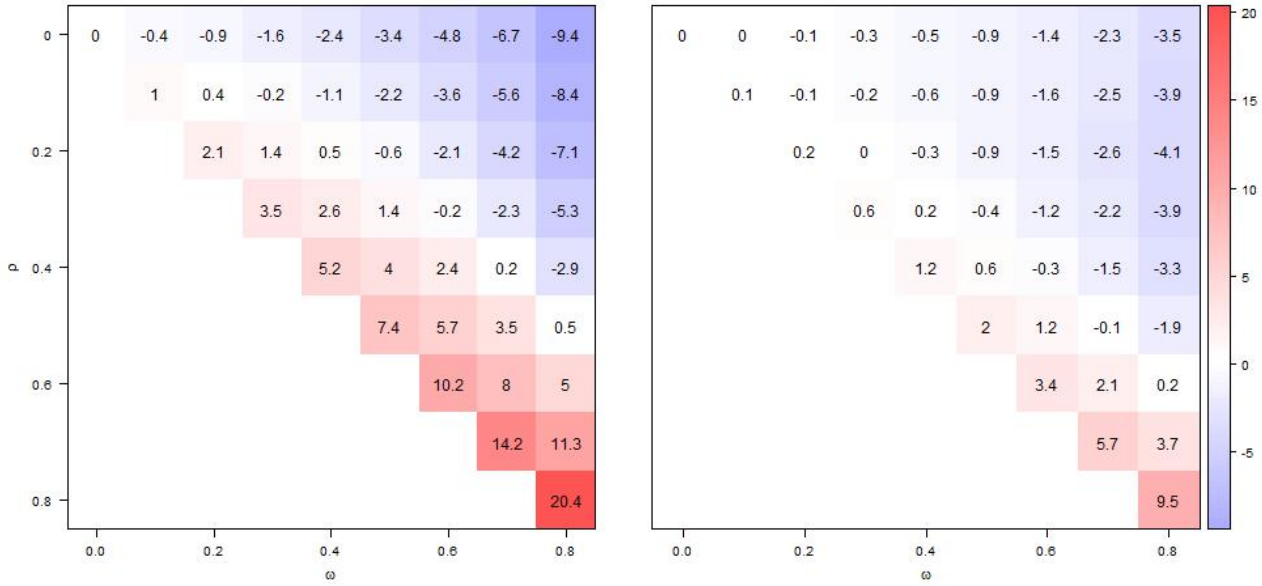
Current season's VE 60%



(a) (b)

Figure S4.1: VE estimates with variable VE and seasonal influenza risk in the past and a current VE of 60% (a) unadjusted for prior season's vaccination status (model 1) and (b) adjusted for prior season's vaccination status (model 3), when current and prior vaccination status is accurately measured; 10 000 simulations per parameter combination. Blue cells indicate negative bias (VE estimates too low), while shades of red indicate positive bias (VE estimates too high).

Current season's VE 20%



(a) (b)

Figure S4.2: VE estimates with variable VE and seasonal influenza risk in the past and a current VE of 20% (a) unadjusted for prior season's vaccination status (model 1) and (b) adjusted for prior season's vaccination status (model 3), when current and prior vaccination status is misclassified ($\sigma = 90\%, \zeta = 95\%$); 10 000 simulations per parameter combination. Blue cells indicate negative bias (VE estimates too low), while shades of red indicate positive bias (VE estimates too high).

S5. Sensitivity analysis for accuracy of vaccination status assessment and bias in VE

To investigate the role specific assumptions regarding accuracy of vaccination status assessment for current and prior season, we varied sensitivity and specificity values: 80,85,90,95 and 98%, both for VE estimates not adjusted (model 1, Figure S5.1) and adjusted for prior season's vaccination status (model 3, Figure S5.2). Per parameter setting, we ran 1000 simulations.

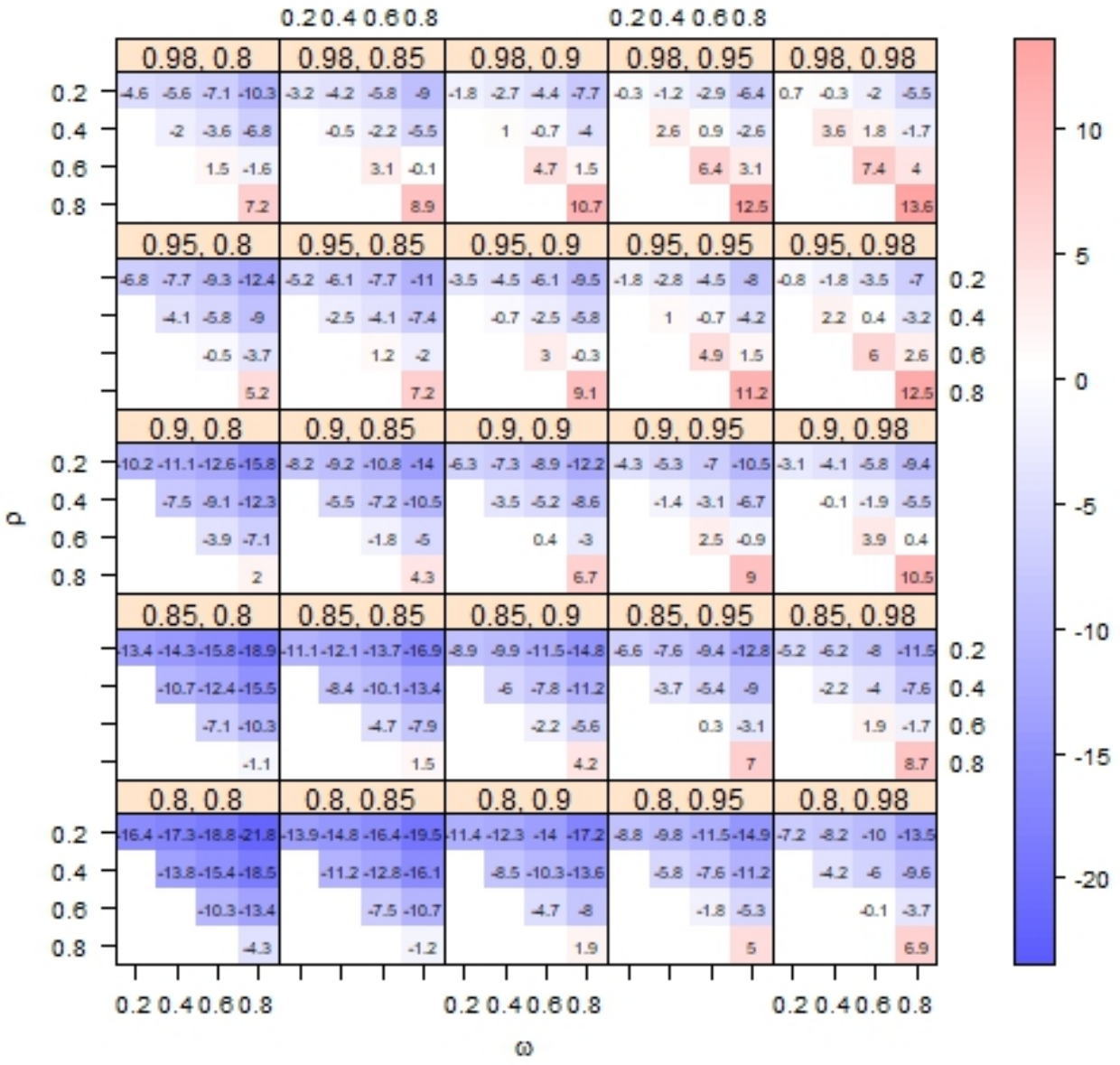


Figure S5.1: Bias in VE estimates that are not adjusted for prior season's vaccination status (model 1), by sensitivity and specificity of current and prior season's vaccination status assessment. 1000 simulations per parameter combination. Shades of blue indicate negative bias (VE estimates too low) and shades of red indicate positive bias.

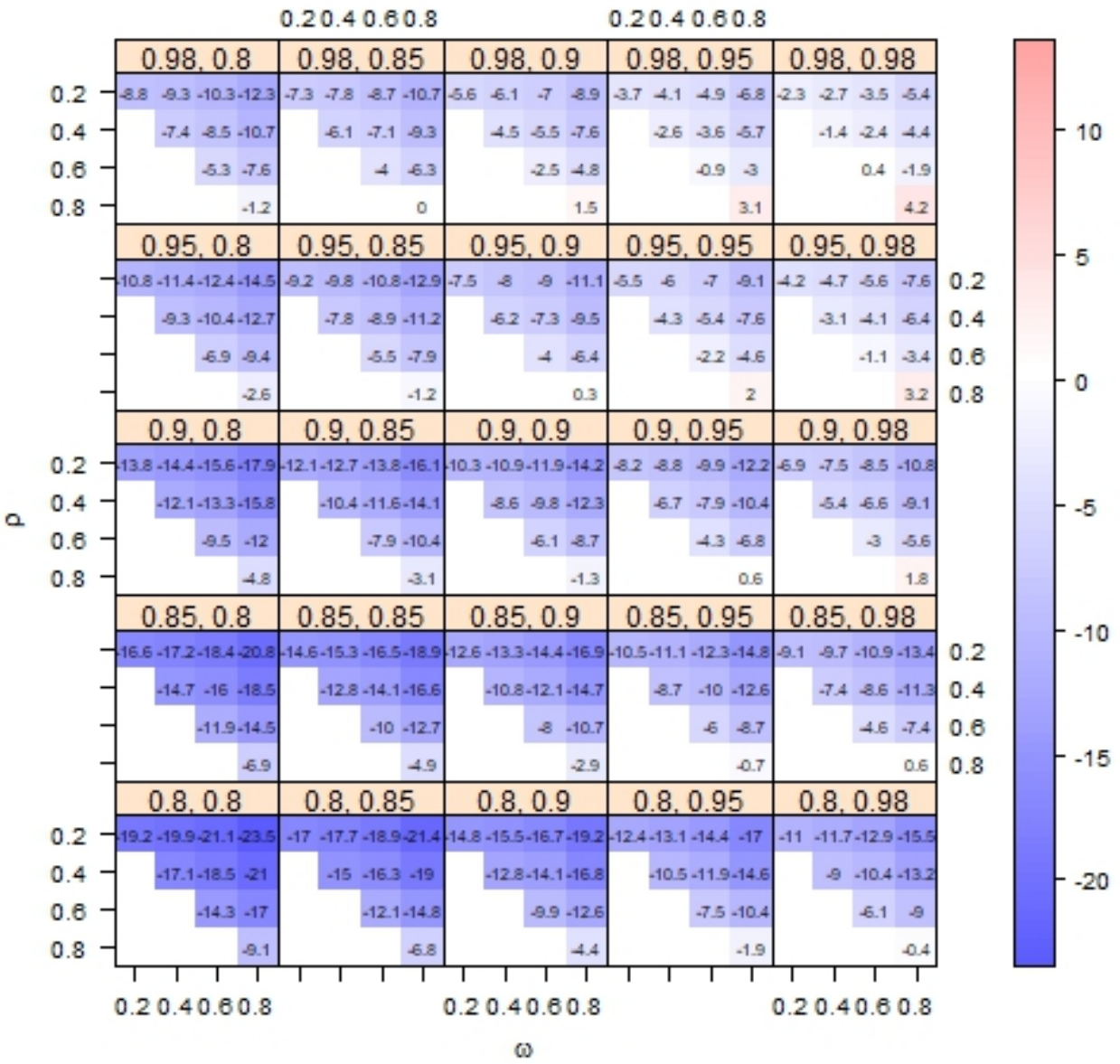


Figure S5.2: Bias in VE estimates that are partially adjusted for prior season's vaccination status (model 3), by sensitivity and specificity of current and prior season's vaccination status assessment. 1000 simulations per parameter combination. Shades of blue indicate negative bias (VE estimates too low) and shades of red indicate positive bias.