Investigating the Legacy of the 1918 Influenza Pandemic in Age-Related Seroepidemiology and Immune Responses to Subsequent Influenza A(H1N1) Viruses Through a Structural Equation Model

Cheryl X. P. Chuah, Rachel L. Lim, and Mark I. C. Chen

Table of Contents

Web Appendices 1–8.......................................................................................... 1–19
Web Tables 1–4.................................................................................................. 20–23
Web Figures 1–7............................................................................................... 24–30
Web Appendix 1

In this section, we deal with how a structural equation model (SEM) can be used to estimate the titer-mediated, titer-independent and combined effect of a vaccine (which correspond to the indirect, direct and total effects of a particular factor in the terminology of structural equation models), and potentially combine information from various sources of data. The motivation for the use of such methods would be to estimate the effect of correlates of protection for a given vaccine. This may be particularly necessary in the setting of an influenza pandemic where we would not have trials on outcomes for a vaccine (which may include imperfect vaccines that are available prior to the pandemic) until a sizeable number of pandemic infections have already occurred. Therefore, while the data presented here focus on antibody titers measured by hemagglutination inhibition (HI) assays, we wish to point out that the methods presented here can potentially be extended to other correlates of protection besides antibody titers in mediating protection. The methods can use the observational datasets we present in the main manuscript, vaccine trial data if available, or potentially even mix and match vaccine trial data with different kinds of observational data.

In the Web Material that follows, we will focus mainly on the ability of SEM to adequately predict a titer-mediated effect for a vaccine while disentangling potential confounding by age. We do so by firstly presenting the results from the use of our SEM approach to model titer distributions. We then show a set of results from our real dataset, then move on to results from a series of experiments on synthetic data to show when the model accurately estimates the underlying parameters used to generate the synthetic data, and when it is less accurate. Finally, we show an example of how data from a totally separate study
might be useful in providing estimates of titer-mediated effects in the context of an influenza pandemic. The synthetic data used in this study can be made available from the corresponding author upon request.
Web Appendix 2

As discussed in the main text, since HI titers mediate protection, and a greater proportion of vaccinated individuals have higher titers, then vaccinated individuals may experience a risk reduction for infection relative to non-vaccinated individuals mediated by HI titers (Web Figure 1). To investigate this, we used the built-in functions of the Generalized Structural Equation Modeling (GSEM) package in Stata 15 (STATA Corp.; College Station, USA), with the model structure depicted in Web Figure 2. As explained in the main text, the SEM comprised two halves. Firstly, as titers had a right-skewed distribution, we grouped titers into 3 categories (<10, 10 and ≥20), and estimated the vaccine’s association with titers using ordinal regression ($\beta^T_V$). In Web Table 1, Equation 1 shows how the coefficients from the ordinal regression model can be used to then predict the proportions in each titer category by vaccination history. The second half of the SEM investigated the relationship of titers (titers 10 and ≥20 versus <10 as $\beta^S_T$ and $\beta^S_T$ respectively) and vaccination ($\beta^S_V$) on the risk of seroconversion through failure time regression, with time to seroconversion as defined in the main manuscript. Equation 2 shows the underlying formula for the concept shown in Web Figure 1, where the risk in the non-vaccinated ($r^M_U$) and vaccinated groups ($r^M_U$) is in part mediated by their respective distribution of HI titers. The vaccinated group would have more individuals in the higher titer categories (10 and ≥20) than the non-vaccinated group, and the risk of infection in those with higher titer categories is potentially lower than in the reference category. We can multiply the proportions in the titer categories with the corresponding relative risk (i.e. $exp(\beta^S_T)$ and $exp(\beta^S_T)$), and then by the risk in the reference titer category (<10, $exp(D)$ where D is a constant derived from the failure time regression
model). Summing the respective risks by the titer categories in the non-vaccinated and vaccinated groups then gives $r_U^M$ and $r_V^M$ respectively, and the ratio of the latter over the former (Equation 2c) is then the risk reduction from the vaccine mediated by titers. The titer-independent effect of the vaccine on risk reduction, $exp(\beta_V^S)$, is estimated directly in the failure time regression model which simultaneously adjusts for the effect of titers on risk of infection (Equation 3a). The combined effect is then simply the product of the titer-independent risk reduction and the titer-mediated risk reduction (Equation 3b). The coefficients were computed following implementation of the GSEM model using the “nlcom” command in Stata (see command files attachment, “sem_vaccine_data.xlsx”, “Commands, real data” worksheet for commands, and “real_data” worksheet for the actual data.)
Web Appendix 3

The synthetic data modeled a simple situation with just two age groups, fixing the two groups to have different levels of vaccination (as shown in Web Table 2). Since a key limiting factor in our real data was the number of adults in our oldest birth cohort, we make reference to the sample size of the older age group when describing the results from our synthetic datasets, and fixed the default sample size at 50 older adults and 2,450 younger adults. We used the distribution of HI titers in the oldest birth cohort versus that for all the other groups as a guide for the distribution of HI titers at the levels in Web Table 2. Also, the association between titer categories and reduced risk of seroconversion approximated those from Table 2 in the main manuscript, as did the resulting size of the titer-mediated effect (from Figure 3 in the main manuscript). This can be calculated using the actual synthesized proportions in the various titer categories (rather than the predicted proportions depicted in equation 1), and the relative risk parameters for the effect of titers in Web Table 2.
Web Appendix 4

Based on the dataset included with the supplementary material (sem_vaccine_data.xlsx; "synthetic_data worksheet"), we can simulate each replicate in two steps using two numbers per replicate, which are randomly generated with a uniform distribution from 0 to 1:

- a titer distribution given the probabilities in Web Table 2 by comparing the first random number to the cumulative probability of being in each titer category given the age and the vaccination status of the individual.
- a probability of infection given randomly generated titer, the product of the baseline risk for that age group, the relative risk from the effect of the titer, and the relative risk from the titer-independent effect of the vaccine.

We then used the same commands as for the real data to perform an SEM, and estimate the titer-unadjusted, titer-independent, titer-mediated effects, as well as the combined effect and the proportion of the combined effect due to titer-mediated effects.

A set of sample commands for the first experiment (used to generate Web Figure 5) is also provided in the same file ("Sample program, synthetic data" worksheet).

As shown in Web Figure 3, an ordinal logistic model can adequately reproduce the distribution of the hemagglutination inhibition titer categories in the real data. However, the confidence intervals are wide for some of the older birth cohorts in particular (due to the small number of observations) and there is some deviation between the modeled and observed titers in the vaccinated group for the 1967–1976 birth cohort. The same ordinal logistic model was also similarly able to give a reasonable fit to the synthetic distribution of HI titers generated by the parameters in Web Table 2.
Web Appendix 5

Web Figure 4B shows the results for the titer-mediated effect using birth cohort stratified titer-independent effects (i.e. by including an interaction term for birth cohort and the effect of the vaccine). There is separation for the birth cohort 1927–1936, given no observations with infections amongst the small number of non-vaccinated participants (n=24) in that birth cohort stratum, so the estimates for this group are omitted. Figure 4 from the main manuscript, where we have the estimates for the titer-mediated effect using a global estimate of the vaccine’s titer-independent effect without stratifying by birth cohort, is shown as Web Figure 4A for comparison. The estimates for the titer-independent effects in Web Figure 4B are extremely wide. Consequently, the estimate for the combined effect (green triangles in both panels) are far wider in Web Figure 4B than in Web Figure 4A.
Web Appendix 6

We next performed a set of experiments with synthetic data simulated using the parameters in Web Table 2 to explore the degree to which sample size is responsible for the wide uncertainty in the estimates. Web Figure 5 simulates scenarios where the effects are either entirely mediated by titers or independent of the titers, and explored how sample size affects our ability to measure effect sizes for the vaccine of similar magnitude to those estimated for the oldest birth cohort (born from 1911 – 1926). When synthesizing datasets where the vaccine had no effect independent of antibody titers (Web Figure 5A), we indeed estimated a coefficient which varied symmetrically around zero. Even for a sample size of only 50 older adults (like what was available in our study), the interquartile range for the coefficient (round markers with error bars) was restricted to between $-0.07$ to $0.06$, with little chance of a significant result in either direction (<2%, orange bars). Moreover, the estimated titer-mediated effect (square markers) were of the same order as what was simulated (coefficient of $-0.23$, or relative risk of 0.79), with the interquartile range for our estimates again being fairly narrow ($-0.32$ to $-0.18$), and in ~70% of the replicates we detected a significantly protective effect (green bars). The point estimate for the combined effect (triangular markers) also coincided with what was simulated in the synthetic data. However, the range of estimates was wider ($-0.35$ to $-0.15$), since this compounds both the uncertainty contributed by the titer-independent and titer-mediated effects. Consequently, a significantly protective effect was detected in only 38% of replicates (blue bars). In contrast, the version of the GSEM that was unadjusted for titers (diamond markers and purple bars) performed poorly, with the median estimate of effect size deviating substantially from what was simulated in the underlying data when the
sample had only 50 older adults. With a greater number of older adults (100 or more), median estimates for the effect of the vaccine coincided with what was simulated in the synthetic data but the interquartile range remained wide. This may have reflected an inadequacy of the model to accurately discern interactions between the effects of age and vaccine on protection, even at relatively large samples sizes.

Conversely, Web Figure 5B shows the results from a synthetic dataset where the protective effect equivalent to a relative risk of about 0.8 arises entirely independent of the HI titers, and occurs in both the younger and older adults. We would have a fair chance of detecting this effect (with 55% of replicates having $P$ values <0.05) in a version of the GSEM that simultaneously adjusts for the effects of age, vaccination status and antibody titers. More importantly, the model correctly estimates that the effects were not mediated by the titers in almost all replicates, with less than 1% of replicates having a coefficient for the titer-mediated effect in the older age group that was significant at $P$<0.05. Estimates for the combined effect symmetrically span the level simulated (interquartile range from −0.29 to −0.16, for a sample size of 50 older adults). Again, the version of the model which includes a term for interaction between age and vaccine unadjusted for HI titers resulted in inaccurate and wide-ranging estimates (diamond markers and purple bars), particularly at smaller sample sizes.

In Web Figure 6, we fixed the sample size of the synthetic data at 50 older adults and 2,450 younger adults and then varied the titer-independent effect of the vaccine through 4 levels (from coefficient of 0.00 to −0.36, corresponding to relative risks of 1.0, 0.9, 0.8 and 0.7 respectively). For each synthetic dataset, both the older and younger age groups were simulated to have the same titer-independent effect, so that any differences in the
effect of the vaccine by age should arise only through the titer-mediated effects due the age-related differences in the association of the vaccine with HI titers. Web Figure 6A shows that, regardless of the level of the titer-independent effect, median estimates for the titer-mediated effect for the vaccine are consistently near the level simulated in the synthetic data, and 90% of the estimates would fall below zero (indicating a protective effect). Between immediately adjacent levels that were simulated (e.g. ind-RR=1.0 and ind-RR=0.9), the 90th percentile of estimates for the titer-independent effects overlap substantially, but the GSEM would in most replicates distinguish a scenario with $\beta_{0,v} = \beta_{1,v} = -0.36$ (or a relative risk of 0.7) from the lack of an effect (i.e. $\beta_{0,v} = \beta_{1,v} = 0.00$). This is also shown in Web Figure 6B where close to 90% of replicates $\beta_{0,v} = \beta_{1,v} = -0.36$ give an estimate for the titer-independent effect (round markers with error bars) that is significantly less than zero (orange bar). On the other hand, Web Figure 6B also shows that regardless of the size of the titer-independent effect, the model would deduce a significantly protective titer-mediated effect in about 70% of replicates. It is harder to detect a titer-independent effect of equivalent magnitude to the titer-mediated effect. For instance, at $\beta_{0,v} = \beta_{1,v} = -0.22$ (or relative risk of 0.8 for the titer-independent effect), both the titer-independent and titer-mediated effects are of nearly similar magnitude. However, a smaller proportion of replicates are significant at $P<0.05$ for the former than the latter (orange bars vs green bars), with only 50% of replicates showing a protective titer-independent effect significant at $P<0.05$ vs 70% for the titer-mediated effect. Unsurprisingly, a combined effect is increasingly detected with greater ease (i.e. more replicates show significant protection at $P<0.05$) as we increase the level of titer-independent protection.
In Web Figure 7, we investigate if the model is able to detect interactions between the age group and the titer-independent effect of the vaccine in protecting against seroconversion. We assume the same 4 levels of titer-independent effects, with this occurring only in the older age group; the vaccine was assumed to have no titer-independent effect in the younger age group. Using a synthetic data set with the default sample size of 50 older and 2,450 younger adults, Web Figure 7A charts the distribution of \( P \) values from testing for an interaction term between age group and the effect of the vaccine. Even if the vaccine had a substantial titer-independent effect in the older age group exceeding the level that was titer-mediated (\( \beta_{0,v} = -0.36 \), or relative risk of 0.7), significance testing comparing a model with and without the interaction term would only give \( P \) values <0.05 and <0.10 in about 7% and 16% of replicates respectively. For such a sample size, the results in a model that adjust for titers are not substantially different in a model that does not adjust for the titers. In fact omitting the term for titers results in less replicates having \( P \) values <0.05 (though more have \( P \) values <0.10). Furthermore, Web Figure 7B shows that including an interaction term to model a difference in the titer-independent effect of the vaccine by age results in wide ranges for estimates of the titer-independent effect (filled round marker), even at \( \beta_{0,v} = -0.36 \) when it would be on the average stronger than the titer-mediated effect in the underlying synthetic data. Consequently, estimates for the combined effect (filled triangular marker) would also vary widely. It is also not possible to accurately determine the net difference in effect by age by including an interaction term without adjusting for the HI titers, as the resulting estimates (filled diamond marker) also vary widely, with none of the tests for interaction being significant at \( P < 0.05 \) (blue bars of lighter shade). On the other hand, omitting the
term to model an interaction between vaccine and age results in estimates of the titer-independent effect being close to zero (unfilled round marker), since only a small proportion (2%) of the synthetic sample are in the older age. However, were a much larger sample size available (1,000 older and 49,000 younger adults), we would detect interactions between vaccine and age significant at $P<0.05$ in ~60% and at $P<0.10$ in ~70% of replicates for $\beta_{0,v} = -0.36$. This is in a model that does not adjust for the HI titers (Web Figure 7C, last stacked bar). However, it should be noted that this is for a combined effect size (filled triangular marker, Web Figure 7D) from the titer-independent and titer-mediated pathways summing to a coefficient of $-0.6$ (or relative risk equivalent of about 0.55).

Web Figure 7 also emphasizes the difficulty of directly detecting age-related differences in protection that arise only through the titer-mediated pathway (from age-related differences of the vaccine in inducing titers). In the scenario with no titer-independent effect (i.e. $\beta_{0,v} = 0.00$), we would detect interactions between vaccine and age significant at $P<0.10$ in <20% of replicates when not adjusting for the HI titers, even for the extremely large sample size simulated in Web Figure 7C. Web Figures 7A and C also shows that we are unlikely to conclude that there are age-related differences in titer-independent effects when none were simulated (i.e. $\beta_{0,v} = \beta_{1,v} = 0.00$), with only a small minority of replicates significant at $P<0.05$ in when testing for interaction at the two extremes of sample sizes. Web Figures 7B and 7D also show that, regardless of the sample size of the synthetic dataset, the estimate of the titer-independent effect in the older age group would have been centered around zero when no titer-independent effects were simulated (filled round marker for ind-RR=1.0). While the interquartile range was extremely wide,
very few replicates gave an estimated coefficient that was significantly above or below zero even when using the much larger sample size (orange bars of lighter shade in Web Figure 7D).
Web Appendix 7

We also explored if our estimates for the association of titers with seroconversion (which is being used as an imperfect proxy for infection) was reflective of the protection conferred by having different levels of antibodies detectable by hemagglutination inhibition assays. For this, we retrieved data for 88 cases (1) which had been confirmed to have influenza A(H1N1)pdm09 infection by reverse-transcriptase polymerase chain reaction (RT-PCR), and which also had HI assays taken before day 5 of symptoms (i.e. before the rise in antibodies generated by the infection). This was then used as an alternative to seroconversion events to define infections. Web Table 3 compares the profile of these cases against characteristics of cohort participants aged ≤60 years. The cases were less likely to be female, and mostly young adults aged <30 years. The distribution of antibody titers also differed significantly, with titers ≥20 being under-represented amongst cases, whether compared to all cohort participants or just those from the community. The cases were also different in terms of their characteristics and titer distributions from a potential set of controls, which we defined as those in the community cohort who neither seroconverted nor displayed any acute symptoms suggestive of an acute respiratory infection during the fortnightly telephone interview during our study period (2).

A case control analyses confirmed that the titers were associated with protection (Web Table 4). This association persisted even after adjusting for differences in age and gender, with the associations for the titer levels modelled being somewhat stronger than those from the cohort study. However, caution is needed for the estimated effect of having titers ≥20, since this was based on very small numbers (with only a single case having titers ≥20). However, using titers ≥10 vs titers <10 in the multivariable analyses for the case
control analyses also gave a significant odds ratio for the effect of titers (OR = 0.59, 95% CI 0.43, 0.79, $P=0.001$). Repeating the analyses in the main manuscript with titers $\geq 10$ vs titers $<10$ (in those aged $\leq 60$ years) gave a very similar corresponding relative risk of 0.59 (95% CI 0.45, 0.78, $P=0.001$)).
Web Appendix 8

Several insights arise from the additional analyses presented in our supplementary material. Firstly, we show that ordinal logistic regression is an appropriate way of modeling the skewed distribution of antibody titers observed in our study population. Alternative ways to predict the proportions in each titer category include multinomial regression, as well as assuming the proportion in each successive titer category follows a logistic function in the method proposed by Siber et al (3). Vaccine associated differences in the predicted proportions in various titer categories from the ordinal logistic regression could then be used to estimate the titer-mediated effects of the vaccine. As performed for the analyses in our main manuscript, titers could be modeled as a categorical variable in the second half of the structural equation model to investigate the association between titer categories and seroconversion (or infection). Predicted proportions in each titer category can in turn be also converted into the predicted difference in means between vaccinated and unvaccinated individuals, as an alternative input variable for structural equation modeling of the titer-mediated effect. However, this implicitly assumes that the titers have a near-linear effect on risk of infection, which may not always be true.

Secondly, the first set of experiments with our synthetic data shows that our sample size was likely adequate when using SEM to detect predicted titer-mediated effects of the magnitude that we estimated (Web Figure 5A). Conversely, in scenarios where we simulated no titer-mediated effects, the SEM approach also did not conclude there was any such effect in the vast majority of replicates. However, due to the uncertainty contributed by estimates of the titer-independent vaccine effect, the overall range of
estimates for the combined effect was wider, and may not reach the arbitrary cut-off value of $P<0.05$ for significance given our study’s sample size. In contrast, the SEM performed less well in generating reliable estimates for titer-independent effects of similar magnitude (Web Figure 5B).

Thirdly, the next two sets of experiments with synthetic data, which varied the level of titer-independent effects, shows that these could be detected when they are consistent across the age group (Web Figure 6), but are difficult to detect if it occurs only in the older age group (Web Figure 7). Generally, the titer-independent effects would require larger sample sizes to estimate accurately than titer-mediated effects of similar magnitude. Also, Web Figures 6 and 7 emphasize that the SEM approach could still detect the simulated titer-mediated effect with reasonable accuracy regardless of the titer-independent effects assumed, which is unsurprising (for reasons we will discuss below). Additional conclusions from the experiments underlying Web Figure 7 are that, at the effect sizes simulated, it is unlikely that we can directly observe age-related differences in vaccine effectiveness. In particular, the titer-unadjusted effect essentially reflects what observational studies of vaccine effectiveness (which would not have data on titers) may find. Conventional ways of testing for interaction in such circumstances would be unlikely to give $P$ values $<0.10$ for either age-stratified titer-unadjusted or titer-independent effects, at least not for the given sample size in the oldest birth cohort in our study. Correspondingly, estimates on the size of such effects would likely have wide confidence intervals.

In general, our conclusions are that the estimates of the age-stratified titer-mediated effects can be still be reliable for an imperfect vaccine with the effect sizes that we
simulated, even for the small numbers in the oldest birth cohort of our study. On the other hand, estimates for age-related differences in titer-independent effects would be much less reliable. Attempts to estimate age-related differences for the effect of the vaccine without incorporating information about antibody titers (i.e. titer-unadjusted) would not help the matter and in our synthetic data could not yield conclusive estimates.

Close examination of the underlying formulas for estimating the titer-mediated effect and our experiments with synthetic datasets give us some reasons why the above is so. The predicted effect from the titer-mediated pathway is based on combining the estimates from the first half of the SEM equation (i.e. the ordinal logistic model on the effect of the vaccine on titers) with the effect of titers on risk of seroconversion from the second half of the SEM equation. The estimate for the association of the vaccine with titers was helped by having about equal numbers of vaccinated and unvaccinated individuals in the older age groups in our study. Moreover, the estimates on the association of titers with the risk of seroconversion were not stratified by age, and were hence drawing on the entire sample size (including those in younger birth cohorts). In contrast, estimates of the titer-independent effect of the vaccine is dependent on the proportion who had seroconversion (<20% of the study population). The number of seroconversion events is particularly small when stratifying by age and vaccination status, the issue being further compounded by the lower risk of infection in older age groups.

Predictions from the SEM that the vaccine conferred some protection in the older birth cohorts is hence predicated on two key factors. Firstly, the vaccine has no titer-independent effect on increasing the risk of infection (rather than decreasing the risk of infection). For the reasons mentioned, and as shown through the synthetic data
experiments, the titer-independent effects are difficult to estimate accurately, and age-stratified estimates would be unreliable. However, at least at the level of the full study population, our data shows that there is no strong evidence that the vaccine increases or decreases the risk of seroconversion. The other key assumption then is that antibody titers are associated with protection and have a similar effect across the age groups. Admittedly, sample sizes would not permit stratified analyses for titers by age. Also, given the outcome of interest used for the analyses presented in our main manuscript was based on seroconversion, there are concerns that these findings are biased due to the ceiling effect (i.e. that those with higher baseline titers are less likely to seroconvert upon infection) (4). However, as shown in Web Table 3 and Web Table 4, where we present results from a group of 88 cases that were independently confirmed to have influenza A(H1N1)pdm09 on RT-PCR assays, these cases had a distribution of titers on HI assays that was different from the population, being significantly less likely to have antibodies at higher titers (≥1:20). Since only one case had titers ≥1:20, the strength of the association could not be accurately estimated, and we did not proceed to use these parameters to predict the titer-mediated effect. However, were the point estimates in the range of what the case control study suggests, then the relative risk resulting from the titer-mediated effect alone would be about 0.6 for the oldest birth cohort.

The possibility of combining data from such a case control study with data on the association of the vaccine with titers raises an intriguing prospect of how SEM could be applied in a pandemic situation. The relationship between the seasonal influenza vaccine on titers to a pandemic virus can be established with banked sera from past vaccine trials as was done by Hancock et al (5). In turn, antibody titers to a pandemic strain could be
simultaneously measured in infected individuals presenting early in the pandemic, as well as an appropriate set of controls or samples from the general population, matched by age and gender (or if not matched, then adjusted for statistically). The difference in the distribution of antibody titers between the cases and the controls (or general population) would then provide some idea on how titers potentially correlate with protection, and therefore allow us to estimate the titer-mediated effect of existing vaccines.
### Web Table 1: Equations used to predict distribution of titers, the titer-mediated effect, titer-independent effect and the combined effect of the vaccine

| Equation | Description / Formula |
|----------|-----------------------|
| 1        | Distribution of titers by vaccination history \( (x=0 \text{ for non-vaccinated and } x=1 \text{ for vaccinated}) \) |
| 1a       | \( P_0 = \left( 1 + \exp(x.\beta_V^T-C_1) \right)^{-1} \) |
| 1b       | \( P_1 = \left( 1 + \exp(x.\beta_V^T-C_2) \right)^{-1} - P_0 \) |
| 1c       | \( P_2 = 1 - P_1 - P_0 \) |
|          | where \( P_0, P_1 \) and \( P_2 \) are proportions in titer categories <10, 10 and \( \geq 20 \) respectively, and \( C_1 \) and \( C_2 \) are constants for cut-points for the titer categories as estimated from the ordinal logistic regression. |
| 2        | Risk by vaccination history due titer-mediated effect and relative risk in vaccinated vs unvaccinated |
| 2a       | \( r_0^M = \exp(D). \left( P_0^U + \exp(\beta_S^U) \right) \left( P_1^U + \exp(\beta_S^U) \right) \) |
| 2b       | \( r_1^V = \exp(D). \left( P_0^V + \exp(\beta_S^U) \right) \left( P_1^V + \exp(\beta_S^U) \right) \) |
| 2c       | \( RR^M = r_0^M / r_1^V \) |
|          | where \( r_0^M \) and \( r_1^V \) sum the risk of infection in various titer categories for non-vaccinated and vaccinated groups respectively; \( P_0^U, P_1^U \) and \( P_0^V, P_1^V \) are proportions in corresponding titer categories in non-vaccinated, and likewise \( P_0^U, P_1^U \) and \( P_0^V, P_1^V \) in the vaccinated group; \( D \) is the constant from exponential failure time regression and cancels out when estimating \( RR^M \), the risk reduction due to the titer-mediated effect. |
| 3        | Combined effect from titer-mediated and titer-independent effects |
| 3a       | \( RR^I = \exp(\beta_S) \) |
| 3b       | \( RR^C = RR^I \cdot RR^M \) |
|          | where \( RR^I \) is risk reduction from the titer-independent effect of the vaccine and \( RR^C \) is the combined effect; \( \ln(RR^M) \) and \( \ln(RR^C) \) converts the relative risk into the coefficients presented in Figure 3 of the main manuscript. |

\( RR^M \): Titer-mediated effect on risk reduction.  
\( RR^I \): Titer-independent effect on risk reduction.  
\( RR^C \): Combined effect on risk reduction.
| Parameter                                           | Age Group |           |
|-----------------------------------------------------|-----------|-----------|
|                                                     | 0 = Older | 1 = Younger |
| Proportion of Total Samples                         | 0.02      | 0.98      |
| Proportion Vaccinated                               | 0.60      | 0.30      |
| Base Risk of Infection*                             | 0.10      | 0.20      |
| Distribution of Titers in the Non-Vaccinated        |           |           |
| 0 = <10                                             | 0.80      | 0.85      |
| 1 = 10                                              | 0.05      | 0.05      |
| 2 = ≥20                                             | 0.15      | 0.10      |
| Distribution of Titers in the Vaccinated            |           |           |
| 0 = <10                                             | 0.45      | 0.75      |
| 1 = 10                                              | 0.15      | 0.10      |
| 2 = ≥20                                             | 0.40      | 0.15      |
| Effect of Titers (vs <10)                           |           |           |
| 1 = 10                                              | Default RR = 0.90; RR = 1.0 for Web Figure 5B |
| 2 = ≥20                                             | Default RR = 0.30; RR = 1.0 for Web Figure 5B |
| Titer-Independent Effect of Vaccine                 |           |           |
| Web Figure 5A                                       | RR\text{I} = 1.00 | RR\text{I} = 1.00 |
| Web Figure 5B                                       | RR\text{I} = 0.79 | RR\text{I} = 0.79 |
| Web Figure 6A and 6B                                | RR\text{I} = 1.0, 0.90, 0.80 and 0.70 |
| Web Figure 7A to 7D                                 | RR\text{I} = 1.0, 0.90, 0.80 and 0.70 | RR\text{I} = 1.0 |

RR: Risk reduction.
RR\text{I}: Titer-independent effect on risk reduction.
*This represents participants who were not vaccinated and had samples with HI titers of 0 against all viruses.
Web Table 3: PCR confirmed cases versus cohort participants aged ≤60 years of age with valid samples

| Characteristic          | Cases, n=88 | All Cohort Participants, n=2203 | Community Cohort Participants, n=737 | Controls†, n=314 |
|------------------------|-------------|---------------------------------|--------------------------------------|------------------|
| Female Gender          | 23 (26.1)   | 651 (29.6)                      | 737 (59.2)                           | 195 (62.1)       |
| Age Group              |             |                                 |                                      |                  |
| 17 to 30               | 55 (62.5)   | 1368 (62.1)                     | 149 (20.2)                           | 40 (12.7)        |
| 30 to 39               | 8 (9.1)     | 261 (11.9)                      | 141 (19.1)                           | 54 (17.2)        |
| 40 to 49               | 12 (13.6)   | 352 (16.0)                      | 290 (39.4)                           | 135 (43.0)       |
| 50 to 60               | 13 (14.8)   | 222 (10.1)                      | 157 (21.3)                           | 85 (27.1)        |
| Titers‡                |             |                                 |                                      |                  |
| <10                    | 80 (90.9)   | 1792 (81.3)                     | 648 (87.9)                           | 262 (83.4)       |
| 10                     | 7 (8.0)     | 153 (7.0)                       | 42 (5.7)                             | 25 (8.0)         |
| ≥20                    | 1 (1.1)     | 258 (11.7)                      | 47 (6.4)                             | 27 (8.6)         |
| P value*               | 0.001       | 0.079                           | 0.034                                |                  |

Numbers in brackets are column percentages.
†Controls were community cohort participants which neither seroconverted nor had acute respiratory infection symptoms.
‡Titers in samples taken before day 5 of symptoms in cases; in both cases and cohort participants, titers to A/California/7/2009(H1N1)pdm09 were measured by hemagglutination inhibition assays.
*P value by chi-squared test comparing cases and that grouping of cohort participants.
**Web Table 4: Case control study on effect of titers on risk of infection in individuals aged ≤60 years of age**

| Characteristics | Case Control Study | Failure Time Regression Model in Main Manuscript (Figure 1) |
|-----------------|--------------------|----------------------------------------------------------|
|                 | Univariate Analysis† | Multivariable Analysis† | OR (95% CI) | P value | OR (95% CI) | P value | RR (95% CI) | P value |
| Female Gender   | 0.22 (0.13, 0.37) | <0.001 | 0.29 (0.16, 0.53) | <0.001 | 0.90 (0.62, 1.31) | 0.589 |
| Age Group (vs 17 to 30 years) | | | | | | |
| 30 to 39        | 0.11 (0.05, 0.25) | <0.001 | 0.11 (0.04, 0.26) | <0.001 | 0.90 (0.62, 1.31) | 0.589 |
| 40 to 49        | 0.06 (0.03, 0.13) | <0.001 | 0.07 (0.03, 0.14) | <0.001 | N.A.* |
| 50 to 60        | 0.11 (0.05, 0.22) | <0.001 | 0.11 (0.05, 0.24) | 0.015 |
| Titers (vs <10) | | | | | | |
| 10              | 0.92 (0.38, 2.20) | 0.846 | 0.83 (0.29, 2.36) | 0.731 | 0.89 (0.62, 1.28) | 0.521 |
| ≥20             | 0.12 (0.02, 0.91) | 0.040 | 0.05 (0.01, 0.39) | 0.004 | 0.42 (0.29, 0.60) | <0.001 |

N.A.: Not applicable.

RR: Risk reduction.

OR: Odds ratio.

†Controls were community cohort participants which neither seroconverted nor had acute respiratory infection symptoms.

*Not Applicable as birth cohorts were used.
Web Figure 1. Framework for estimating titer-mediated and combined effects of the vaccine on risk of seroconversion. The first and last stacked bars give the distribution of antibody titers in the vaccinated and unvaccinated cohort respectively, while the colors correspond to the titer category, with blue being titer <10, green being titer = 10 and yellow being titer ≥20. Section (A) depicts the titer-mediated effect of vaccination in individuals on risk of infection, and (B) depicts the combined effect of vaccination (titer-mediated and titer-independent) on risk of infection, while (C) depicts the baseline titer-mediated effect on risk of infection in the absence of vaccination. The titer-mediated and titer-independent effect on risk reduction are labelled $RR^m$ and $RR^i$ respectively, with combined effect denoted as $RR^c$. 
Web Figure 2: Path diagram for structural equation model (SEM). The first half of the SEM estimates the relationship of the factors represented with the distribution of hemagglutination inhibition assay titers using ordinal logistic regression, while the second half estimates the relationship with seroconversion for the same variables plus the titers (as a categorical variable) using exponential failure time regression. Arrows denote paths within the SEM with double lines indicating a multichotomous factor. The pathways for the effect of the vaccine on seroconversion is shown using red lines, with the undashed and dashed arrows denoting the titer-independent and titer-mediated effects respectively. *Subgroup identifiers for the separate military units and long-term facilities; the circles with double lines represent the latent variables used to separately model potential clustering of observations by subgroup for first half of the model on titers, and the second half on time to seroconversion. **LTCF: Long-term care facility.
Web Figure 3: Comparison of observed (colored bars) and predicted (points) distribution of titers from the ordinal logistic portion of the GSEM, stratified by influenza vaccination in the last year (as specified in legend below panels) and birth cohort (panels A to I).
Web Figure 4: Age-stratified association of the vaccine with the risk of infection, with titer-unadjusted (blue diamonds), titer-mediated (red circles), combined relative risks (green triangles), and titer-independent effects (purple squares). The blue bar shows the proportional contribution of the titer-mediated effect to the combined effect. Error bars denote the 95% confidence intervals. UL: Upper limit. LL: Lower limit. In (A), the combined effects are estimated using unstratified estimates of the titer-independent effect, while in (B) the combined effects use the birth cohort stratified estimates of the titer-independent effect. Due to the coefficients for the combined effect being positive (i.e. having RR>1.0) in some birth cohorts, the proportional contribution of the titer-mediated effects is a negative number and hence not shown. Also, the estimates for 1927–1936 could not be estimated as there were no observations with infections amongst the small number of non-vaccinated participants (n=24) in that stratum.
Web Figure 5: Effect of sample size on estimates of the titer-independent, titer-mediated, combined and titer-unadjusted effect of the vaccine. In (A), the synthetic data simulates a titer-mediated vaccine effect which protects against seroconversion, with a coefficient of −0.23 (i.e. relative risk 0.79) in the older age group. In (B), there is no titer-mediated effect but the vaccine has a similar independent effect in both the older and younger age groups with a coefficient of −0.23. Round, square and triangular markers are coefficients (with error bars denoting the interquartile range) for titer-independent, titer-mediated and combined effects respectively in a model that includes age, vaccine and hemagglutination inhibition titers, while diamond markers give coefficients for the vaccine effect unadjusted for titers in a model that includes age and vaccine plus the interaction between these two variables. Orange, green and blue bars give the corresponding proportion of replicates where the estimates have a $P$ value of <0.05. UL: Upper limit.
Web Figure 6: Effect of varying levels of titer-independent vaccine effect on the estimates of the titer-independent, titer-mediated, combined effect of the vaccine. In all scenarios, the synthetic data sample size is kept constant at 50 older and 2,450 younger adults, and the titer-mediated effect is kept constant at a coefficient of −0.23 (i.e. relative risk of ~0.79). The titer-independent vaccine effect is assumed to be the same in both older and younger adults. (A) presents the joint estimates for different levels of the titer-independent vaccine effect expressed as a relative risk (RR) [i.e. ind-RR, for values 1.0, 0.9, 0.8 and 0.7 as green, orange, blue and red respectively]. Unfilled markers and filled markers of darker shades respectively denote results where titer-independent and titer-mediated effects are significant at P<0.05, while non-significant results are in lighter shades. B) gives the estimated coefficients for the 4 levels of titer-independent effects, with the proportion that is titer-mediated is shown as text below each set. Round, square and triangular markers are coefficients (with error bars denoting the interquartile range) for titer-independent, titer-mediated and combined effects respectively while orange, green and blue bars give the corresponding proportion of replicates where the estimates have a P value of <0.05.
Web Figure 7: Ability to detect age-related differences for varying levels of titer-independent vaccine effect with two extreme sample sizes (50 older and 2,450 younger adults in A and B and 1,000 older and 49,000 younger adults in C and D). In all panels, the titer-mediated effect is constant in both older and younger adults (and equivalent to a coefficient of $-0.23$ or relative risk of $0.79$ in the older adults), but the titer-independent effect varies for older adults (and is constantly set to RR=1.0 for younger adults). We performed a test for interaction by assessing if inclusion of a term between age and the titer-independent effect of the vaccine significantly improved model fit. This was done for models which included terms for age, vaccine and hemagglutination inhibition titers, and models which were unadjusted for titers but included terms for age and vaccine. Panels A and B show the proportion of replicates with $P$ values <0.05 and <0.10 from the test for interaction, stratified by whether the models adjusted for hemagglutination inhibition titers (indicated on horizontal axes). Panels C and D shows the sets of estimated coefficients from models which exclude and include the interaction term between age and the titer-independent effect of the vaccine (unfilled and filled markers respectively, and also indicated on the horizontal axes). Markers show point estimates for coefficients in the older age group (with error bars for inter-quartile range) for the titer-independent, titer-mediated and combined effect of the vaccine in a model which adjusts for titers, and the net effect of the vaccine when unadjusted for titers (round, square, triangular and diamond markers respectively). The orange, red, green and blue bars give the corresponding proportion of replicates where the estimates have a $P$ value <0.05, with darker and lighter shades of the same color for models without and with interaction. Ind-RR: Titer-independent vaccine effect expressed as a relative risk.
References

1. Chen MI, Barr IG, Koh GCH, et al. Serological response in RT-PCR confirmed H1N1-2009 influenza a by hemagglutination inhibition and virus neutralization assays: an observational study. *PloS One*. 2010;5(8):e12474.

2. Jiang L, Lee VJ, Lim WY, et al. Performance of case definitions for influenza surveillance. *Eurosurveillance*. 2015;20(22):21145.

3. Siber GR, Chang I, Baker S, et al. Estimating the protective concentration of anti-pneumococcal capsular polysaccharide antibodies. *Vaccine*. 2007;25(19):3816–3826.

4. Petrie JG, Ohmit SE, Johnson E, et al. Efficacy studies of influenza vaccines: effect of end points used and characteristics of vaccine failures. *J. Infect. Dis*. 2011;203(9):1309–1315.

5. Hancock K, Veguilla V, Lu X, et al. Cross-reactive antibody responses to the 2009 pandemic H1N1 influenza virus. *N. Engl. J. Med*. 2009;361(20):1945–1952.