

Statistical  
influenza  
Models for antibody data  
modelling of?  
or generalisable?

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

February 21, 2020

## 1 Introduction

Probably need a bit more background - Sheena?

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. Does this need a Ref.?

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. 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.

titres are interval-based  
that is, ...

For influenza vaccines, 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 to 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].

*Cite paper, paper 23 → [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 we have seen used in the literature: Cox proportional hazards regression; ordinary 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.

including... w/ Ref's

## 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 receive 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. 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 about 7 months for signs of influenza-like illness. Symptoms reported by any household member prompted swabbing and influenza testing by PCR for all household members [5]. The protective titre could be estimated using a Cox proportional hazards model where the outcome was time to infection or the end of the study [4].

We direct the reader to [5] for full details of the study, but provide a brief outline here.

This is vague. Did they do this?

### 2.2 The Ha Nam household study

The Ha Nam study has followed 270 households in Hanam, Vietnam since 2007. 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 summarised in Figure 1.

depicted/demonstrated/stylised

This needs to be clearer/more explicit. We can't measure when an ind. is in close contact w/ an infectious ind.?

I'm unclear about what Fig. 1 is showing. Neither of these studies are challenge studies?

### 3 COX PROPORTIONAL HAZARDS

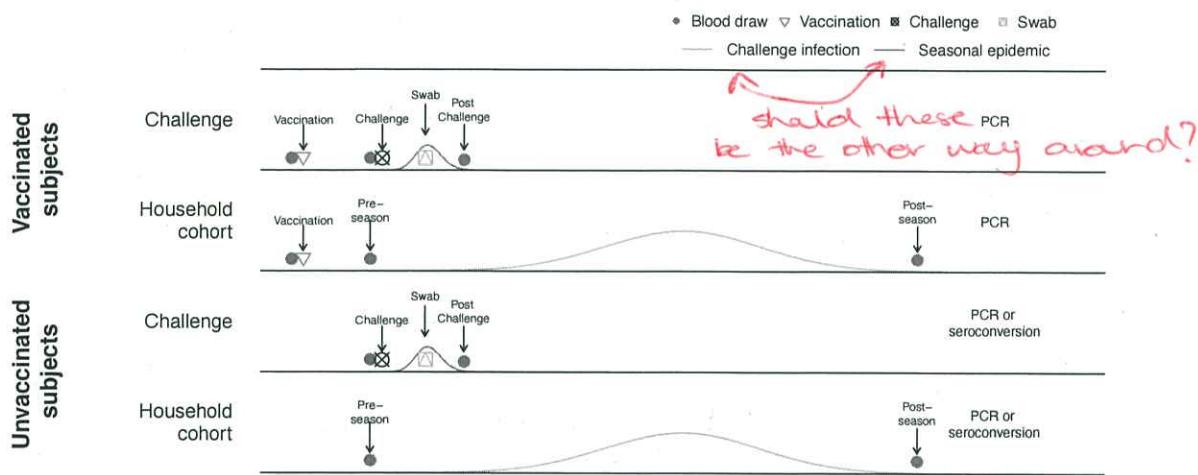


Figure 1: Example study designs for the estimation of protective titres.

### 3 Cox proportional hazards

In Cox proportional hazards regression, the outcome considered 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).

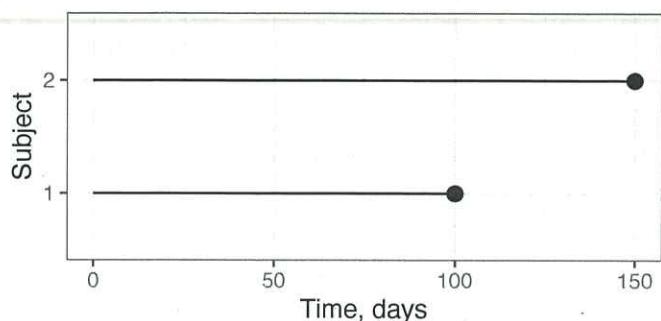


Figure 2: An example of time-to-event data. Two subjects are followed from time 0. Subject 1 experiences the even at time 100. Subject 2 experiences the event at time 150.

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.

If subjects are not at risk of the event 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.

### 3 COX PROPORTIONAL HAZARDS

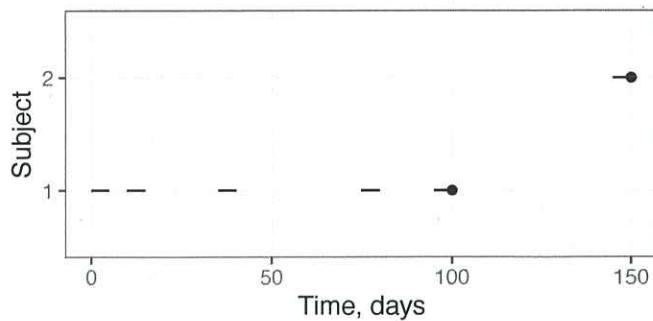


Figure 3: An example of time to event data. Two subjects are followed from time 0. Subject 1 experiences the event at time 100. Subject 2 experiences the event at time 150. Subject 1 was at risk of the event for a total of 5 days. Subject 2 was at risk of the event for a total of 25 days.

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 controlled. If follow-up time is completely unrepresentative of true time at risk, the Cox model will not produce reliable 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 reliable results.

✓ observed / measured?

To investigate this, we generated a simulated dataset, based on the following model:

*be precise.*

$$T \sim \text{Exponential}(\text{rate} = \lambda)$$

$$h(t) = \lambda$$

$$\log \lambda = -3 - 1.5X_{\text{logit}}$$

*logit*

Where  $T$  is the survival time,  $h$  is the hazard and  $X_{\text{logit}}$  is the true logit measurement simulated from  $N(2, 2^2)$ .

*need to make sure you have described this earlier over an X day period window.*

Each individual was assigned a proportion of time they were exposed to the virus. This proportion was generated from  $\text{Beta}(10, 10 \frac{1-m}{m})$  where  $m$  is the expected proportion for the population. When  $m$  was 1, the proportion assigned was always 1 to represent the ideal context of all follow-up time being time at risk. The maximum time of follow-up was set to 200 (days). If an individual's true survival time (a random number generated by  $\text{Exponential}(\text{rate} = \exp(-3 - 1.5X_{\text{logit}}))$ ) was at most 200 times their “time at risk proportion”, then they were “infected”. Otherwise, they were “not infected”.

*This is not clear.*

Recorded time for each individual was 200 for everyone not infected (i.e., they were followed for the entire follow-up period and did not experience the event) and the true survival time divided by “time at risk proportion” for everyone infected (i.e. they were followed for that amount of time before experiencing the event).

*censored?*

Each simulation included 10,000 observations and was run 10,000 times at different values of the expected proportion of time at risk. The Cox proportional hazards model was fit to each simulated dataset. From 10,000 simulations at each value of the expected proportion

*10,000 times for each of the n propn. dim at risk values?*

*1%, 10%, 20%, 30%, ..., 100%.*

### 3 COX PROPORTIONAL HAZARDS

at risk, the mean of the estimated coefficient (a measure of its expected value at that proportion) and its standard deviation (a measure of its expected standard error at that proportion and sample size) were saved. The results are summarised in Figure 4.

In the ideal case (proportion of time at risk set to 1) the estimate is unbiased and has the smallest error. As the proportion of time at risk decreases, bias is introduced while the error remains almost constant. When the proportion is small (less than 20%), bias appears to decrease while error increases.

The maximum bias was less than 4% away from the true value. The standard error increases notably only when the expected proportion at risk is very small (less than 10%). This shows that under the assumption that the time of follow-up is proportional to time at risk, the Cox model can be expected to produce reliable results.

?

↑ I'd say approx.  
<30%  
or 40%

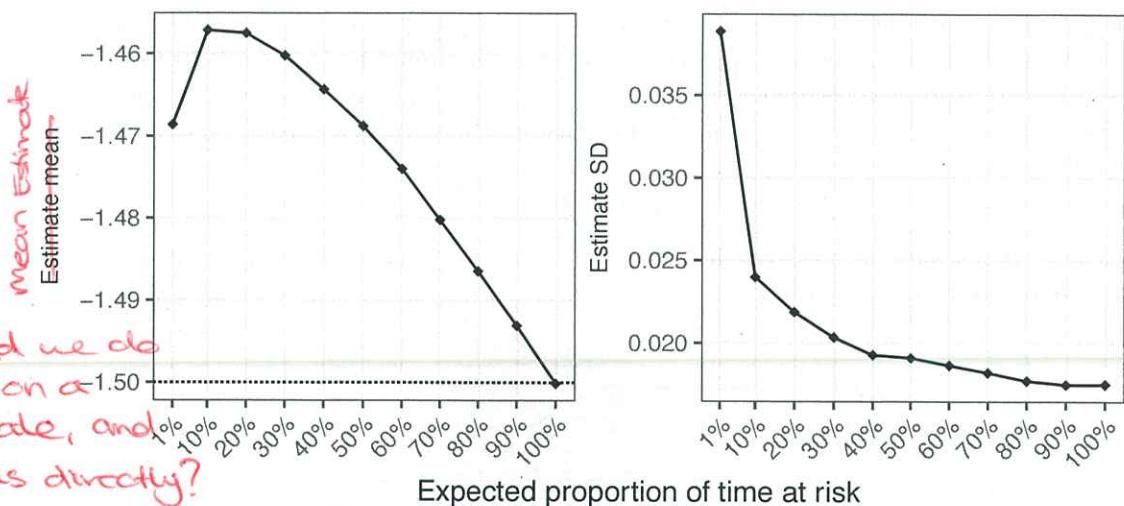


Figure 4: The results of time-to-event simulations. For each proportion of time at risk, the mean of the estimated coefficient (left panel) is shown as well its standard deviation (right panel) obtained from 10,000 simulations. Points represent the values of expected proportion for which the simulations were performed. The dotted horizontal line is the true value of the estimated parameter.

in (a)

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

The assumption will likely not hold if, with a seasonal disease, there are people in the sample whose follow-up starts before the season. An illustration is 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

5

Can we also consider heterogeneity in exposure?  
i.e., Beta(0.5, 0.5), or similar. Vary  $a + b$  ( $\text{Beta}(a,b)$ ) to  
get greater/lesser propn's ~~not~~ exposed for  $x < 0.2$  or  $> 0.8$

### 3 COX PROPORTIONAL HAZARDS

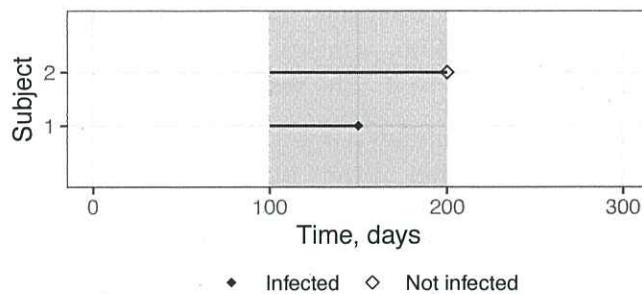


Figure 5: An illustration of a pattern of follow-up where the assumption of true time at risk being proportional to time of follow-up is likely to hold. The shaded region marks the period of time when the disease is active. Both subject's follow-up starts at the beginning of the disease season (activity). 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).

*(I really think we should use "censored" language.)*

like the titres have a smaller effect than they actually do thus biasing the estimate of titre effect towards the null.

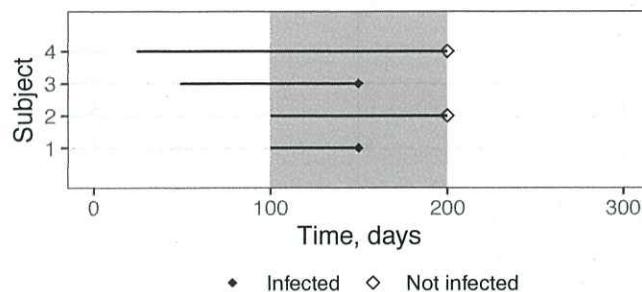


Figure 6: An illustration of a pattern of follow-up where the assumption of true time at risk being proportional to time of follow-up is likely to not hold. The shaded region marks the period of time when the disease is active. Subject 1 gets infected at 150 days, at which point their follow-up would 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.

*(a proportion of the a simulation study)*

To demonstrate this problem, we performed some additional simulations using the same procedure as described above, except subjects were randomly chosen to have had their follow-up started earlier. For those 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. Bias towards the null increases rapidly even when only a small (10%) proportion

*for all subjects*

## 4 LOGISTIC REGRESSION

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 ~~so~~ 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).

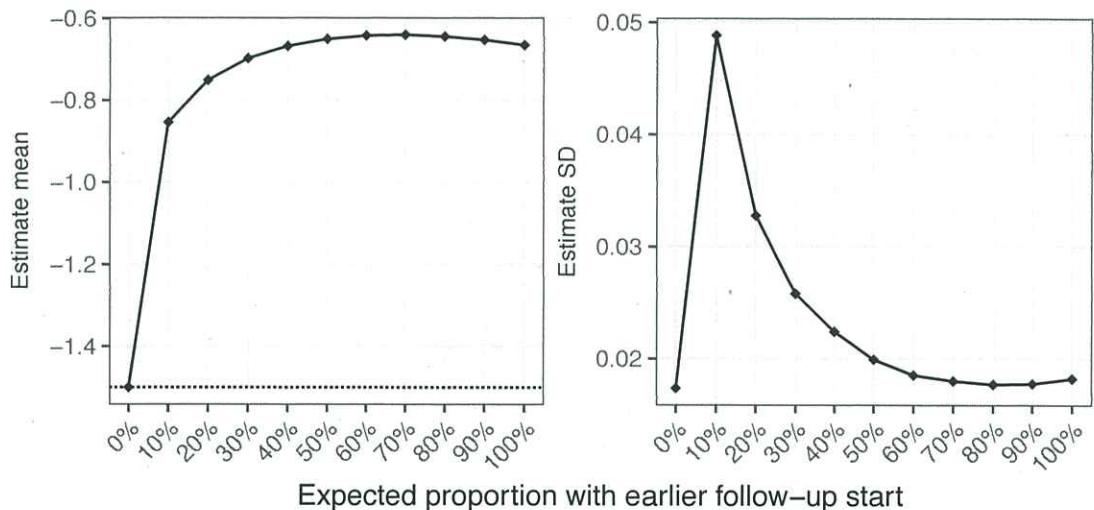


Figure 7: The results of time to event simulations. For each proportion of the sample with earlier follow-up start, the mean of the estimated coefficient (left panel) is shown as well as the standard deviation of that coefficient (right panel) from 10,000 simulations. Points represent the values of expected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

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

*Describe  $\beta_0$  &  $\beta_T$  here (once), so you don't need to again*

A potentially large problem with the application of this model to antibody data is that low antibody titres do not necessarily guarantee infection (which is one of the model's assumptions). This assumption of baseline risk of 1 can be justified if it is shown 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. 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.

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

#### 4 LOGISTIC REGRESSION

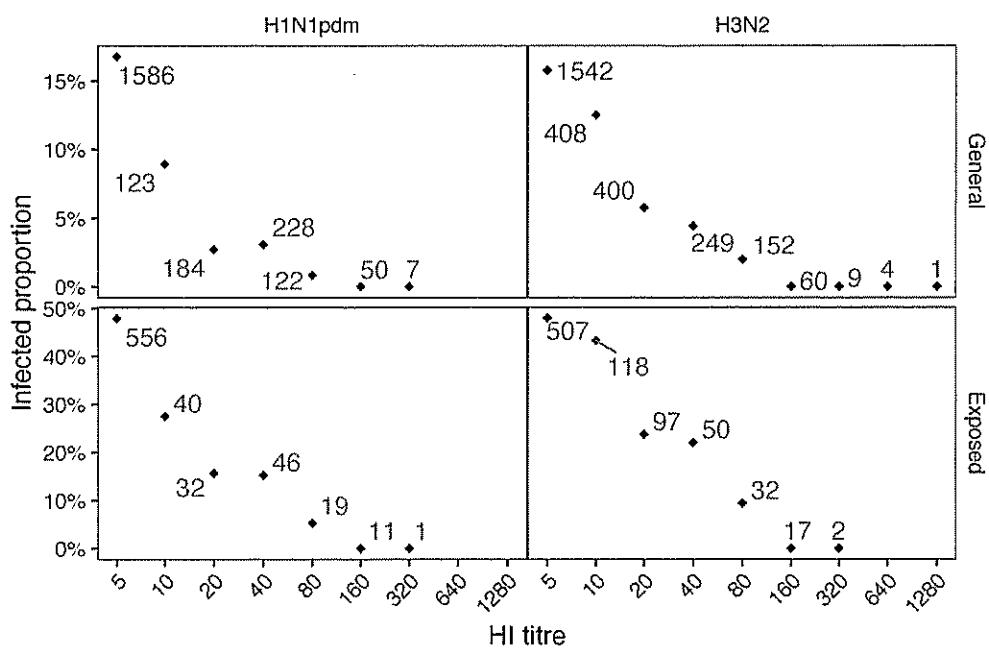


Figure 8: Ha Nam cohort data. Subjects were grouped by virus and measured HI titre. Numbers next to points are the total number of observations in the corresponding group. The top row is all observations. The bottom row are the observations from households with at least one infection in a given season.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

*column*

## 5 SCALED LOGISTIC REGRESSION

across the titre range; An illustration is in Figure 9. Thus, ignoring the baseline risk assumption will lead to biased estimates.

depicted

poor model-fit, and hence, estimates of titre protection

b) You can prob. word this better than I have...

I would lead here with "estimating the baseline risk in the model leads to the scaled logistic reg<sup>n</sup> model" and accounts for this issue"

--- or something.

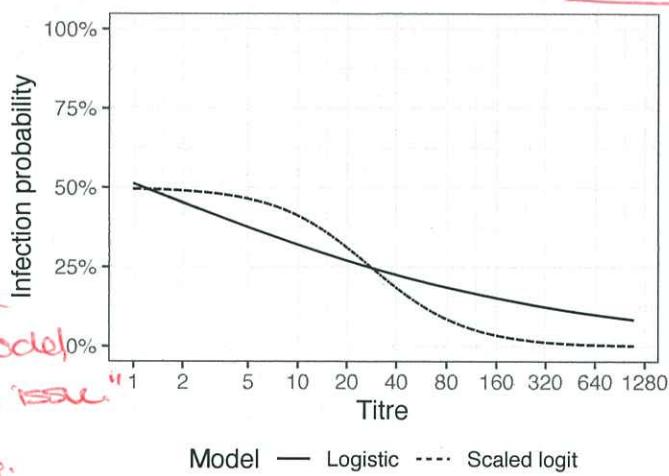


Figure 9: An illustration of a bad fit of the logistic model. The dotted line (overlaps the dashed line) is the true probability curve  $0.5 \frac{\exp(5 - 1.5X_{\text{logtitre}})}{1 + \exp(5 - 1.5X_{\text{logtitre}})}$ . The solid line is the expected fitted probability curve from standard logistic regression. The dashed line (overlaps the dotted line) is the expected fitted probability curve from scaled logistic regression. The expected curves were obtained from 10,000 simulations each with 10,000 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.

→ Can we also add CI bands to this

Fig. to show diff. in SE. will also help distinguish true + estimated scaled LR lines.

You've already said this.  
Are there two lines?  
May be do colour, or dotted/dashed

### 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 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  $\beta_i$  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

(also too strong)  
I'd say that this assumption is not unreasonable  
not that it is justified

## 5 SCALED LOGISTIC REGRESSION

*Hence, for the same sample size, the scaled*

*is estimating an additional parameter. This leads to larger standard errors, and potential convergence problems. Figures 10 and 11 summarise 10,000 simulation results from the model*

$$\frac{0.5}{\exp(-5+1.5X_{\text{logit}})}$$

*with  $X_{\text{logit}}$  simulated from  $N(2, 2^2)$ . Logistic and scaled logit models were fit using maximum likelihood estimation.*

*Can you justify back p this claim? or say it later*

*Don't use \frac with inline text.*

*can you show estimates of par. too.*

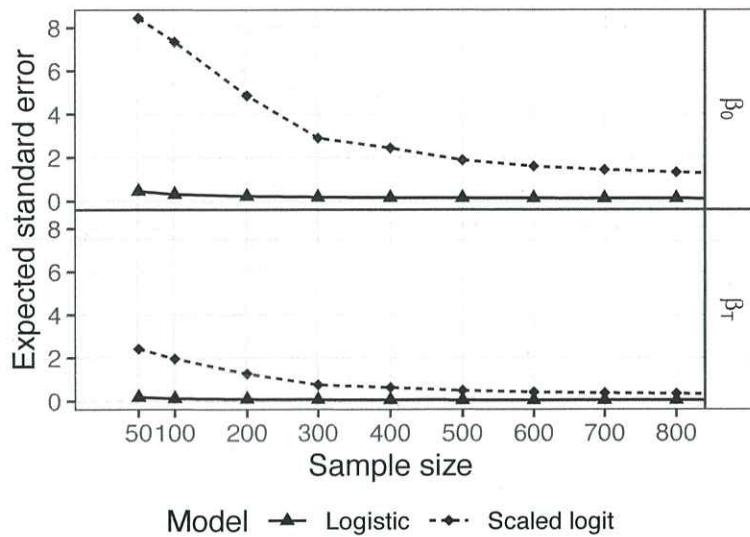


Figure 10: Simulation results of fitting the standard logistic and the scaled logit models to the same simulated datasets. 10,000 datasets were simulated. Shown is the mean standard deviation (estimate of the expected standard error) of the  $\beta$  regression estimates from the standard logistic (the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

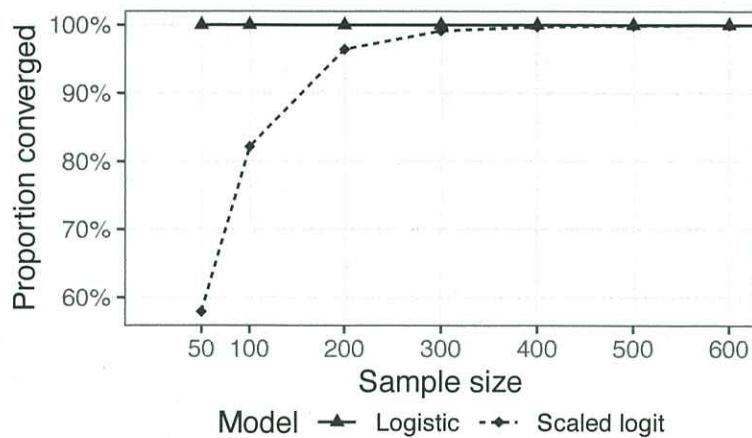


Figure 11: Simulation results of fitting the standard logistic and the scaled logit models to the same simulated datasets. 10,000 datasets were simulated. Shown is the proportion of successful attempts to fit the standard logistic (the solid line) and the scaled logit (the dashed line) model. Points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

what was the reason for nonconvergence?  
could you rectify by changing defaults on  
e.g., no. Iterations, tolerances, etc?

## 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.

not suitable

### 6.1 Scaled logit fit

The scaled logit model was fit in a Bayesian way to better handle missing data and censored titre measurements. The full specification is

$$\begin{aligned}
 & Y_i \sim \text{Bernoulli}(p_i) \\
 & p_i = \frac{\lambda}{1 + \exp(\beta_0 + \beta_T T_i)} \\
 & T_i \sim N(\mu, \sigma^2) \text{ truncated}([L_i, H_i]) \\
 & \mu \sim N(2, 2^2) \quad \sigma \sim \text{Exponential}(\text{rate} = 0.1) \\
 & \lambda \sim \text{Uniform}(0, 1)
 \end{aligned}$$

where  $\text{TN}_{[a,b]}(\mu, \sigma^2)$  is a Normal dist<sup>n</sup> truncated to  $[a, b]$ . Where  $i$  is observation index,  $Y_i$  is infection status (0 — infected (both symptomatic and asymptomatic), 1 — not infected),  $T_i$  is true log HI titre,  $L_i$  is the low bound of observed HI interval (e.g., for observation of 20 this is  $\log(20)$ ),  $H_i$  is the high bound of observed HI interval (e.g., for observation of 20 this is  $\log(40)$ ).

While I agree Bayesian is better, this comes completely out of the blue since everything else has been frequentist to now.

An obs. question from a reviewer would be: do the same issues occur in a Bayesian framework?

The prior distributions for the mean and standard deviation of log HI titres were chosen to be broad but still somewhat reflect the expected distribution in the general population

⊕ split model structure and prior distributions into sep. parts.

— mean high enough to allow a fair proportion ( $>20\%$ ) to have titres above detectable level ( $\log(10)$ ) and standard deviation low enough to not result in titres above highest detectable level ( $\log(1280)$ ) to be more prevalent than titres between  $\log(640)$  and  $\log(1280)$ .

*in  $[0,1]$*

The prior distribution for the baseline risk parameter  $\lambda$  gives equal probability to all ~~of its~~ possible values. The prior distributions for the logistic curve parameters  $\beta_0$  and  $\beta_T$  was ~~were~~ chosen to be spread around all of their plausible values —  $\beta_0$  is likely below 0 (otherwise the probability of infection would be close to 0 and not change across the observed titre range) and not too low (e.g. a value below -30 would result in the probability of infection being equal to  $\lambda$  and not changing across the observed titre range); while  $\beta_T$  is likely above 0 (since titres are protective) but not too high since it would result in a very steep curve (e.g. a titre of 20 may have the expected infection probability of  $\lambda$  while a titre of 40 has the expected infection probability of 0).

*Nice justification of priors!*

The fitted infection curve is shown in Figure 12. The fitted protection curve is shown in Figure 13.

*The fitted prob. of infection curves are shown in Fig. 13,  
and corresponding <sup>prob. of</sup> protection curves in Fig. 12.*

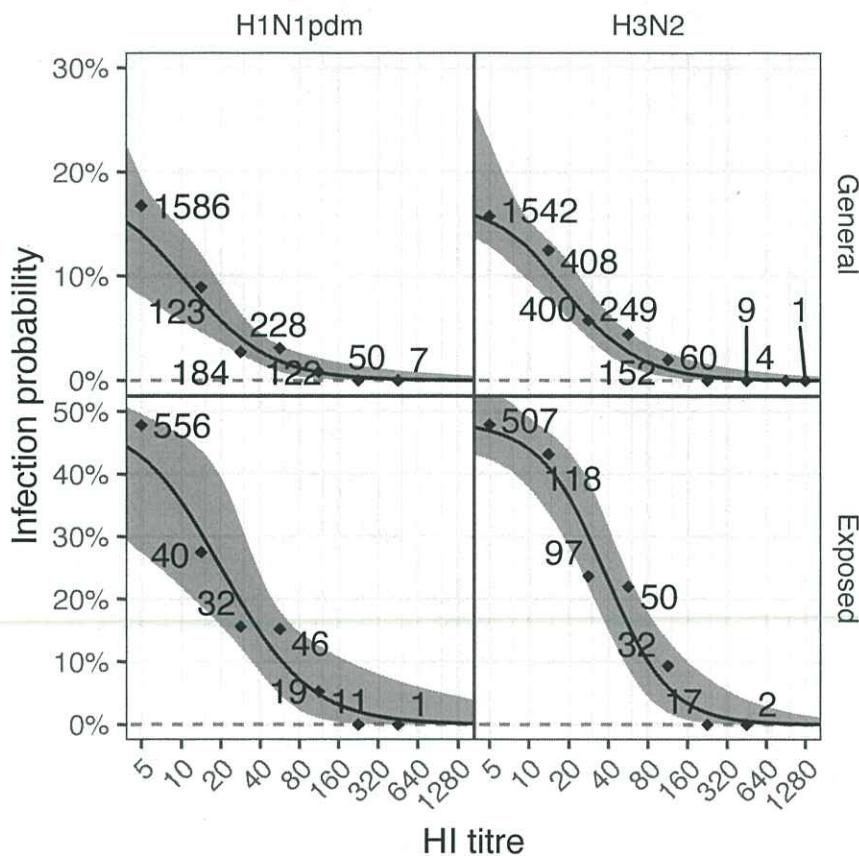
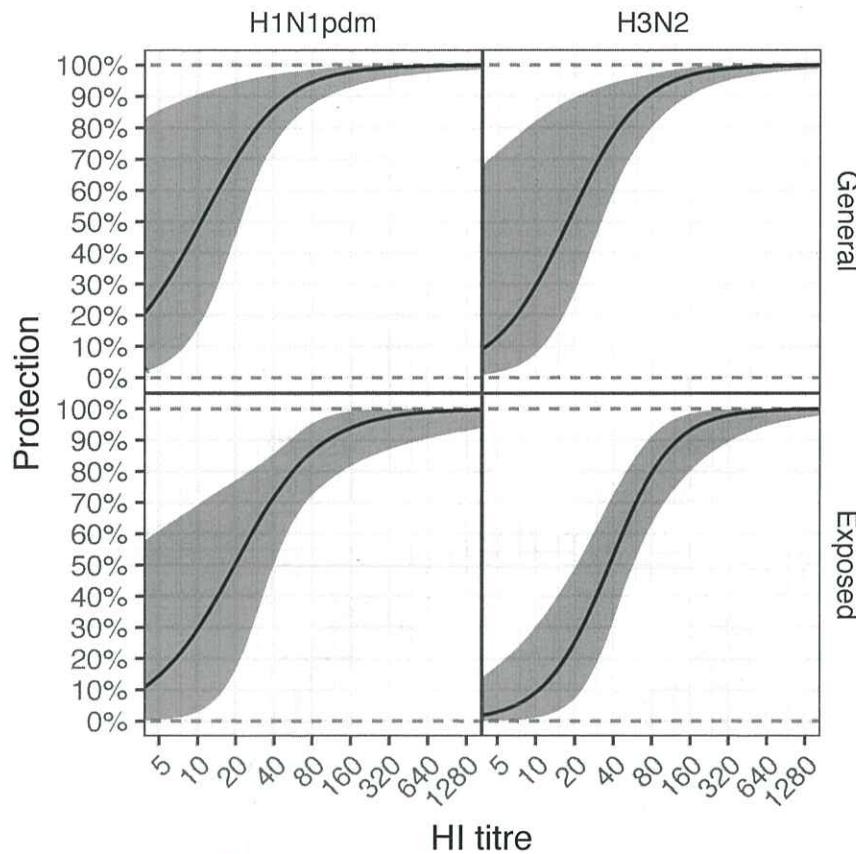


Figure 12: Fitted infection curves and credible intervals from the scaled logit model Bayesian fit to Ha Nam data (also shown in Figure 8). The solid line is the median of the posterior distribution. The shaded region is the 95% credible interval. The dashed line is the prior distribution, i.e. the bounds of the shaded region that would have been obtained if the data contained no information to estimate model parameters. The upper bound is above 90% and therefore not visible. The points are the infected proportions at the corresponding titre measurements. The numbers next to the points are the total sample size of the corresponding groups. Note that the titre measurements were shifted to midpoints of the corresponding intervals to better represent average underlying titres, e.g. the measurement of 20 was shifted to the log-scale midpoint between 20 and 40 (except the measurement of 5 which represents undetectable titres and the measurement of 1280 which represents titres above the highest detectable level).

is there  
also a  
dashed line  
at 1?  
Perhaps not  
helpful here  
could generate  
a figure of  
the prior  
Sep. for  
appendix?

was this just for the Fig., or in the analysis? You said earlier that the Bayesian approach was adopted to better deal with interval-censoring, so I assume Fig. only?



*prob. of*  
Figure 13: Fitted protection curves and credible intervals from the scaled logit model Bayesian fit to Ha Nam data (also shown in Figure 8). The solid line is the median of the posterior distribution. The shaded region is the 95% credible interval. The dashed line is the prior distribution, i.e. the bounds of the shaded region that would have been obtained if the data contained no information to estimate model parameters.

Limiting the sample to just those exposed to the virus (at least one household infection in a season) improved the precision of the estimates. Despite a fairly large sample, reasonably precise estimates of 50% protection can only be obtained for the H3N2 virus (inferring from the model fitted to the exposed subset). 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 credible 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 14. (ba)

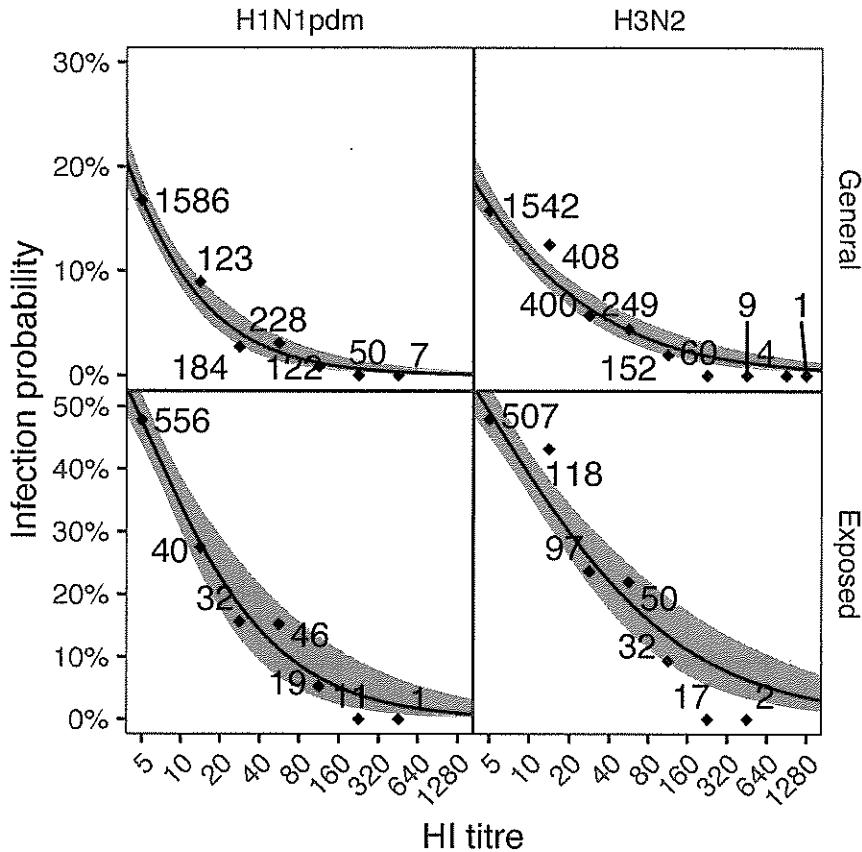


Figure 14: Fitted infection curves and confidence intervals from the standard logistic model fit to Ha Nam data (also shown in Figure 8) using maximum likelihood with no accounting of censored titres (observations of 5 (below detectable) and 1280 (above detectable) were unchanged, all other observations 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.

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 15.

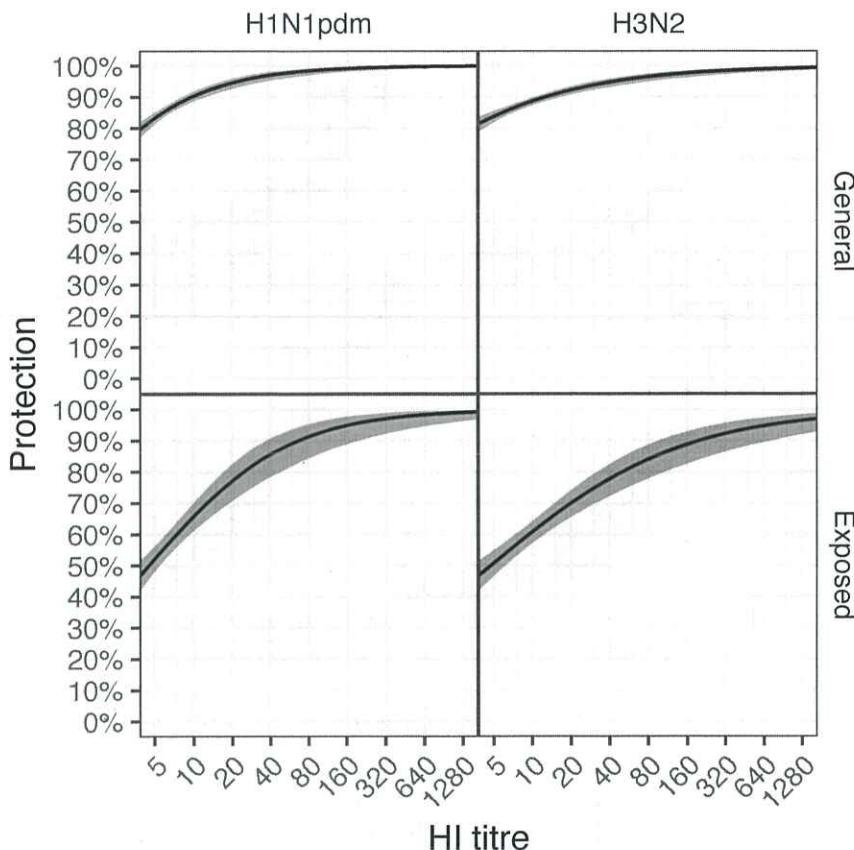


Figure 15: Fitted protection curves and confidence intervals from the standard logistic model fit to Ha Nam data (also shown in Figure 8) using maximum likelihood with no accounting of censored titres (observations of 5 (below detectable) and 1280 (above detectable) were unchanged, all other observations 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.

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). Subtracting these relative-to-5 infection probabilities from 1 generates curves shown in Figure 16

*This should appear before your first protection curve.*

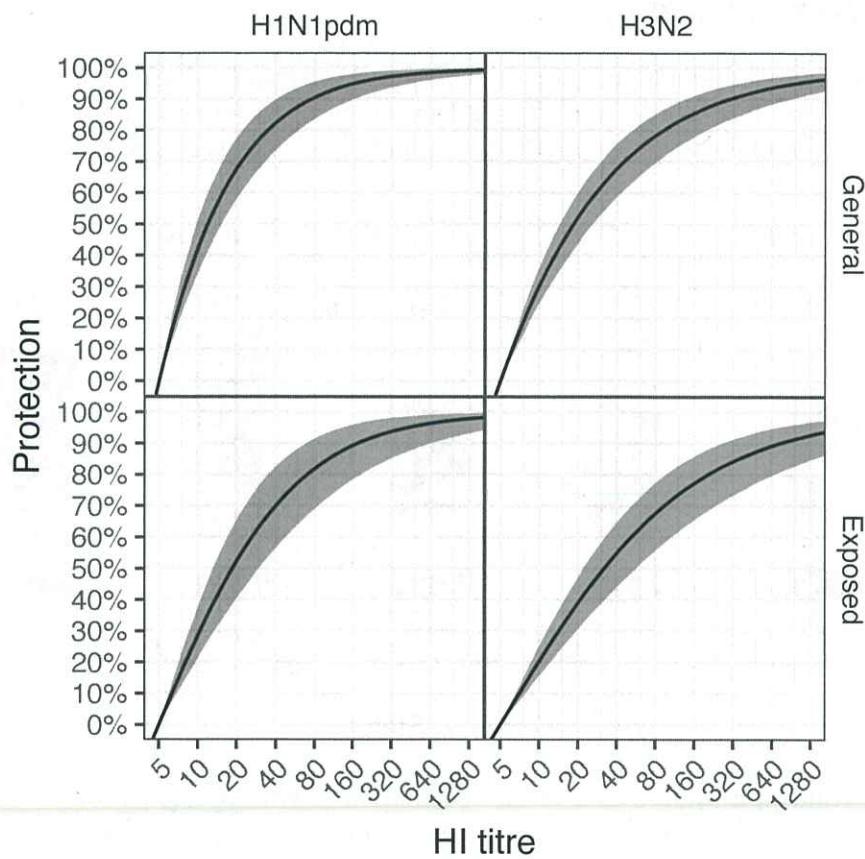


Figure 16: Fitted relative-to-5 protection curves and confidence intervals from the standard logistic model fit to Ha Nam data (also shown in Figure 8) using maximum likelihood with no accounting of censored titres (observations of 5 (below detectable) and 1280 (above detectable) were unchanged, all 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.

While the relative-to-5 protection curves (Figure 16) appear more plausible than the relative-to-baseline protection curves (Figure 15), 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 16) 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.

*→ This hampers the interpretability + comparison of results.*

I'd suggest that these results also suggest an increased confidence in the results (small errors), that is misplaced: a result of a strong assumption on the data that is unfounded.

## 7 Kiddyvax application

*Repeating*  
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 17.

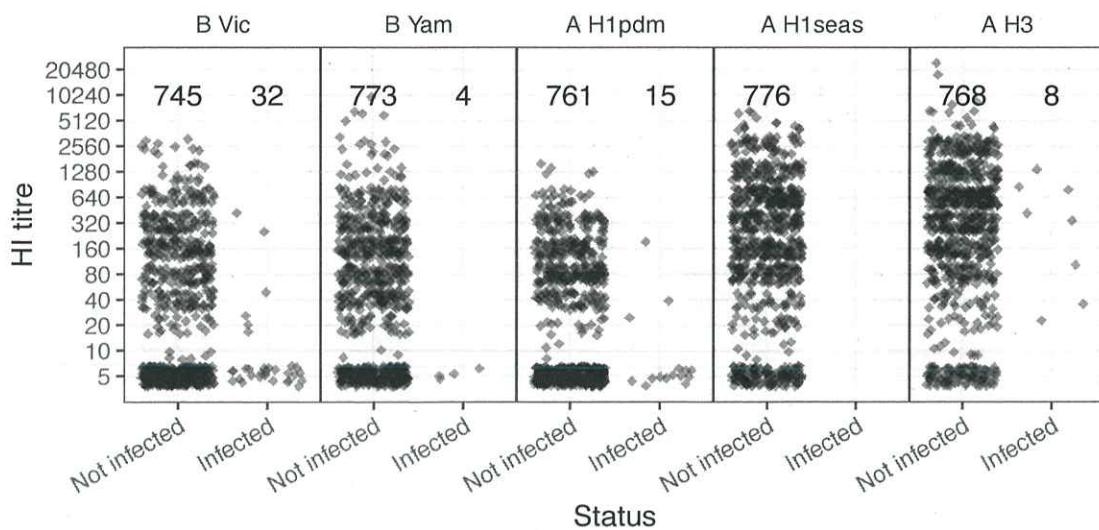


Figure 17: Kiddyvax study data. Post-vaccination titres are shown for those who got infected (PCR-confirmed symptomatic infection) over the course of the study and those who did not. Panels correspond to the five tested viruses. *B/Victoria, ...*

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.

### 7.1 Scaled logit fit

The same scaled logit model was fit in the same way using the same prior distributions as for the Ha Nam data. The protection curves are in Figure 18. The credible intervals are very broad due to the fact there are very few infections in the sample.

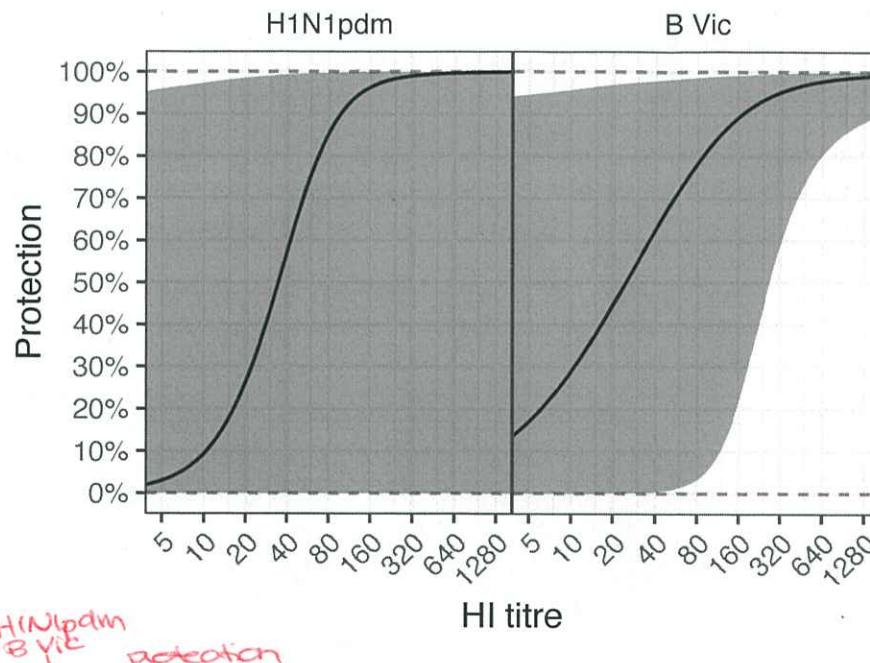
### 7.2 Standard logistic fit

The relative protection curves are in Figure 19. The same considerations regarding the unjustified assumption of baseline risk of 1 apply here as they did with Ha Nam (Figure 17 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 taken to be the time from start of follow-up to end of follow-up.

*The model was*



why does this  
act so bad?

Figure 18: Fitted infection curve and credible interval from the scaled logit model Bayesian fit to subsets of kiddyvax data (shown in Figure 17). The solid line is the median of the posterior distribution. The shaded region is the 95% credible interval. The dashed line is the prior distribution, i.e. the bounds of the shaded region that would have been obtained if the data contained no information to estimate model parameters.

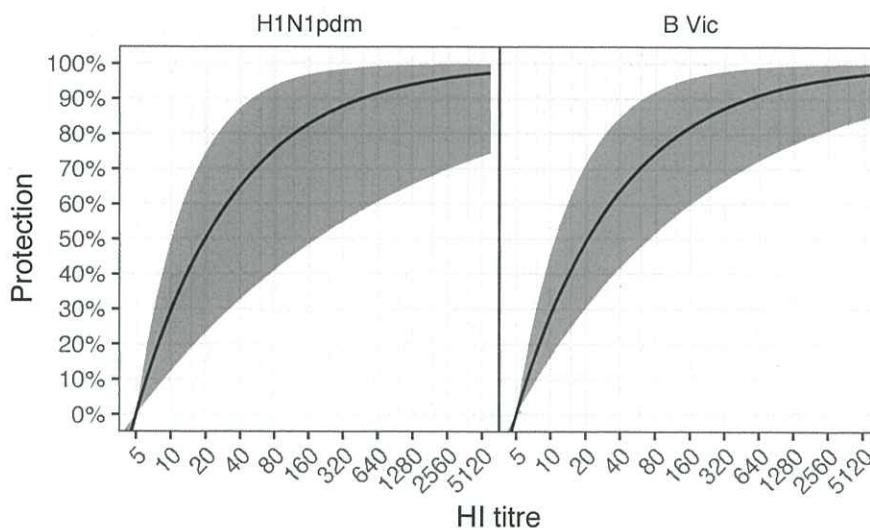


Figure 19: Fitted relative-to-5 protection curves and confidence intervals from the standard logistic model fit to kiddyvax data (shown in Figure 17) using maximum likelihood with no accounting of censored titres (observations of 5 (below detectable) were unchanged, all other observations were moved 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.

This should be earlier  
when you first  
introduce Cox PH.

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

where  $h$  is the hazard function and  $X_{\text{logtitre}}$  is the post-vaccination titre measurement on the log scale. The resulting protection curves (relative to the titre of 5) are in Figure 20. Although the point estimates are similar to those recovered from the model; however standard errors are inflated.

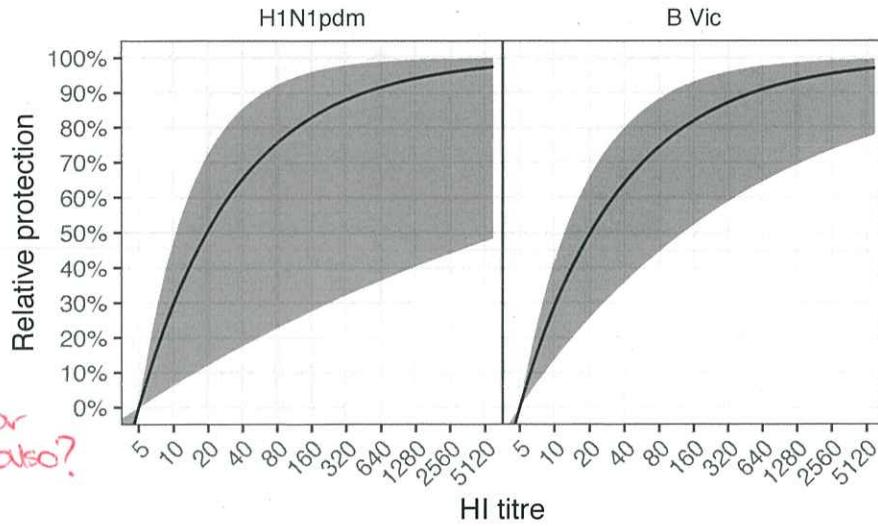


Figure 20: Fitted relative-to-5 protection curves and confidence intervals from the Cox proportional hazards model fit to kiddyvax data (shown in Figure 17) with no accounting of censored titres (observations of 5 (below detectable) were unchanged, all other observations were moved to the midpoints of 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 <del>immunity</del> <sup>infection</sup>
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.

Need to show protection curves for the sim studies so we can make that comparison.

high (near) probability of infection

bit vague.

## Bibliography

- [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>.

All code is in [github.com/khvorov45/model-comparison](https://github.com/khvorov45/model-comparison)