Protection curves of Ng 2013

Arseniy Khvorov

11/12/2019

The published protection curves in Ng (2013) were not generated correctly.

The paper in question is

Ng S, Fang VJ, Ip DK, et al. Estimation of the association between antibody titers and protection against confirmed influenza virus infection in children. J Infect Dis. 2013;208(8):1320–1324. doi:10.1093/infdis/jit372

And can be accessed here

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3778972/

The data and the code from the paper are under the header "9. The association between antibody titers and protection against influenza" here

http://web.hku.hk/~bcowling/influenza/kiddivax_study.htm

1 Data

The data is from the Kiddyvax study. Subjects were recruited, vaccinated and monitored for a year for flu infection. The relevant data are their post-vaccination HI titres, whether they got infected (PCR-confirmed) while they were followed up and when the infection occured.

2 Published results

The published results are from the Cox proportional hazards model fitted separately to the H1 and B viruses. The model is of the form

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

Where h is the hazard function, t is timepoint, h_0 is baseline hazard, X_{logtitre} is the individual's log HI titre at time t and P is influenza risk (proxy measure) at time t.

They had a waning titre model to estimate $X_{\text{logtitre}}(t)$.

The proxy measure for flu activity $P(t)$ was extrapolated from here

http://web.hku.hk/~bcowling/influenza/data/influenza_proxy_1998to2013.csv

The time-varying covariates were handled by creating multiple rows per individual where each row is a day of follow-up. Each row then contains the individual's estimated titre on that day, estimated flu proxy on that day and whether the individual got infected on that day. For those who got infected, the day they got infected was their last row in this trasformed data.

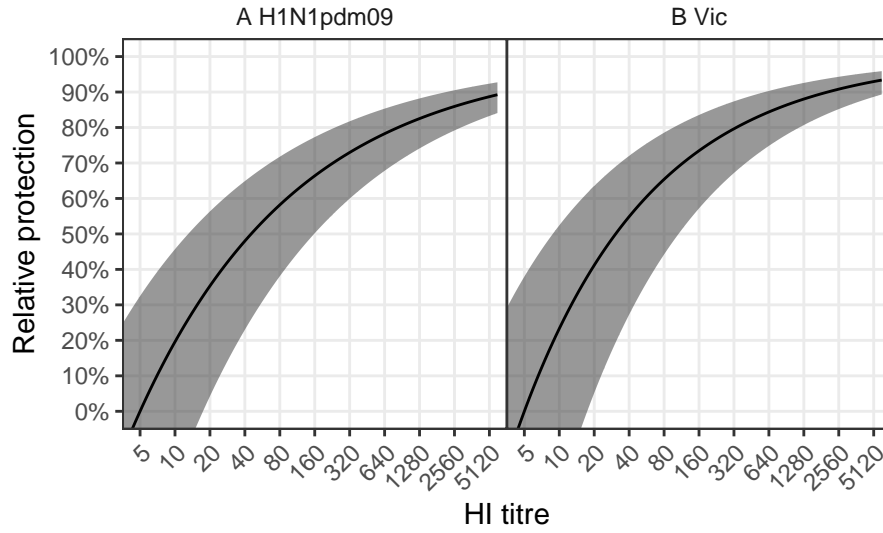The published protection curves are reproduced in Figure 1

Figure 1: Published protection curves of Ng (2013)

3 What I did

I took the transformed data to be my starting point and I used the same model. This means that there are potentially questionable things that I did not question, these include:

- Extrapolation of flu proxy
- Inclusion of flu proxy into the model
- The waning titre model
- Using a waning titre model with no account of censored observations

4 The problem with the original

The code that generated one of the original curves is below (the same code generates the other curve, just with the variable names changed)

```
beta <-vic.vic$coef
se <-sqrt(sum(vcov(cox.td.B)))
est <-data.frame(titer=log2(5*2^(0:9)),proxy=2, mu=NA, ci1=NA, ci2=NA)
est$mu <-exp(beta[1] * est$titer + beta[2]*est$proxy )
est$ci1 <-exp(beta[1] * est$titer + beta[2]*est$proxy  - 1.96*se)
est$ci2 <-exp(beta[1] * est$titer + beta[2]*est$proxy + 1.96*se)
est$d <-(est$mu[1]-est$mu)/est$mu[1]
est$d.ci1 <-(est$mu[1]-est$ci2)/est$mu[1]
est$d.ci2 <-(est$mu[1]-est$ci1)/est$mu[1]
```

The problem is this line

```
se <-sqrt(sum(vcov(cox.td.B)))
```

It sets the standard error to a number (square root of the sum of entries in the variance-covariance matrix) which is used to generate the interval bounds across the entire covariate range.

The linear predictor l part of the model has this form

$$l = \beta_T X_{\text{logtitre}} + \beta_P P$$

What we want is the variance of l which is

$$V(l) = X_{\text{logtitre}}^2 V(\beta_T) + P^2 V(\beta_P) + 2X_{\text{logtitre}} P \ Cov(\beta_T, \beta_P)$$

Which is only equal to the sum of entries of the variance-covarinace matrix when both X_{logtitre} and P are 1.

5 Correcting confidence intervals

The quantity that we are looking for is protection r which is

$$r = 1 - \exp(l)$$

Note that the curves in Figure 1 are supposed to show protection relative to the titre of 5.

So then the target quantity is

$$r = 1 - \exp(\beta_T(X_{\text{logtitre}} - \log(5)) + \beta_P P)$$

The code chunk in Section 4 extracts this by doing this wierd manipulation

```
est$d <-(est$mu[1]-est$mu)/est$mu[1]
```

Which is

$$r = \frac{\exp(\beta_T \log(5) + 2\beta_P) - \exp(\beta_T X_{\text{logtitre}} + 2\beta_P)}{\exp(\beta_T \log(5) + 2\beta_P)}$$
$$= 1 - \exp(\beta_T(X_{\text{logtitre}} - \log(5)))$$

Meaning P (flu proxy) is set to 0, so the variance of the linear part of this quantity is simply

$$V(r^{linear}) = (X_{\text{logtitre}} - \log(5))^2 V(\beta_T)$$

So both the quantity and its correct variance could have been extracted from the model fit directly

```
est$r_linear <- coef(vic.vic)[[1]] * (est$titer - log2(5))
est$r_var <- (est$titer - log2(5))^2 * vcov(vic.vic)[1, 1]
```

And then used to generate probability-scale quantities

```
est$d <- 1 - exp(est$r_linear)
est$d.ci1 <- 1 - exp(est$r_linear + qnorm(0.975) * sqrt(est$r_var))
est$d.ci2 <- 1 - exp(est$r_linear - qnorm(0.975) * sqrt(est$r_var))
```

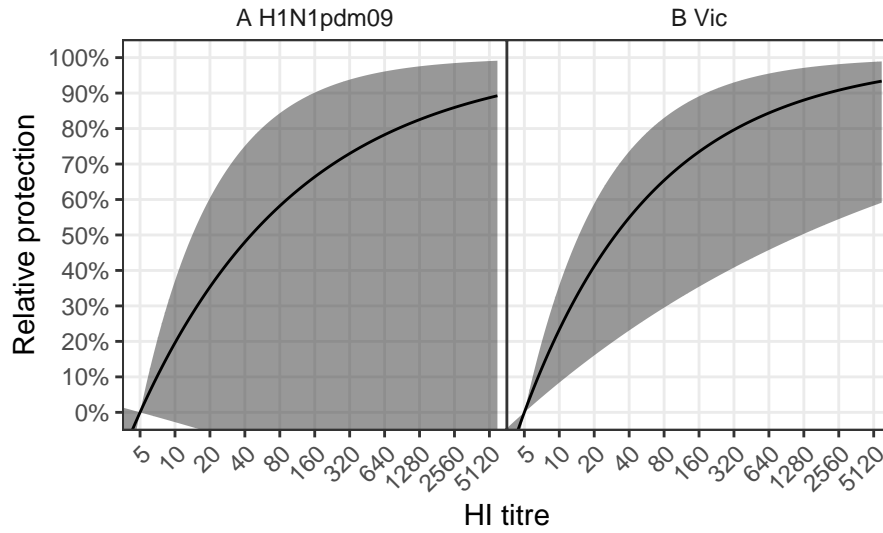The curves that result from this correction are in Figure 2

Figure 2: Ng (2013) curves but with corrected CIs

6   Correcting model fit

There is another problem with the above intervals which is how the model was fit.

```
vic.vic<-cox.td.B <- coxph(Surv(t,event)~postvax.b+proxy,data=kdata.td.b)
```

This assumes that every row from the dataset is an independent right-censored observation. There are 773 observations in `kdata.td.b` but 244,382 rows. So fitting a model like this implies that we have 244,382 subjects in the study who we followed anywhere from 0 to 413 days.

To fit this properly we need to create the start and the end timepoints for each row and cluster on individual id.

```
library(dplyr)
kdata.td.b <- kdata.td.b %>%
  group_by(hhid) %>%
  mutate(
    start = t,
    end = lead(t, default = max(t) + 1)
  ) %>%
  ungroup()
vic.vic <- coxph(
  Surv(start, end, event) ~ postvax.b + proxy + cluster(hhID),
  data = kdata.td.b
)
```

The protection curves that result from the corrected model call and the corrected intervals are in Figure 3.
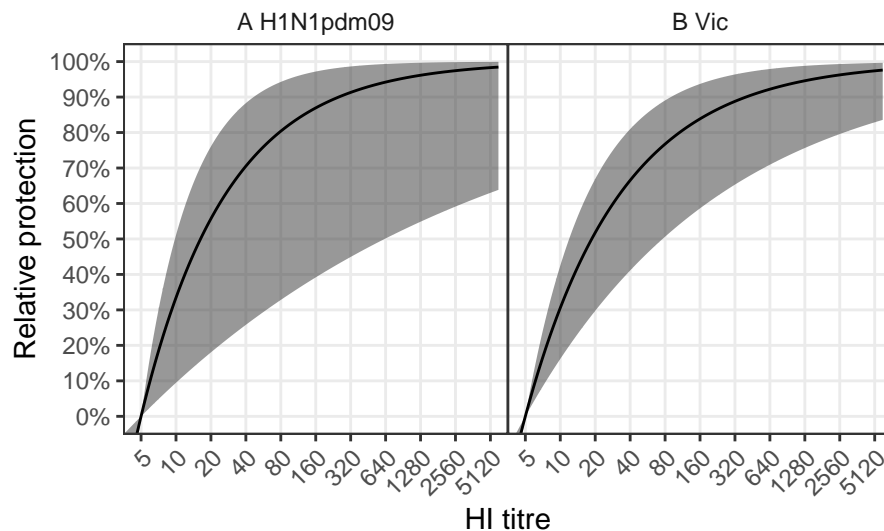

Figure 3: Ng (2013) curves with corrected fit and CIs

The curves I generated by fitting a simple one-covariate one-row-per-observation right-censored Cox model with no proxy or waning titres are in Figure 4. They look remarkably similar to Figure 3.
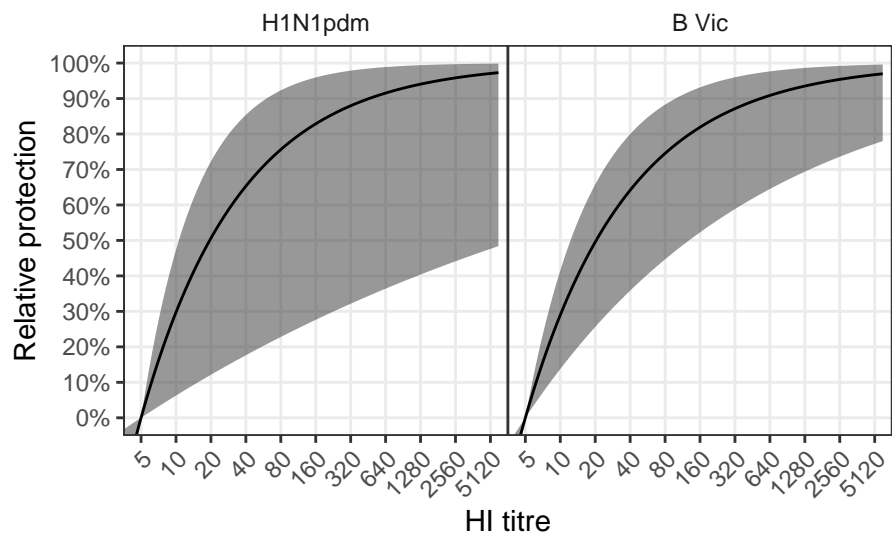
Figure 4: Protection curves from simple one-covariate one-row-per-observation right-censored Cox model with no proxy or waning titres