Statistical modelling of antibody data

Arseniy Khvorov, David J. Price, Annette Fox, Sheena G. Sullivan

March 10, 2020

1 Introduction

In the study of infectious diseases it is useful to have some measure of whether a person is infected, immune or susceptible to infection. This may be important, for example, to understand disease prevalence, population-level susceptibility or for evaluation of vaccines. However, direct measurement of immunity is often not possible and instead some *correlate* of protection is used. For viral infectious diseases, an oft-used correlate is the serum antibody titre, which provides a measure of the amount of antibody that recognizes a particular epitope.

Antibody titres have several limitations. Titres are the inverse of the greatest dilution of antibody that inhibits virus in serial dilutions, with higher values indicating greater inhibition. There is no true zero value, nor is there a true measure of the maximum value; titres can only be as low as the minimum starting concentration and as high as the maximum dilution. Furthermore, within dilution intervals, the true concentration that inhibits virus is unknown; only the upper and lower bound of each dilution is known, i.e. titres are interval-censored. In addition, antibody titres are merely a surrogate measure and their sensitivity and specificity may be imperfect, such that reduced titres may not always correspond to increased susceptibility.

For influenza, the haemagglutination inhibition (HI) antibody titre is an established correlate of protection. Indeed, the annual reformulation of influenza vaccines is partially dependent on demonstration that circulating viruses are no longer inhibited by vaccine-induced antibodies indicated by HI titres below a certain threshold [1]. And annual re-licensing of updated formulations is dependent on demonstrating that a vaccine induces HI titres above this same threshold [2]. The HI threshold commonly used is a titre of 1:40, thought correspond to a 50% reduction in risk. This figure is derived from cohort studies among vaccinated or infected individuals who have been followed for infection, and among whom the median titre associated with protection (no detected infection) is calculated [3, 4].

Several methods have been proposed for the analysis of antibody titre data and the calculation of protective thresholds. Each has its own set of assumptions that makes it more or less appropriate for the data being analysed. Here we will consider 3 models that are used in the literature: a Cox proportional hazards regression; a logistic regression; and a scaled logit model. Using simulations and data from two published studies, we discuss these models' assumptions, limitations and situations in which they may not be appropriate.

2 Study designs and motivating examples

Establishment of the threshold at 40 is based on human challenge studies from the 1960s in which volunteers were either randomised to received vaccine or not, or were challenged with virus [3]. Blood samples were collected pre-challenge. For vaccinated individuals, challenge occurred at least 14 days after vaccination. Nasal swabs taken 48-hours after challenge were used to determine infection by virus culture, and for unvaccinated volunteers challenged with live virus, infection was additionally indicated by pre-to-post-challenge sero-conversion (4-fold rise in titre). In both cases, the protective titre was estimated from the pre-challenge geometric mean titre for uninfected participants.

While challenge studies permit close observation of participant responses to infection under highly controlled conditions, the infectious dose administered may be unnaturally high. Several observational studies have been established to understand influenza transmission in more realistic conditions. In these studies, participants are determined to be at risk of infection because one or more of their close contacts (i.e. household members) has been identified as influenza-infected or because influenza has been known to be circulating. Infection may be determined by laboratory testing of respiratory samples, or, if unvaccinated, by sero-conversion. In this paper, we will illustrate our findings using two established household studies.

2.1 The Kiddivax study

The Kiddivax study was a randomized controlled trial undertaken in Hong Kong in 2009-2010 in which 796 children aged 6-17 years were randomized to receive inactivated vaccine. We direct the reader to [5] for full detail of the study but provide a brief outline here.

Blood samples were taken pre- and 1-month post-vaccination to estimate their vaccine-induced antibody titres. The children and their household contacts were followed for approximately 7 months for influenza-like illness. Symptoms reported by any household member prompted swabbing and influenza testing by PCR for all household members [5]. The protective titre was estimated using a Cox proportional hazards model where the outcome was time to infection or the end of the study [4].

2.2 The Ha Nam household study

The Ha Nam study has followed 270 households in Ha Nam Vietnam since 2007. We direct the reader to [6] for full description of the study but provide a brief outline here.

Sera are collected (bi)annually, at intervals spanning RT-PCR-confirmed transmission periods, determined by surveillance data. Households are followed weekly for influenza-like illness and all household members swabbed when any member shows symptoms. Unlike the Kiddivax study, where the time at risk can be measured, in the Ha Nam study and other household studies of natural infection, the true time at risk is unknown. These differences are depicted in Figure 1.

We will be using study data from 2009 to 2015 consisting of the subjects' pre-season HI titre measurements against H1N1pdm09 and H3N2 viruses as well as the subjects' infection status

with those viruses.

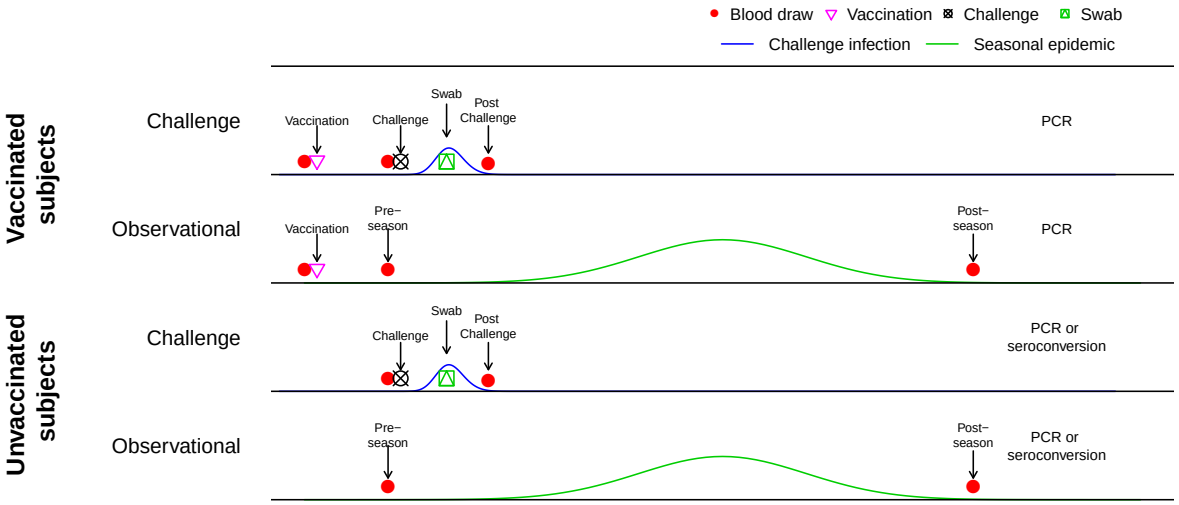

Figure 1: Example study designs for the estimation of protective titres.

3 Cox proportional hazards

The Cox PH is a common approach to analysing time-to-event data [7]. It assumes that the subjects being followed up are at risk of an event (e.g., death). This risk can change over time and it is called "hazard". The hazard then is a function of time which proportionately changes with covariates. For example, consider a study of how tumour size affects survival. For simplicity, say there are two sizes — large and small. It may be found that large tumours increase the risk of death by 2 as compared to small tumours. This would mean that the hazard for the large-tumour group is twice the hazard for the small-tumour group at any time.

For the Cox PH regression, the outcome is time-to-event. This time should normally be time at risk of an observable event. As an illustration, in Figure 2, two subjects are followed until both experience an event (e.g., infection).

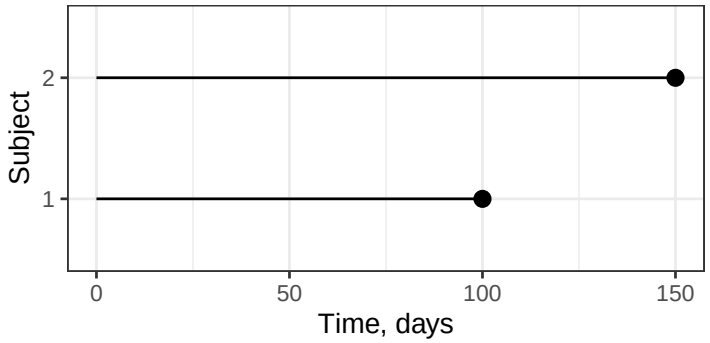

Figure 2: An example of time-to-event data. Two subjects are followed from time 0. Subject 1 experiences the even at time 100. Subject 2 experiences the event at time 150.

Since subject 2 experienced the event later (i.e., "survived" for longer), the covariate pattern (e.g., antibody titre) of subject 2 would be considered by the model to be more "protective" than that of subject 1. That is, their covariates would be consistent with a reduced hazard. For the purposes of this discussion, we will assume that a subject can only experience one event after which they are no longer at risk for the remainder of the follow-up time. For seasonal infectious diseases this may be a reasonable assumption if the follow-up is confined to one season and infection grants immunity for the rest of the season (or is fatal).

If subjects are not at risk for all of their follow-up time (e.g., not exposed to the virus), then the total follow-up time may be misleading as illustrated in Figure 3. Taking the actual time at risk into account would lead to the opposite conclusion in this example — it is subject 1 who is more "protected" since they were at risk for longer.

In infection data, true time at risk is unobservable because the risk of infection among susceptible individuals depends on exposure to the pathogen, which cannot be measured. If follow-up time is completely unrepresentative of true time at risk, the Cox model will not produce meaningful results. However, if the total time of follow-up is assumed to be, on average, proportional to the total time at risk (e.g., subjects who are followed up for longer can be expected to have been exposed for longer) then the situation illustrated in Figure 3 will be "averaged out" and the model may still produce meaningful results.

To investigate the behaviour of the estimates when time of follow-up is not the same as time

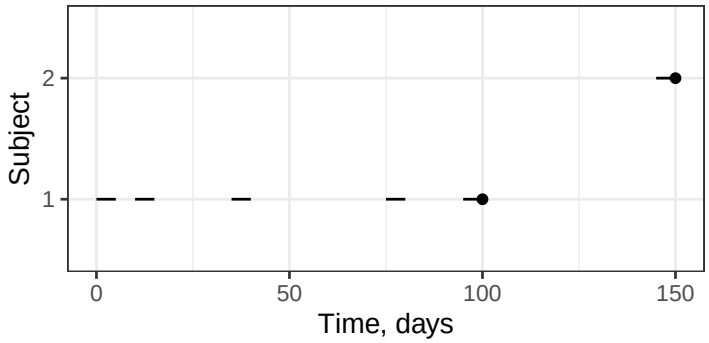

Figure 3: An example of time-to-event data. Two subjects are followed from time 0. Subject 1 experiences the even at time 100. Subject 2 experiences the event at time 150. Subject 1 was at risk of the event for a total of 5 days. Subject 2 was at risk of the event for a total of 25 days.

at risk but is proportional to it, we generated a simulated dataset, based on the following model:

$$T \sim \text{Exponential}(\text{rate} = \lambda)$$
$$h(t) = \lambda$$
$$\log\lambda = -3 - 1.5X_{\text{logtitre}}$$

Where T is the survival time, h is the hazard function and X_{logtitre} is the true log-titre measurement simulated from $N(2, 2^2)$.

Maximum time of follow-up was set to 100 days. Each individual was assigned a proportion of the follow-up time they were exposed to the virus (time-at-risk proportion). This proportion was generated from Beta$(\mu\kappa, (1-\mu)\kappa)$ where μ is the expected proportion for the population and κ is inversely proportional to the variance of the proportion. Smaller κ values result in larger heterogeneity of exposure in the population. When $\mu = 1$, the proportion assigned was always 1 to represent the ideal context of all follow-up time being time at risk. An individual's follow-up time was 100 times their time-at-risk proportion.

A random number generated for each individual by Exponential$(\text{rate} = \exp(-3 - 1.5X_{\text{logtitre}}))$ represented the time it would take that individual to experience the event (get infected), that is "survival" time. If survival time was greater than time at risk, the individual was "not infected" and their follow-up time was 100 (i.e., this was a right-censored observation). If survival time was smaller than time at risk, the individual was "infected" and their follow-up time was their survival time divided by their time-at-risk proportion.

Each simulation included 10,000 observations and was run 10,000 times at each of the following values of μ: 0.01, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.9, 1 and at each of the following values of κ: 0.5, 1, 10, 100, 1000. The Cox proportional hazards model was fit to each simulated dataset. From 10,000 simulations at each combination of values of μ and κ, the mean of the estimated coefficient and its standard deviation were saved. The results are summarised in Figure 4.

In the ideal case ($\mu = 1$) the estimate is unbiased and has the smallest error. As the proportion of time at risk decreases, bias is introduced. More bias is introduced when κ is small, i.e.

when there is a larger amount of heterogeneity of exposure. When exposure is very similar for all members of the population (large κ), the model estimate has minimal bias across the range of μ.

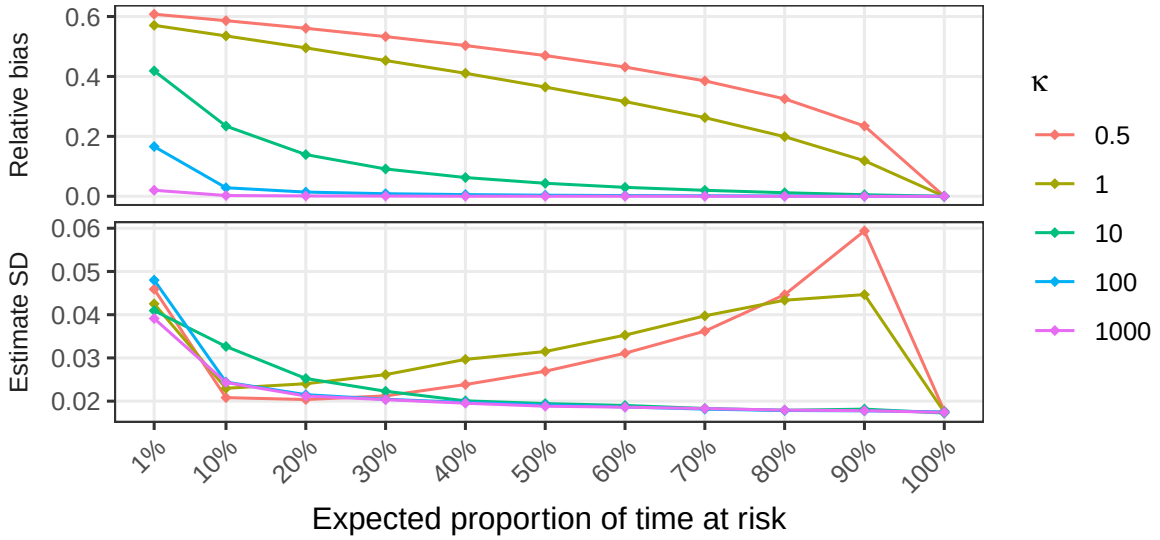

Figure 4: The results of fitting a Cox PH model to the time-to-event simulations. For each combination of μ and κ, the relative bias of the mean of the estimate (top panel) is shown as well the standard deviation of the estimate (bottom panel) obtained from 10,000 simulations. Points represent the values for which the simulations were performed.

The results in Figure 4 show that the Cox PH model can produce (almost) unbiased estimates when time-at-risk is not the same as time-of-follow-up when two assumptions are satisfied. First is that time-at-risk is proportional to time-of-follow-up. Second is that exposure in the population is not overly heterogeneous, e.g. if subjects are expected to have been exposed for 20% of their follow-up time then each individual subject was exposed for close to 20% of their follow-up time (corresponds to simulations at $\kappa = 1000,\ \kappa = 100$) rather than having 20% of the subjects almost always exposed and 80% almost never exposed (corresponds to simulations at $\kappa = 0.5,\ \kappa = 1$).

The assumption of time at risk being proportional to time of follow-up is likely to hold when everyone in the sample is followed through a period of similar disease activity. There may still be subjects who are not exposed (or over-exposed), but as long as the level of exposure does not correlate with the covariates of interest (e.g., antibody titres), then with a large enough sample size the average exposure for any covariate pattern (e.g., antibody titre level) should be proportional in the same way to the time of follow-up. If the disease is seasonal, the start of follow-up can be set to the start of the season. For those who do not get infected, end of follow-up can then be the end of the season. For those who do get infected, end of follow-up can be the infection time assuming that infection grants immunity for the rest of the season. An illustration is in Figure 5.

This assumption will likely not hold if, with a seasonal disease, there are people in the sample whose follow-up starts before the season as shown in Figure 6. For those with earlier follow-up start, their follow-up time will be large regardless of their titres (since they spend a proportion of that time not being at risk). This should make it seem like the titres have a smaller effect than they actually do thus biasing the estimate of titre effect towards the null.

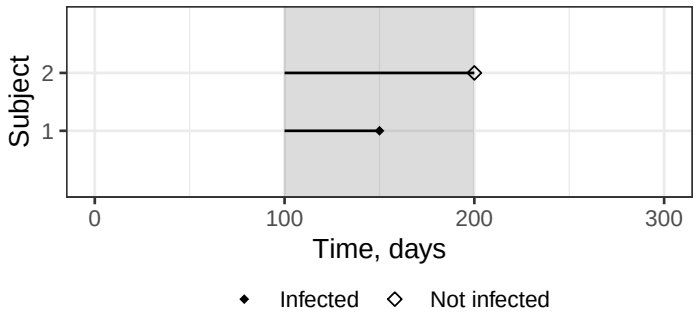

Figure 5: An illustration of a pattern of follow-up where the assumption of true time at risk being proportional to time of follow-up is likely to hold. The shaded region marks the period of time when the disease is active. Both subject's follow-up starts at the beginning of the disease season (activity). Subject 1 gets infected at 150 days, their follow-up would end there assuming infection grants immunity (their total recorded time of follow-up would be 50 days). Subject 2 does not get infected through the season, their time of follow-up ends at the end of the season (their total recorded time of follow-up would be 100 days).

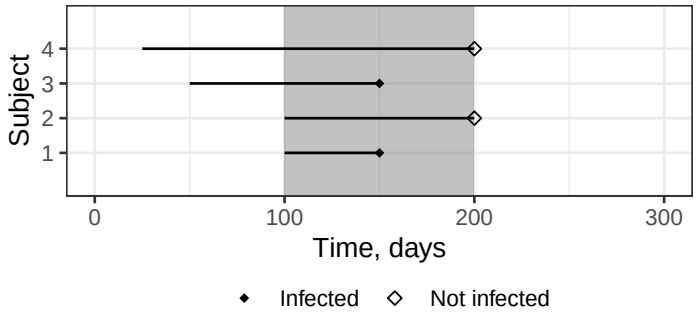

Figure 6: An illustration of a pattern of follow-up where the assumption of true time at risk being proportional to time of follow-up is likely to not hold. The shaded region marks the period of time when the disease is active. Subject 1 gets infected at 150 days, at which point their follow-up would end if it can be assumed that infection grants immunity (their total recorded time of follow-up would be 50 days). Subject 2 does not get infected within the season, their follow-up ends at the end of the season (their total recorded time of follow-up would be 100 days). For subjects 3 (recorded follow up of 100 days) and 4 (recorded follow-up of 175 days) follow-up commenced prior to season onset.

To demonstrate this problem, we performed a simulation study using the same procedure as described above, except a proportion of the subjects were randomly chosen to have their follow-up started earlier. For those randomly chosen, a uniform random number between 0 and 200 (days) was added to their recorded follow-up time. The proportion of time at risk was set to 1. The bias resulting from varying the proportion of the sample with earlier follow-up is shown in Figure 7. The estimate is unbiased when follow-up starts at the start of the season for all subjects. Bias towards the null increases rapidly even when only a small (10%) proportion of the sample has started follow-up prior to the season. For this reason, it would be advisable to record follow-up time in such a way that true unobservable time at risk can be reasonably expected to be proportional to the recorded time of follow-up (i.e., the longer a subject is followed, the longer they are likely to have been at risk).

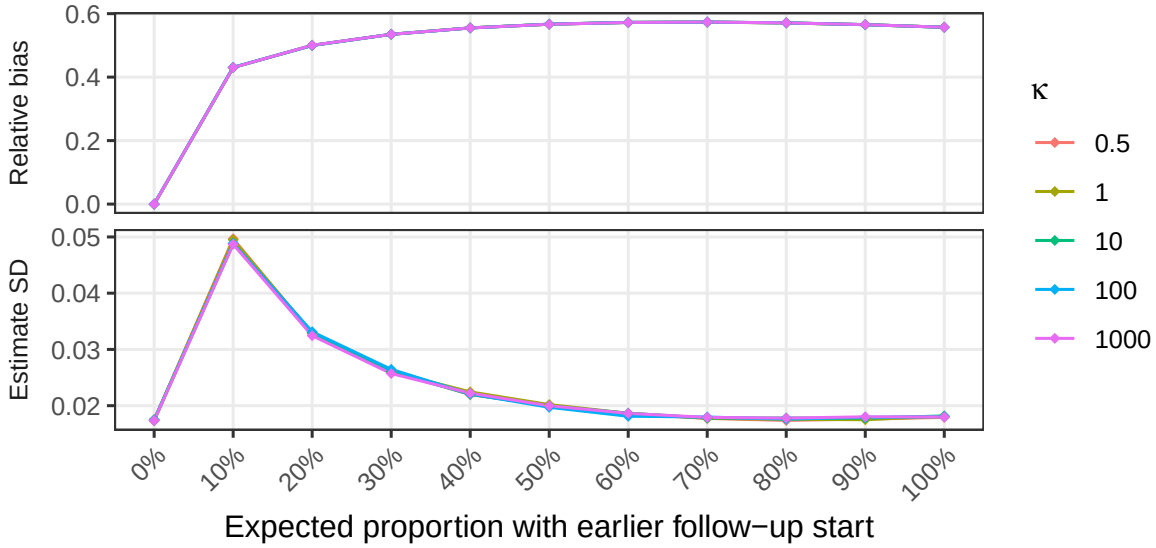

Figure 7: The results of time-to-event simulations. For each proportion of the sample with earlier follow-up start, the mean of the estimated coefficient (left panel) is shown as well as the standard deviation of that coefficient (right panel) from 10,000 simulations. Points represent the values of expected proportion for which the simulations were performed. The dotted horizontal line is the true value of the estimated parameter.

4 Logistic regression

The logistic model assumes that the probability of outcome follows a logistic curve from 1 (at low covariate values assuming a protective covariate) to 0 (at high covariate values). If there is only one covariate which is the antibody titre measurement, then the model is:

$$P(Y=1) = \frac{\exp(\beta_0 + \beta_T X_{\text{logtitre}})}{1 + \exp(\beta_0 + \beta_T X_{\text{logtitre}})}$$

Where β_0 is the log-odds of infection when $X_{\text{logtitre}} = 0$ and β_T is the log-odds ratio of infection of subjects with a given X_{logtitre} compared to subjects subjects whose X_{logtitre} is 1 unit lower.

A potentially large problem with the application of this model to antibody data is that low antibody titres do not necessarily correlate with a high probability of infection (which is one of the model's assumptions). This assumption of baseline risk of 1 can be justified if it is the case that subjects with low antibody titres always (or almost always) get infected. If a large proportion of subjects with low titres do not get infected, this assumption is not justified and this model will produce a poor fit. As an example, this assumption is not justified in the Ha Nam data as shown in Figure 8. In both the general and the exposed populations, less than half the subjects with the titres below detectable levels (recorded as 5) were infected.

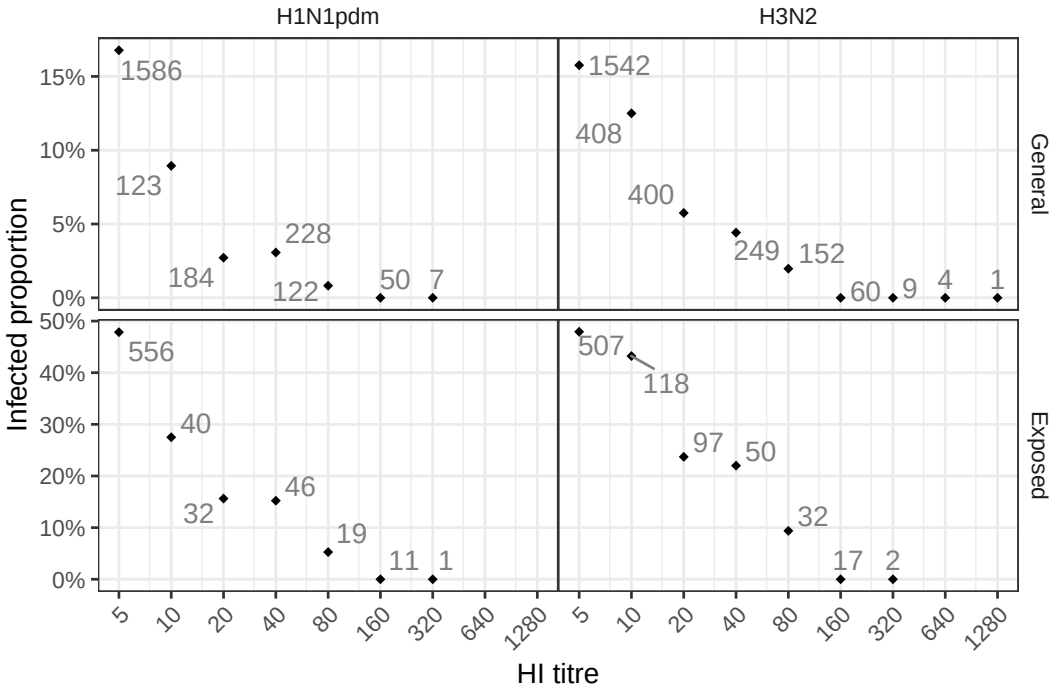

Figure 8: Ha Nam cohort data. Subjects were grouped by virus and measured HI titre. Numbers next to points are the total number of observations in the corresponding group. The top row is all observations. The bottom row are the observations from households with at least one infection in a given season. The left column is observations for the H1N1pdm virus, the right row — for H3N2 virus.

If the baseline risk cannot justifiably be assumed to be 1, the fitted probability curve will be flatter than the true curve, thereby misrepresenting the true probability of infection across the titre range as depicted in Figure 9. Thus, ignoring the baseline risk assumption will lead to poor model fit and hence biased estimates of titre effect.

Estimating the baseline risk in the above model leads to a "scaled logistic" model which accounts for the issue [8].

5 Scaled logistic regression

This model is the same as logistic regression except that it estimates the baseline probability of outcome (i.e. the probability at low covariate values assuming a protective covariate) as

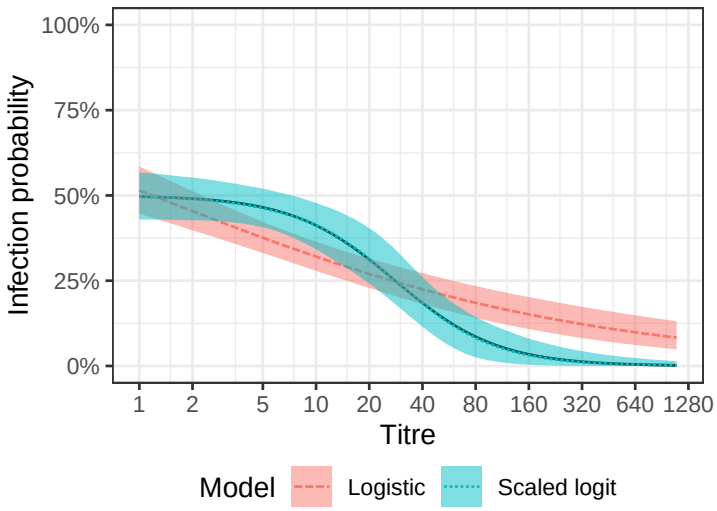

Figure 9: An illustration of a bad fit of the logistic model. The solid line ($0.5\frac{\exp(5-1.5X_{\text{logtitre}})}{1+\exp(5-1.5X_{\text{logtitre}})}$) is the true probability curve where the baseline risk is 0.5. The dashed line is the expected fitted probability curve from standard logistic regression. The dotted line is the fitted probability curve from scaled logistic regression. The shaded regions are the 95% confidence intervals. The expected curves and CIs was obtained from 10,000 simulations each with 500 observations simulated from the true curve. The models were fitted to each simulated dataset and the means of the regression estimates were taken from all simulations.

opposed to assuming that it is equal to 1. If there is only one covariate which is the antibody titre measurement then the model is

$$P(Y=1)=\frac{\lambda}{1+\exp(\beta_0+\beta_T X_{\text{logtitre}})}$$

Note that with the above parameterisation, the β parameters are negated relative to logistic regression.

The model still assumes that low probability bound is 0 (i.e. that high titres guarantee immunity in a univariate model). This assumption is justified if there is a reasonable number of people in the sample who have high titres and none (or very few) of whom get infected. In the Ha Nam data (Figure 8), there is a total of 131 observations with titres above 160. None of those subjects got infected in the corresponding season justifying the assumption that high titres guarantee immunity.

Compared to logistic regression, the scaled logit model requires a larger sample size. This is because the scaled logit model attempts to use the same amount of information to estimate one more parameter. This leads to larger standard errors and potential convergence problems. Figures 10 and 11 summarise 10,000 simulation results from the model $\frac{0.5}{\exp(-5+1.5X_{\text{logtitre}})}$ with X_{logtitre} simulated from $N(2,2^2)$. Logistic and scaled logit models were fit using maximum likelihood estimation.

Standard errors of the β regression estimates (β_0 being the intercept and β_T being X_{logtitre} coefficient) were consistently higher with the scaled logit model. In order to reliably converge

(under the chosen true parameter values) the scaled logit model required the sample size to be over 500. The standard logistic model always converged.

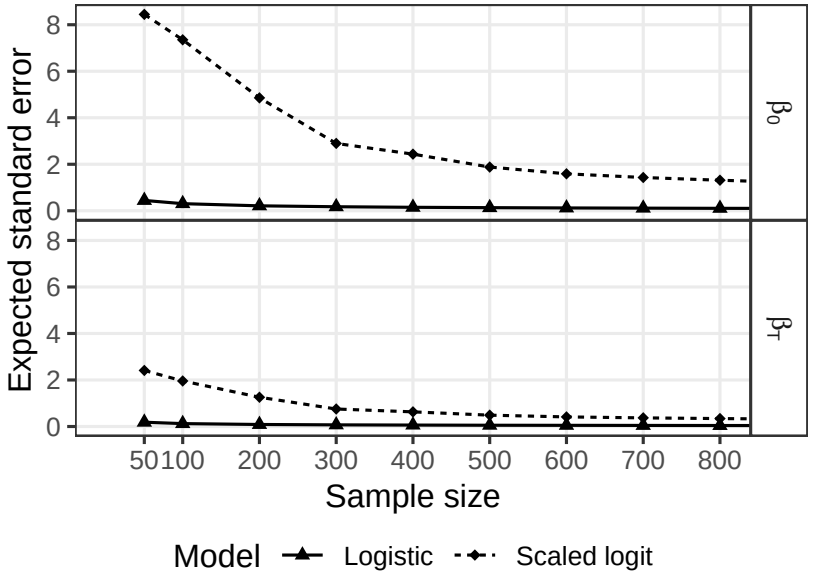

Figure 10: Simulation results of fitting the standard logistic and the scaled logit models to the same simulated datasets. 10,000 datasets were simulated. Shown is the mean standard deviation (estimate of the expected standard error) of the β regression estimates from the standard logistic (the solid line) and the scaled logit (the dashed line). Points indicate parameter values at which the simulations were performed.

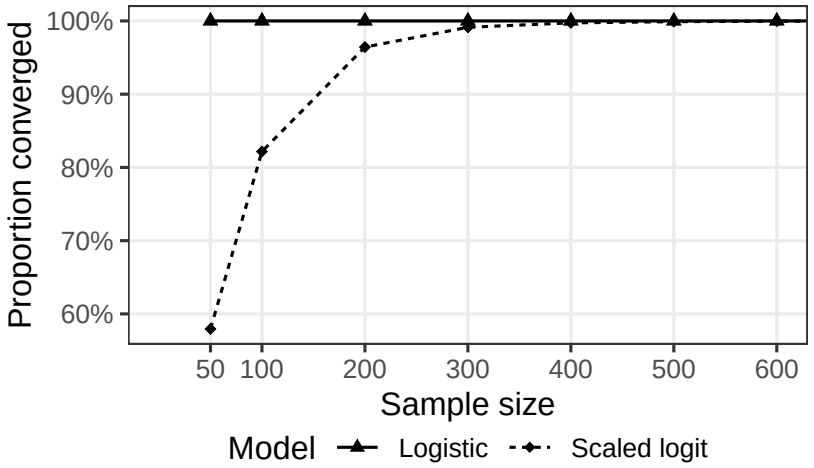

Figure 11: Simulation results of fitting the standard logistic and the scaled logit models to the same simulated datasets. 10,000 datasets were simulated. Shown is the proportion of successful attempts to fit the standard logistic (the solid line) and the scaled logit (the dashed line) model. Points indicate parameter values at which the simulations were performed.

6 Ha Nam application

This section shows an application of the scaled logit model to the Ha Nam cohort which is plotted in Figure 8. For this cohort, the Cox model is inapplicable due to absence of reliable data on disease activity. The standard logistic model is inappropriate due to unjustified assumption of baseline of 1.

6.1 Scaled logit fit

The scaled logit model was fit using the maximum likelihood method. The fitted protection curves are in Figure 12.

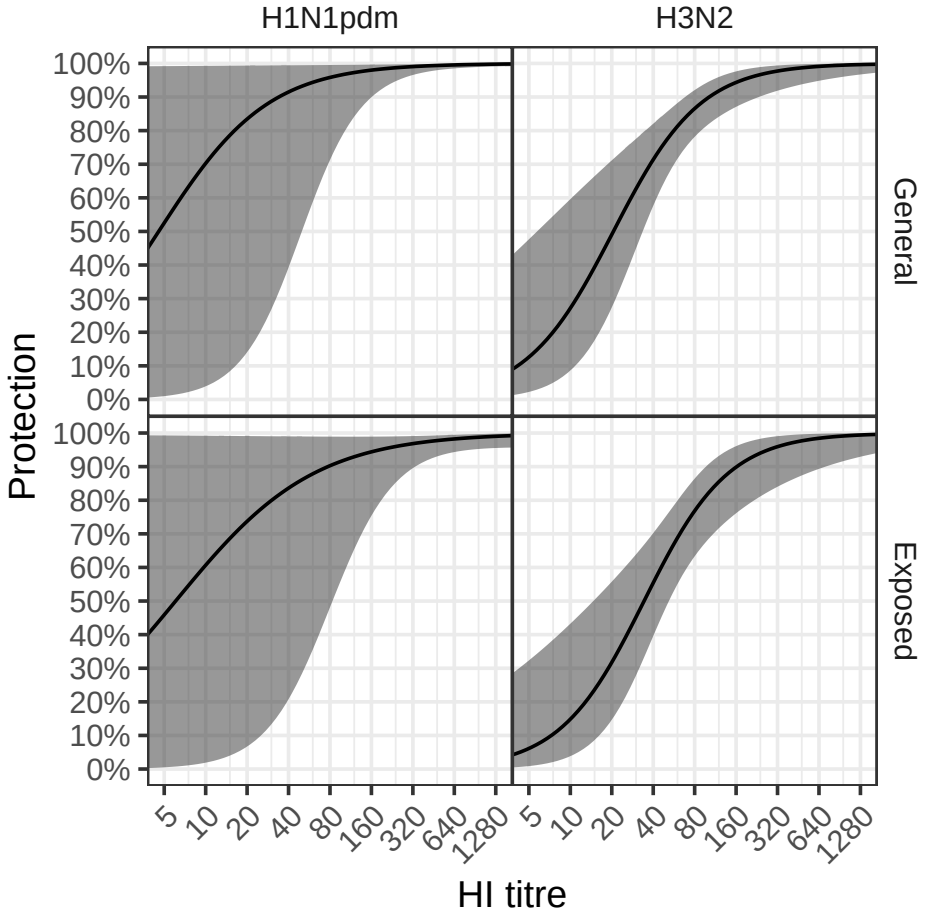

Figure 12: Fitted protection curves and confidence intervals from the scaled logit fit to Ha Nam data (also shown in Figure 8) using the maximum likelihood method. The solid line is the point estimate. The shaded region is the 95% confidence interval.

There is not enough data in the H1N1pdm subset to generate useful estimates. For H3N2 subset, limiting the sample to just those exposed to the virus (at least one household infection in a season) improved the precision of the estimates. The general lack of precision in the estimates is likely due to the number of parameters in the model (3 parameters with 1 covariate) and the censored nature of HI titre measurements, particularly the fact that any

titre below 10 is undetectable which means that it is impossible to distinguish between many of the observations (e.g. some subjects with undetectable titres may have had a true titre of 9 while others — of 2, but both were recorded as 5) which makes it harder to estimate the baseline probability (as seen in the confidence bounds increasing at small ($<$10) titres).

6.2 Comparison to standard logistic fit

As mentioned before, the standard logistic model is inappropriate due to unsatisfied assumption of baseline of 1. It was fit to the same Ha Nam data using maximum likelihood with no accounting for censored titres (observations of 5 (below detectable) and 1280 (above detectable) were unchanged, all other observations were moved to the midpoint of the corresponding censored interval on a log scale). The fitted infection curve is in Figure 13. -ba

While the infection curves appear to fit the data well, it is problematic to generate a protection curve from these results. There are two options for the protection curve. One is to use the same procedure as was used in the scaled logit fit — divide the fitted infection probabilities by the baseline (here assumed to be 1, so the fitted values would not change) and subtract the resulting relative-to-baseline infection probabilities from 1. The protection curves resulting from this procedure are in Figure 14.

The other option for generating a protection curve is to calculate the fitted probability of infection at a give titre and divide that by the fitted probability of infection at the titre of 5 (or any other) thereby generating a curve that shows relative-to-5 infection probabilities (as opposed to relative-to-baseline). The variance of this quantity may be estimated by using the bootstrap method. Subtracting these relative-to-5 infection probabilities from 1 generates curves shown in Figure 15.

While the relative-to-5 protection curves (Figure 15) appear more plausible than the relative-to-baseline protection curves (Figure 14), both result from fitting a model with an unsatisfied assumption and neither method of generating protection curves from logistic regression model reliably produce accurate results.

The relative-to-5 curve (Figure 15) presents an additional problem. The curve shows how much "better" different titres are at protecting against infection than the titre of 5. There is nothing inherently special about this threshold of 5. Its choice is based on the lower dilution of 10. Pre-treatment of the sera necessitates dilution, so the lower bound can never be 1, but could be something more than 1 such as $<$5. Nevertheless, the curves may look substantially different if a different threshold (e.g. 10 or 1) is chosen.

7 Kiddyvax application

The data for this study includes post-vaccination HI titres of subjects who were followed for up to one year for flu infection. The infection status was determined by PCR which was done for everyone who experienced symptoms. This data is shown in Figure 16.

Analyses were done on the data for B Vic and A(H1pdm) viruses in the same way as for the Ha Nam data with the addition of the Cox proportional hazards model.

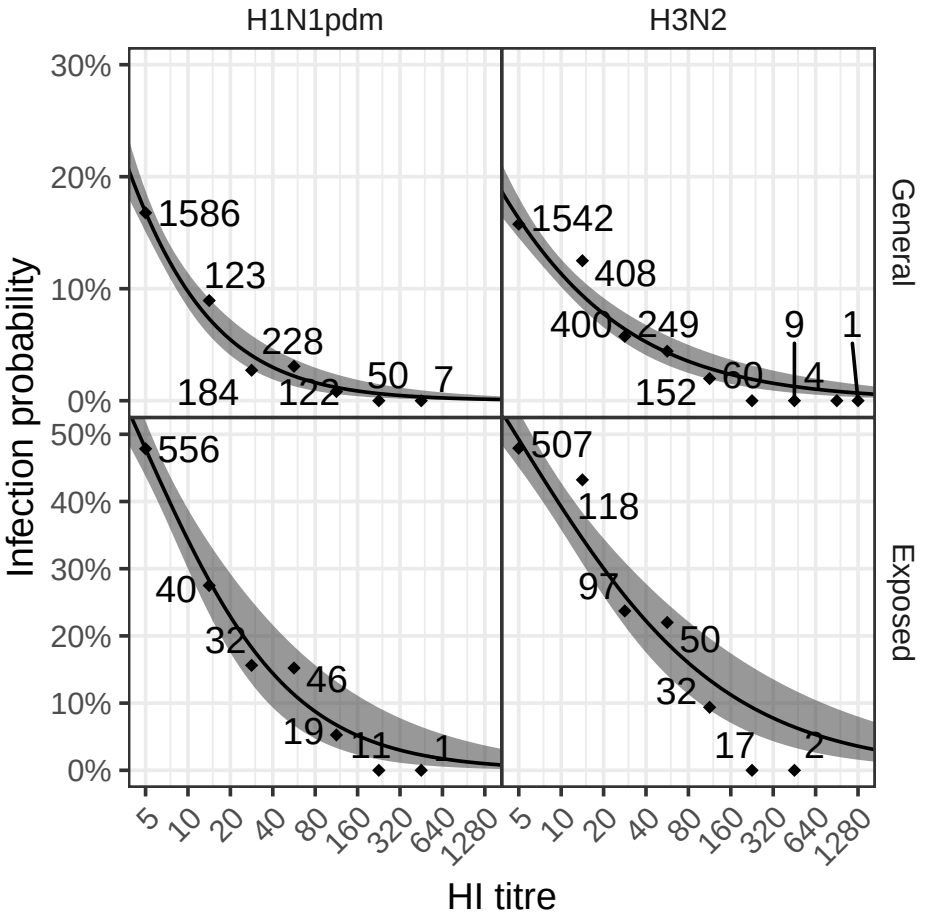

Figure 13: Fitted infection curves and confidence intervals from the standard logistic model fit to Ha Nam data (also shown in Figure 8) using maximum likelihood with no accounting of censored titres (observations of 5 (below detectable) and 1280 (above detectable) were unchanged, all other observations were moved to the midpoints of the corresponding censored intervals on a log scale). The points are the infected proportions at the corresponding modified titre measurements (i.e. interval midpoints). The solid line is the point estimates. The shaded region is the 95% confidence interval. The numbers next to the points are the total sample size of the corresponding groups.

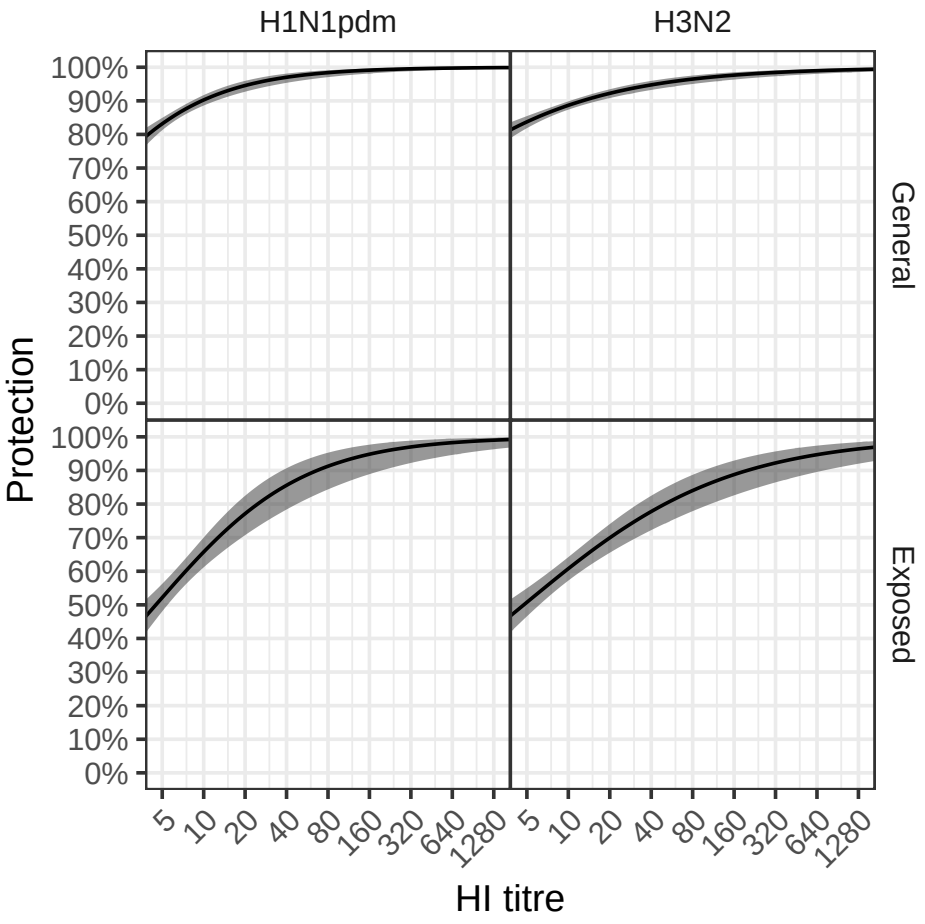

Figure 14: Fitted protection curves and confidence intervals from the standard logistic model fit to Ha Nam data (also shown in Figure 8) using maximum likelihood with no accounting of censored titres (observations of 5 (below detectable) and 1280 (above detectable) were unchanged, all other observations were moved to the midpoints of the corresponding censored intervals on a log scale). The solid line is the point estimates. The shaded region is the 95% confidence interval.

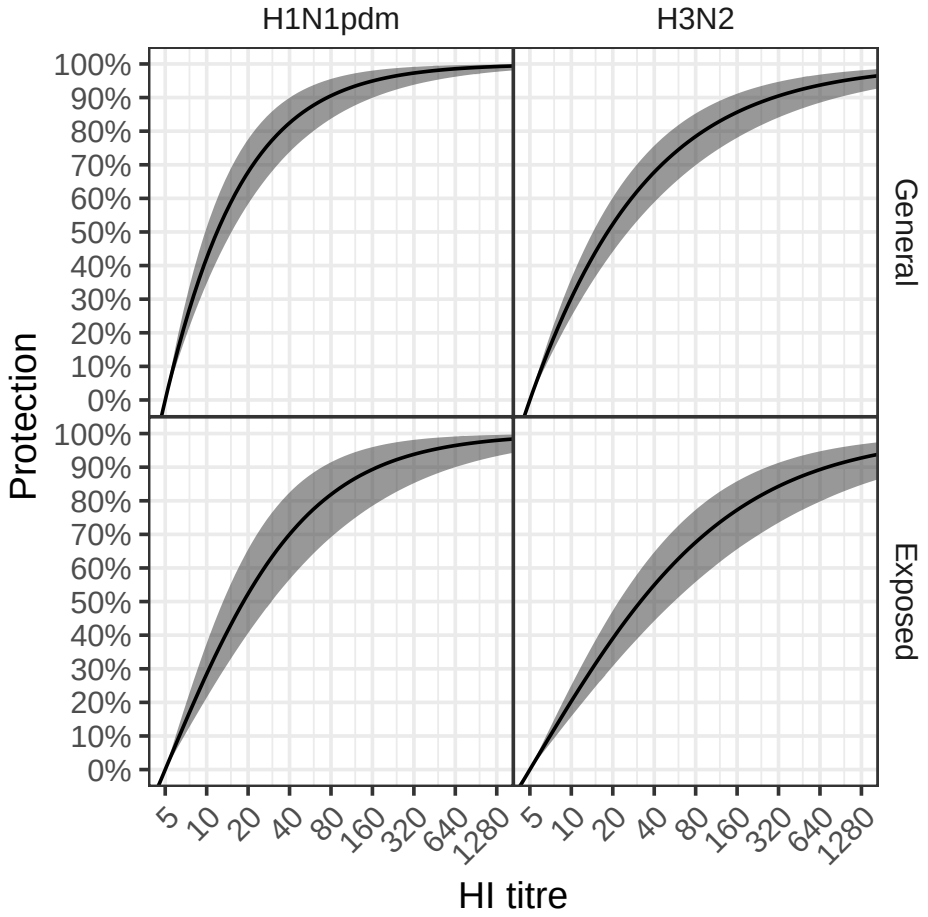

Figure 15: Fitted relative-to-5 protection curves and confidence intervals from the standard logistic model fit to Ha Nam data (also shown in Figure 8) using maximum likelihood with no accounting of censored titres (observations of 5 (below detectable) and 1280 (above detectable) were unchanged, all other observations were moved to the midpoints of the corresponding censored intervals on a log scale). Solid lines are the point estimates, with shaded regions the 95% confidence interval. The bounds for the confidence interval were obtained by using the bootstrap method (10,000 samples).

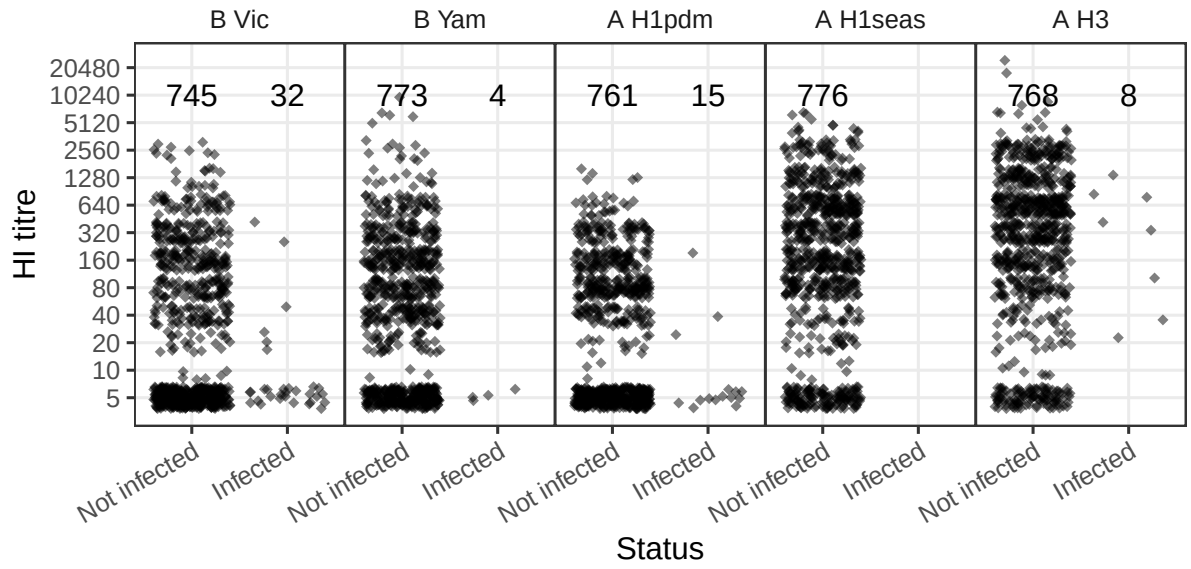

Figure 16: Kiddyvax study data. Post-vaccination titres are shown for those who got infected (PCR-confirmed symptomatic infection) over the course of the study and those who did not. Panels correspond to the five tested viruses.

7.1 Scaled logit fit

The same scaled logit model was fit in the same way as for the Ha Nam data. The protection curves are in Figure 17. The model fails to produce useful estimates due to there being very few infections in the sample.

7.2 Standard logistic fit

The relative protection curves are in Figure 18. The same considerations regarding the unjustified assumption of baseline risk of 1 apply here as they did with Ha Nam (Figure 16 shows many uninfected subjects with undetectable titres).

7.3 Cox proportional hazards fit

For infected individuals, the time at risk was taken to be the time from start of follow-up to infection. For those who did not get infected the time at risk was take to be the time from start of follow-up to end of follow-up.

The model was

$$h(t) = h_0 \exp(\beta_T X_{\text{logtitre}})$$

where h is the hazard function and X_{logtitre} is the post-vaccination titre measurement on the log scale. The resulting protection curves (relative to the titre of 5) are in Figure 19. Although

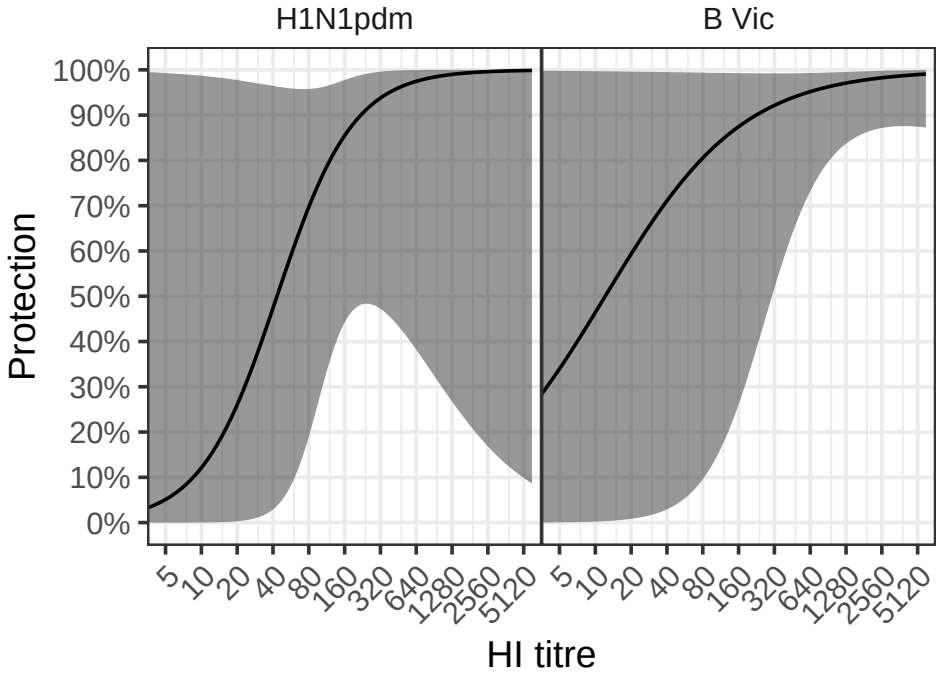

Figure 17: Fitted protection curves and confidence intervals from the scaled logit fit to Kiddyvax data (also shown in Figure 16) using the maximum likelihood method. The solid line is the point estimate. The shaded region is the 95% confidence interval.

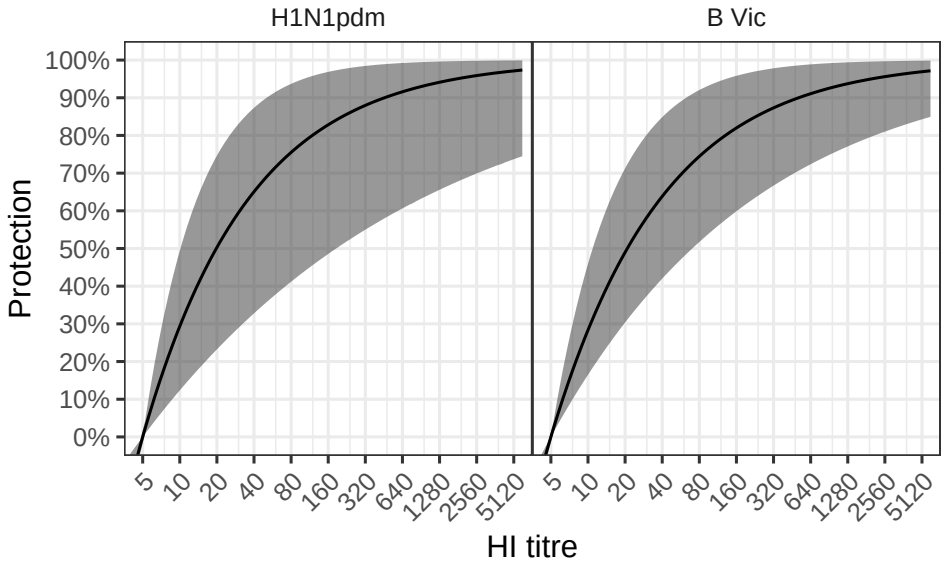

Figure 18: Fitted relative-to-5 protection curves and confidence intervals from the standard logistic model fit to kiddyvax data (shown in Figure 16) using maximum likelihood with no accounting of censored titres (observations of 5 (below detectable) were unchanged, all other observations were moved to the midpoints of the corresponding censored intervals on a log scale). The solid line is the point estimates. The shaded region is the 95% confidence interval. The bounds for the confidence interval were obtained by using the bootstrap method (10,000 samples).

the point estimates are similar to those recovered from the model; however standard errors are inflated.

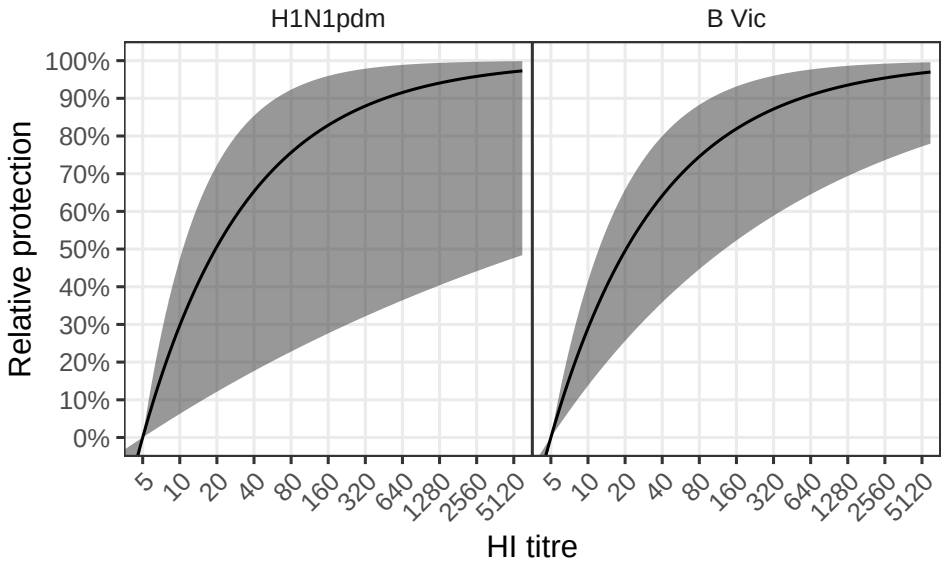

Figure 19: Fitted relative-to-5 protection curves and confidence intervals from the Cox proportional hazards model fit to kiddyvax data (shown in Figure 16) with no accounting of censored titres (observations of 5 (below detectable) were unchanged, all other observations were moved to the midpoints of the corresponding censored intervals on a log scale). The solid line is the point estimates. The shaded region is the 95% confidence interval.

8 Conclusion

Table 1: Summary of the three considered models in terms of their application to antibody data

Model	Potential problem
Cox PH	Biased if follow-up time is not proportional to time at risk for everyone in the sample
Logistic	Biased if low antibody titres do not guarantee immunity
Scaled logit	Requires a large sample size.

In this paper we have explored three different models for estimating protective antibody titres using data from influenza vaccine and infection studies. We have shown that in the presence of good time-to-event data where every subject's follow-up time is at least proportional to their time at risk, the Cox model will likely perform best out the three models explored due to it having the least number of parameters allowing for more precise estimates. Absent such data, logistic regression may be appropriate if the assumption of everyone being infected at low antibody titres can be justified. If this assumption cannot be justified, which is probably the case for influenza studies, the scaled logit model can be applied but the sample needs to be fairly large ($>$500) in order to obtain useful estimates.

References

[1] Barr IG, Russell C, Besselaar TG, Cox NJ, Daniels RS, Donis R, et al. WHO recommendations for the viruses used in the 2013-2014 Northern Hemisphere influenza vaccine: Epidemiology, antigenic and genetic characteristics of influenza A(H1N1)pdm09, A(H3N2) and B influenza viruses collected from October 2012 to January 2013 [Journal Article]. Vaccine. 2014;32(37):4713–25. Available from: https://www.ncbi.nlm.nih.gov/pubmed/24582632.

[2] Wood JM, Levandowski RA. The influenza vaccine licensing process [Journal Article]. Vaccine. 2003;21(16):1786–8. Available from: https://www.ncbi.nlm.nih.gov/pubmed/12686095.

[3] Hobson D, Curry RL, Beare AS, Ward-Gardner A. The role of serum haemagglutination-inhibiting antibody in protection against challenge infection with influenza A2 and B viruses. Journal of Hygiene. 1972;70(04):767–777.

[4] Ng S, Fang VJ, Ip DK, Chan KH, Leung GM, Peiris JS, et al. Estimation of the association between antibody titers and protection against confirmed influenza virus infection in children [Journal Article]. J Infect Dis. 2013;208(8):1320–4. Available from: https://www.ncbi.nlm.nih.gov/pubmed/23908481.

[5] Cowling BJ, Ng S, Ma ES, Fang VJ, So HC, Wai W, et al. Protective efficacy against pandemic influenza of seasonal influenza vaccination in children in Hong Kong: a randomized controlled trial [Journal Article]. Clin Infect Dis. 2012;55(5):695–702. Available from: https://www.ncbi.nlm.nih.gov/pubmed/22670050.

[6] Horby P, Mai LQ, Fox A, Thai PQ, Yen NTT, Thanh LT, et al. The Epidemiology of Interpandemic and Pandemic Influenza in Vietnam, 2007–2010. American Journal of Epidemiology. 2012;175(10):1062–1074.

[7] George B, Seals S, I A. Survival analysis and regression models [Journal Article]. J Nucl Cardiol. 2014;21(4):686–694. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4111957/.

[8] Dunning AJ. A model for immunological correlates of protection. Statistics in Medicine. 2006;25(9):1485–1497.

All code is in github.com/khvorov45/model-comparison