Web Appendix 1

Individual-based transmission hazard model

The individual-based transmission hazard (TH) model described below extends the work of Cauchemez et al. and Tsang et al. to examine influenza vaccine effectiveness in a prospectively followed cohort of households(1-4). The TH model computes the hazard for infection from the community for all individuals across the entire observed influenza season, and, separately, the hazard for infection from the each infected household contact for a 14 day period following the illness onset of the infected contact. Similar to Cox proportional hazards (PH) models, the effects of subject characteristics on these hazards can be evaluated by the inclusion of covariates. Previous analyses have identified age and presence of high risk health conditions as confounders of influenza vaccine effectiveness and these are adjusted for here (5). For each individual, j, in the cohort we observe the following data: y_j, an indicator of RT-PCR confirmed influenza A (H3N2) infection; t_j, the time at which follow-up ends (the day of illness onset for those infected, and the end of the influenza season for those not infected); a_j, subject age; hr_j, an indicator of the presence of high risk conditions; $v_{j,}(t)$, subject vaccination status at time t.

We model the hazard of infection at time t from the community for each cohort member, j, as

$$\lambda_{j,c}(t) = P_c(t) * \psi_c * \exp\{\beta_1 * X_{1j} + \beta_2 * X_{2j} + \beta_3 * X_{3j} + \beta_4 * X_{4j}(t)\} \qquad (1)$$

Equation 1 is comprised of three factors. $P_c(t)$ is a time-varying proxy for the baseline hazard of influenza infection from the community defined by weekly counts of influenza cases reported to the Michigan Department of Health and Human Services Disease Surveillance System (6) standardized to the peak week of activity. The second factor, ψ_c, is a scaling parameter for the baseline community hazard. The last factor is a linear combination of four subject characteristics, with X_{1j} and X_{2j} indicating age groups of 9 to 17 years ($9 \geq a_j < 18$) and ≥18 years ($a_j \geq 18$), respectively, X_{3j} indicates presence of EMR documented high risk condition ($hr_j = 1$), and $X_{4j}(t)$ is an indicator of documented vaccination ($v_{j,}(t) = 1$). Vaccination status was allowed to vary by time with subjects considered to be vaccinated 14 days after the date of documented vaccination. In contrast to previous TH models that modeled time in days during

the relatively short period of household follow-up once an index case had been ascertained, here, t represents the time in days across the entire period of influenza activity observed in the cohort.

We then assume that the serial interval (time in days between symptom onset of prior and subsequent influenza cases) in those households where influenza has been introduced, follows a Weibull distribution

$$f(\tau) = \exp\left\{-\left(\frac{\tau}{\gamma}\right)^{\alpha}\right\} - \exp\left\{-\left(\frac{\tau+1}{\gamma}\right)^{\alpha}\right\} \qquad (2)$$

as has been done in previous models (1-4). In Equation 2, τ is the household exposure time in days from the day of illness onset of the infected household contact through 14 days post illness onset, and α and γ are shape parameters for the Weibull distribution.

We model the daily hazard of infection for each household contact, *j*, from each infected individual, *i*, in the household as

$$\lambda_{i\to j}(\tau) = \lambda_{hh} * f(\tau) * \exp\{\beta_1 * X_{1j} + \beta_2 * X_{2j} + \beta_3 * X_{3j} + \beta_4 * X_{4j}(t)\} \qquad (3)$$

where λ_{hh} is a constant parameter representing the baseline household hazard of infection.

The overall daily hazard of infection for each individual, *j*, is the sum of their hazards from the community and from each infected member of their household, *i*.

$$\lambda_j(t) = \lambda_{j,c}(t) + \sum_i \lambda_{i\to j}(\tau) \qquad (4)$$

Parameter estimation

We let θ denote the set of parameters {α, γ, ψ_c, λ_{hh}, $\beta1$, $\beta2$, $\beta3$, $\beta4$}; all elements in θ were estimated in a Bayesian framework using Markov Chain Monte Chain (MCMC) methods.

The contribution of the likelihood function from each infected individual is the probability that they were infected on their day of illness onset, t_{jo}, multiplied by the probability that they escaped infection each day prior to their illness onset.

$$P\left(y_j = 1, t_j = t_{jo}\right) = \left(1 - \exp\{-\lambda_j(t_{io})\}\right) * \exp\{-\sum_{d=1}^{t_{jo}-1} \lambda_j(d)\} \qquad (5)$$

The contribution to the likelihood function for each individual not infected during the study period is the probability that they escaped infection each day of the influenza season.

$$P\left(y_j = 0, t_j = t_{end} + 1\right) = \exp\{-\sum_{d=1}^{t_{end}} \lambda_j(d)\} \qquad (6)$$

The overall log-likelihood function combining the likelihood functions for both infected and uninfected individuals is

$$L = \sum_{j:y_j=1} \log\left(1 - \exp\{-\lambda_j(t_{io})\}\right) - \sum_j \sum_{d=1}^{t_j} \lambda_j(d) \qquad (7)$$

In Equations 5, 6, and 7 t_j represents the end of the period risk: the day of illness onset for those infected, and the end of the influenza season for those who were not infected.

Non-informative prior distributions were used for each parameter. Additionally, as has been done in previous studies to assure model convergence, we constrained the Weibull shape parameters such that the corresponding serial interval function would result in ≥80% of household secondary infections occurring within 14 days of the onset of illness in the index case (3).

The estimated posterior distribution of the parameters, θ, given the observed data is proportional to the likelihood of the data multiplied by the prior distributions of the parameters

$$P\left(\theta \middle| y_j, t_j, a_j, hr_j, v_j\right) \propto P\left(y_j, t_j, a_j, hr_j, v_j \middle| \theta\right) * P(\theta) \qquad (8)$$

Parameters were updated at each iteration using a Metropolis-Hastings algorithm (7, 8). The MCMC algorithm was run for 15,000 iterations with a burn in of 5,000 iterations. Convergence was visually assessed (Web Figure 1), similarly to (3, 4). Parameter distributions and correlation plots (Web Figure 2) indicated that while mild correlations were observed between some parameters, all parameters were identifiable.

TH model outcomes and simulation

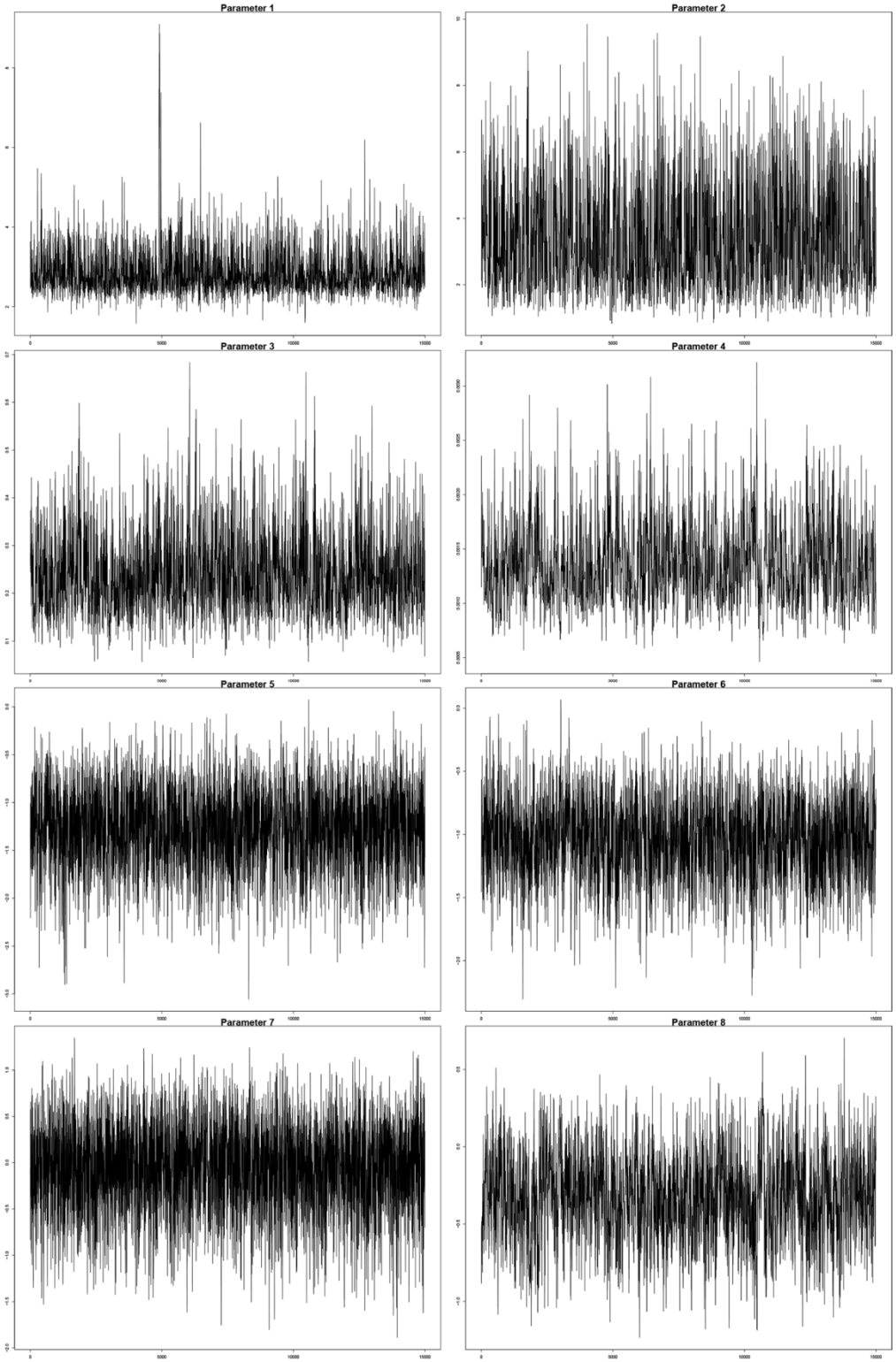
TH model parameter estimates were applied to Equation 1, and daily risks of infection from the community calculated as $1 - \exp\{-\lambda_{j,c}(t)\}$. For each individual, a random value, u, was drawn from a uniform distribution; individuals were considered infected from the community if $u < 1 - \exp\{-\lambda_{j,c}(t)\}$. For each community-acquired infection, the daily risk of household-acquired infection among household contacts was calculated by applying parameter estimates to Equation 3 and calculating daily risk as $1 - \exp\{-\lambda_{i\to j}(\tau)\}$. Household-acquired infections were then simulated by drawing from a random uniform distribution in a similar manner as for community-acquired infections. Additional daily household risk was calculated and infections simulated for the contacts of each subsequent household case allowing for chains of transmission with a maximum length of 4 (primary, secondary, tertiary, and quaternary cases). These simulations were run for 1000 iterations. Median numbers of infections and 95% confidence intervals (CI) were calculated from the distribution of simulation results.

We applied the MCMC algorithm (as described above) to simulated data to determine whether unbiased estimates of the parameters used to simulate the data could be recovered. A total of 5 simulations of community and household-acquired infections in the study population (n=1441) were carried out. We applied the MCMC algorithm to these each simulated dataset to recover the parameters (Web Table 3). To account for random variation in simulated data, we also estimated parameters from a pooled dataset that included all 5 simulations (n=7205). All recovered parameter estimates from the pooled dataset, and nearly all from individual simulations were similar to the parameters used to simulate the data and credible intervals included the simulation parameter. However, estimates of the α parameter, in particular, were systematically higher than the simulation parameter, and credible intervals were relatively wide. This could indicate some level of bias in the inferential framework; however, we would expect this bias to be limited to serial interval estimates given that estimates of the Weibull shape parameters defining the serial interval probability distribution (α and γ) were not correlated with the estimates of other parameters (Web Figure 2). Further, the serial interval distribution estimated from the actual data (2.1 days) was consistent with estimates using standard methods (2.5 days) and previously reported estimates (2.2 days) (9).

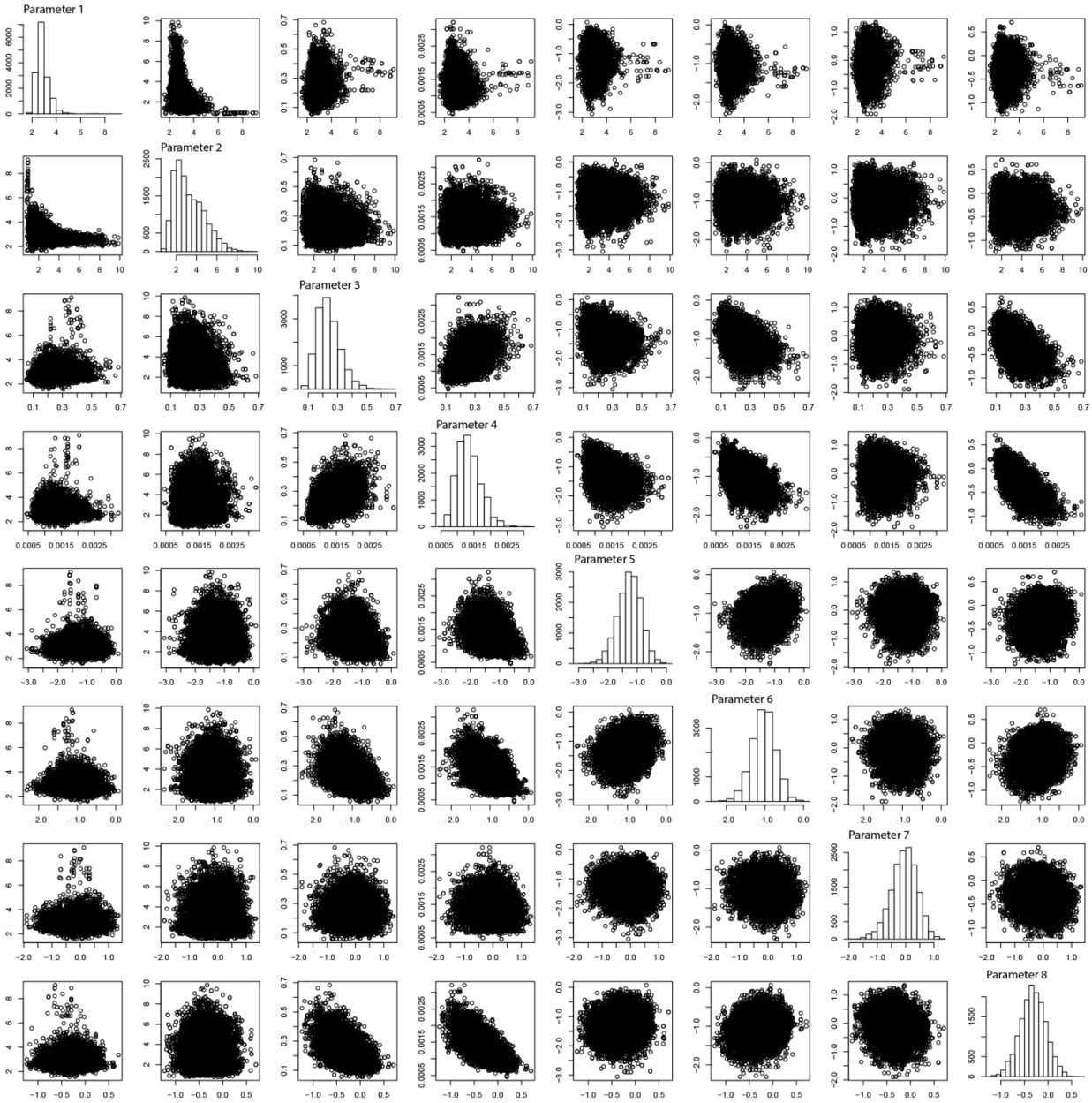
TH model comparison

Overall vaccine effectiveness estimates were produced using the full model described above with age and presence of high risk health conditions considered as confounders; however, the value of reduced models for prediction was explored. Four alternative models were considered including 1) constant community hazard of transmission, 2) no effect of subject age on susceptibility, 3) no effect of high risk health conditions on susceptibility, and 4) no effect of vaccination on susceptibility. Model fit was evaluated by the deviance information criterion (DIC) (10) and a simulation-based chi-square test comparing the number of community-acquired and household-acquired influenza A (H3N2) infections observed in the data and predicted by each model (2, 3).

Web Figure 1. MCMC trace plots for estimation of parameters in the primary vaccine effectiveness model.



Web Figure 2. Distributions and pairwise correlations of parameters estimated in the primary vaccine effectiveness model.



Web Table 1. Parameter estimates for full and alternative individual-based transmission hazard model specifications

	Full Model[a]	Alternative Model 1[b]	Alternative Model 2[c]	Alternative Model 3[d]	Alternative Model 4[e]
α	3.30 (1.33 — 6.59)	3.00 (1.23 — 6.32)	3.32 (1.32 — 6.58)	3.21 (1.28 — 6.38)	3.33 (1.33 — 6.62)
γ	2.87 (2.19 — 4.09)	2.96 (2.23 — 4.33)	2.83 (2.18 — 4.04)	2.88 (2.18 — 4.11)	2.84 (2.21 — 4.02)
λ_{hh}	0.24 (0.11 — 0.42)	0.25 (0.12 — 0.42)	0.12 (0.06 — 0.19)	0.24 (0.12 — 0.43)	0.19 (0.10 — 0.31)
ψ_c	$1.35\text{x}10^{-3}$ ($7.78\text{x}10^{-4}$ — $2.17\text{x}10^{-3}$)	$6.37\text{x}10^{-4}$ ($3.73\text{x}10^{-4}$ — $9.96\text{x}10^{-4}$)	$6.75\text{x}10^{-4}$ ($4.29\text{x}10^{-4}$ — $9.77\text{x}10^{-4}$)	$1.35\text{x}10^{-3}$ ($7.62\text{x}10^{-4}$ — $2.17\text{x}10^{-3}$)	$1.09\text{x}10^{-3}$ ($7.18\text{x}10^{-4}$ — $1.56\text{x}10^{-3}$)
β_1	-1.27 (-2.10 — -0.53)	-1.27 (-2.12 — -0.55)	—	-1.27 (-2.14 — -0.50)	-1.23 (-2.08 — -0.47)
β_2	-1.04 (-1.65 — -0.47)	-1.05 (-1.62 — -0.50)	—	-1.03 (-1.62 — -0.44)	-0.97 (-1.56 — -0.40)
β_3	-0.05 (-1.02 — 0.76)	-0.05 (-1.03 — 0.76)	-0.15 (-1.09 — 0.67)	—	-0.09 (-1.02 — 0.72)
β_4	-0.32 (-0.86 — 0.21)	-0.31 (-0.83 — 0.20)	-0.14 (-0.63 — 0.38)	-0.31 (-0.86 — 0.22)	—

[a]Full model includes covariate terms for age category (<9, 9-17, ≥18), presence of ≥1 electronic medical record documented high-risk health condition, and time-varying vaccination status.

[b]Alternative model 1 is identical to the full model, but with a constant hazard of infection from the community.

[c]Alternative model 2 is identical to the full model, but excludes age category covariates.

[d]Alternative model 3 is identical to the full model, but excludes the high-risk health condition covariate.

[e]Alternative model 4 is identical to the full model, but excludes the vaccination status covariate.

Web Table 2. Comparison of Model Fit for Various Individual Based Transmission Hazard Model Specifications. Household Influenza Vaccine Effectiveness (HIVE) Study 2010-2011.

	Observed Data	Full Model[a]	Alternative Model 1[b]	Alternative Model 2[c]	Alternative Model 3[d]	Alternative Model 4[e]
Infection Order	Cases	Median Cases (95% CI)	Median Cases (95% CI)	Median Cases (95% CI)	Median Cases (95% CI)	Median Cases (95% CI)
Primary	41	43 (31 to 55)	43 (30 to 56)	43 (30 to 57)	43 (32 to 58)	43 (31 to 57)
Secondary	17	15 (7 to 24)	15 (8 to 24)	14 (7 to 23)	15 (8 to 25)	14 (7 to 24)
Tertiary	N/O	3 (0 to 9)	4 (0 to 9)	3 (0 to 8)	3 (0 to 8)	3 (0 to 8)
Quaternary	N/O	0 (0 to 0)	0 (0 to 0)	0 (0 to 0)	0 (0 to 0)	0 (0 to 0)
P-value[f]	-	0.70	0.58	0.76	0.70	0.76
DIC	-	451.64	455.50	456.71	450.91	450.93

Abbreviations: CI, confidence interval; DIC, deviance information criterion; N/O, not observed

[a]Full model includes covariate terms for age category (<9, 9-17, ≥18), presence of ≥1 electronic medical record documented high-risk health condition, and time-varying vaccination status.

[b]Alternative model 1 is identical to the full model, but with a constant hazard of infection from the community.

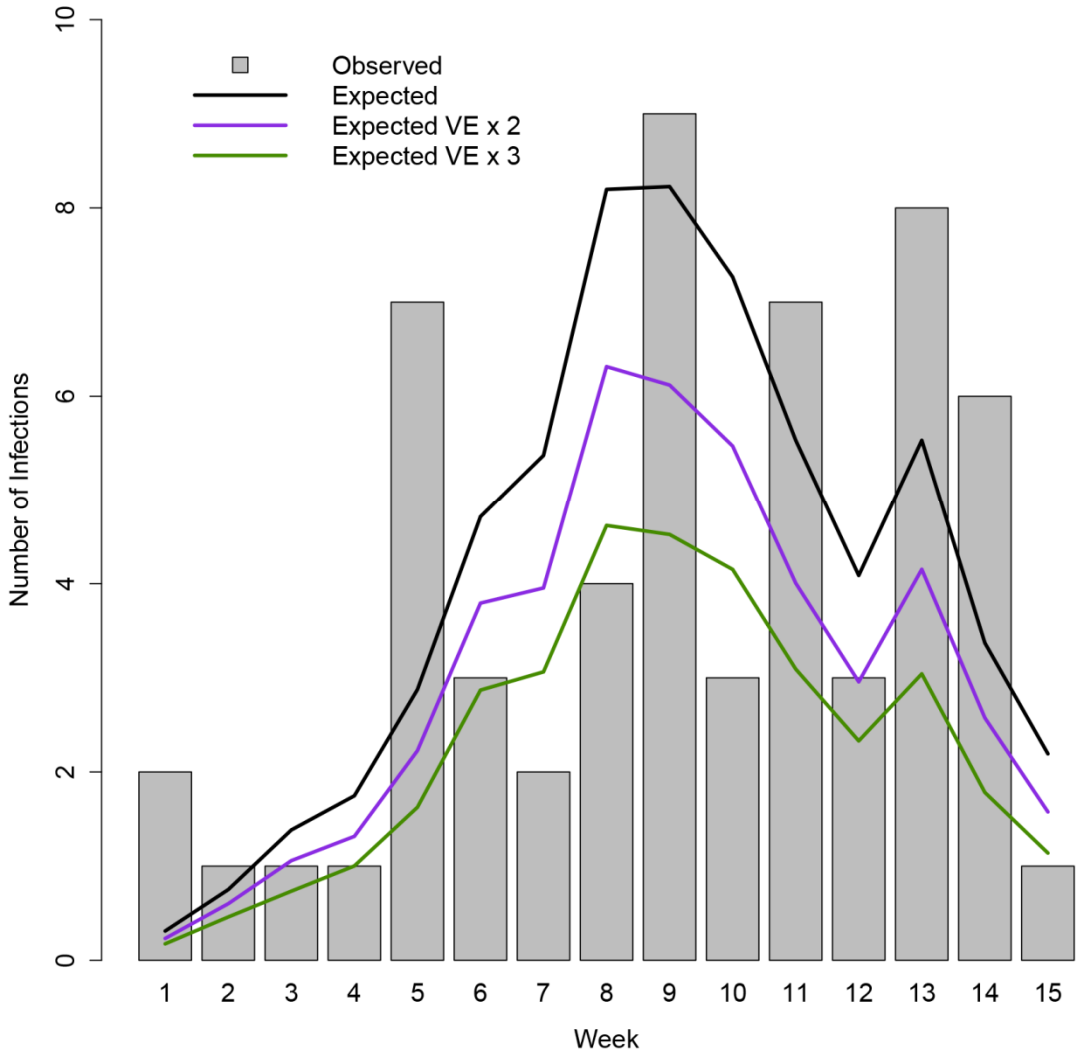
[c]Alternative model 2 is identical to the full model, but excludes age category covariates.

[d]Alternative model 3 is identical to the full model, but excludes the high-risk health condition covariate.

[e]Alternative model 4 is identical to the full model, but excludes the vaccination status covariate.

[f]Simulation-based chi-square test.

Web Figure 3. Weekly Influenza A (H3N2) Infection Counts Observed and Expected with Varying Vaccine Effectiveness. Household Influenza Vaccine Effectiveness (HIVE) Study 2010-2011.



Web Table 3. Recovery of Unbiased Parameter Estimates from Simulated Data[a].

	Simulation Parameters	Simulation 1	Simulation 2	Simulation 3	Simulation 4	Simulation 5	Pooled Simulations
α	3.30	4.71 (2.81, 7.54)	4.08 (2.47, 5.95)	4.35 (2.28, 7.35)	4.15 (2.58, 6.11)	3.00 (1.63, 4.85)	3.61 (2.94, 4.39)
γ	2.87	3.27 (2.78, 3.80)	3.36 (2.85, 3.94)	2.84 (2.40, 3.51)	2.63 (2.28, 3.03)	3.08 (2.51, 3.84)	2.98 (2.75, 3.22)
λ_{hh}	0.24	0.26 (0.15, 0.42)	0.32 (0.16, 0.55)	0.16 (0.08, 0.29)	0.35 (0.21, 0.54)	0.22 (0.13, 0.34)	0.24 (0.18, 0.30)
ψ_c	$1.35\text{x}10^{-3}$	$1.84\text{x}10^{-3}$ ($1.16\text{x}10^{-3}$, $2.72\text{x}10^{-3}$)	$1.28\text{x}10^{-3}$ ($0.78\text{x}10^{-3}$, $2.04\text{x}10^{-3}$)	$1.04\text{x}10^{-3}$ ($0.58\text{x}10^{-3}$, $1.68\text{x}10^{-3}$)	$2.13\text{x}10^{-3}$ ($1.35\text{x}10^{-3}$, $3.13\text{x}10^{-3}$)	$2.12\text{x}10^{-3}$ ($1.39\text{x}10^{-3}$, $3.07\text{x}10^{-3}$)	$1.57\text{x}10^{-3}$ ($1.26\text{x}10^{-3}$, $1.93\text{x}10^{-3}$)
β_1	-1.27	-1.41 (-2.31, -0.64)	-1.24 (-2.03, -0.49)	-1.31 (-2.14, -0.56)	-1.56 (-2.43, -0.86)	-1.32 (-2.00, -0.70)	-1.27 (-1.61, -0.94)
β_2	-1.04	-1.10 (-1.70, -0.54)	-0.84 (-1.44, -0.26)	-1.00 (-1.62, -0.42)	-1.68 (-2.30, -1.10)	-1.11 (-1.66, -0.62)	-1.08 (-1.33, -0.82)
β_3	-0.05	-0.55 (-1.74, 0.40)	0.10 (-0.87, 0.90)	-0.66 (-1.89, 0.31)	-0.18 (-1.20, 0.70)	0.03 (-0.77, 0.72)	-0.13 (-0.55, 0.25)
β_4	-0.32	-0.86 (-1.37, -0.32)	-0.57 (-1.09, -0.03)	0.22 (-0.31, 0.78)	-0.69 (-1.20, -0.17)	-0.48 (-0.92, -0.04)	-0.43 (-0.66, -0.21)

[a]Five simulations of community and household-acquired infections were carried out for the study population (N=1,441). Parameters were recovered from each of the individual simulated datasets as well as a pooled simulation dataset (n=7205) by MCMC.

References

1. Cauchemez S, Carrat F, Viboud C, et al. A Bayesian MCMC approach to study transmission of influenza: application to household longitudinal data. *Stat Med* 2004;23(22):3469-87.
2. Cauchemez S, Donnelly CA, Reed C, et al. Household transmission of 2009 pandemic influenza A (H1N1) virus in the United States. *N Engl J Med* 2009;361(27):2619-27.
3. Tsang TK, Cauchemez S, Perera RA, et al. Association between antibody titers and protection against influenza virus infection within households. *J Infect Dis* 2014;210(5):684-92.
4. Tsang TK, Cowling BJ, Fang VJ, et al. Influenza A Virus Shedding and Infectivity in Households. *J Infect Dis* 2015;212(9):1420-8.
5. Ohmit SE, Petrie JG, Malosh RE, et al. Influenza vaccine effectiveness in the community and the household. *Clin Infect Dis* 2013;56(10):1363-9.
6. Michigan Department of Health & Human Services. Past Michigan Flu Focus Surveillance Reports. 2016. (http://www.michigan.gov/mdhhs/0,5885,7-339-71550_2955_22779_40563-143382--,00.html). (Accessed February 23, 2016).
7. Metropolis N, Rosenbluth AW, Rosenbluth MN, et al. Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics* 1953;21(6):1087-92.
8. Hastings WK. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika* 1970;102(4):97-109.
9. Vink MA, Bootsma MC, Wallinga J. Serial intervals of respiratory infectious diseases: a systematic review and analysis. *Am J Epidemiol* 2014;180(9):865-75.
10. Spiegelhalter DJ, Carlin BP, van der Linde A. Bayesian Measures of Model Complexity and Fit. *Journal of the Royal Statistical Society* 2002;64:583-639.