Report Sheet

Reproduction of Study: Influenza B Vaccine Protection Mediated by Hemagglutination-Inhibition Antibody Titer

Amulya Vemuri, Avery Howard, Bridget Perez, Jenny Surgent

University of Georgia

EPID 4070

Introduction

Influenza B, one of the two strains of the influenza virus responsible for the flu, contributes to a significant rate of morbidity and mortality, particularly in older adults and children. Vaccination used to combat Influenza B, remains as the most effective strategy for virus prevention and severity. However, implantation like vaccine-induced immunity have been recently shown to provide protection but have needed further research (NIH 2024). Hemagglutination-inhibition (HI) antibody titers are generally used as the primary measure of immune response and vaccine efficacy (NIH 2024). Despite their established role in assessing immunogenicity, HI titers direct mediate protection against influenza B remains an area that needs continued research (NIH 2018). The paper examines the relationship between post-vaccination HI titers and protective immunity using data from a randomized clinical trial, placed into R to reproduce the data placebo and non-placebo controlled trials as well as the influenza B vaccine works against hemagglutination-inhibition (HI) antibody titers. Understanding these correlations is essential in understanding vaccine recommendations and strengthening public health preparedness for Influenza B. By determining the limits of HI antibodies that offer protection, researchers can change the vaccine recommendations and improve the overall health strategies.

Methods

The study was a randomized, placebo-controlled trial conducted in Hong Kong between 2009 and 2010 involving children 6–17 years old. Participants were randomly assigned in a 3:2 ratio to receive either a trivalent inactivated influenza vaccine (VAXIGRIP) or a placebo (saline). To measure the children's immune response blood samples were taken before vaccination and one month after. Hemagglutination inhibition (HAI) examinations were used to assess antibody levels. Active surveillance continued until December 2010 with biweekly phone check-ins and home visits for the sick participants, here nose and throat swabs were collected for PCR-confirmed influenza detection. Ethical approvals were secured, with parental consent and written assent required for older children. The statistical analysis focused on understanding how vaccination affected both antibody titers and protection against influenza, differentiating between direct effects (not linked to HAI titers) and indirect effects (mediated through immune response). Regression and Cox proportional hazards models were used, adjusting for factors like age and inverse probability weighting helped estimate the direct impact of vaccination. To ensure reliability, bootstrap resampling (10,000 iterations) was conducted, and sensitivity analyses considered pre-vaccination antibody levels. All analyses were performed using R version 3.4.3.

The results section of the paper began with a plethora of descriptive statistics that ranged from the number of participants to the number of children in which influenza B/Victoria infections were determined. The authors of this study determined these statistics through a series of codes separated into groups of enrollment data, age distribution, infection counts, and vaccine-specific B/Victoria infections. This was then followed with a standard print results section of code. To produce our version of the code, we started by asking Deepseek AI to isolate the part of the code they gave in their file “Results and Figure” that resulted in the information.

After asking Deepseek to adjust for the working directory we were using and the file name I was pulling from, we also asked Deepseek to go a step further with the data. The original study did not have a tabular format for this information so we asked Deepseek to add lines of code that would arrange the data we got from it. The code accomplished this by creating 4 sections for the table to categorize the data by – enrollment flow, mean age by group, total infections by virus type, and B/Victoria breakdown by vaccine; it used the package knitr to create the table. With these amendments made, the first main error we encountered was in the code being able to read the dates in the postvax.date and vaccine.date column due to misinformation. This was rectified with explicit date conversion as.Date() and date formatting %Y-%m-%d, a new column for cleaner filtering. We also ran into issues with the if statement structure. To fix this, the infection data check was separated into a logical variable first, and the column existence checking was simplified. These were the main errors in producing the result values barring errors with indentation or missing parenthesis which were also fixed with Deepseek. Most of the details required were provided by the paper's authors, but due to the issues mentioned above we were not able to entirely rely on their code without using Deepseek to make changes.

Following this, we reproduced the graph from Figure 2B of the paper. The IIV and placebo injection from the graph were administered between 18 September 2009 and 22 January 2010, and the post-vaccination sera were collected between 16 October 2009 and 20 February 2010. 92% of these samples were collected before 31 January 2010. The code used by the authors uses Kaplan-Meier Model fitting with a Base R as its plotting method. It also manually defines the time axis and the percentage labels in the y axis. In our reproduction of the graph, we utilized two AI engines – Chat GPT and Deepseek – and we did not use the code given in the paper. Instead a starting code was achieved by giving Chat GPT access to the data from our study

as well as the description of the graph. This initial code had errors in terms of not being equipped to handle missing data and the code also did not initially use the right model to create the graph. The code also included extra months in the x-axis. These issues were rectified by asking Deepseek to fix any errors, and we were able to replicate the graph with minimal resulting error. The biggest fix was using (ggplot2) – geom_step() – to create the curve as well as scale the axes. There was a recurring warning of 45 rows with missing values or values outside the scale range being given, caught by the geom_step(), which did not hinder the production of the graph. Since we attempted making this graph on our own, there was not a lot of information that was needed from the study that was not available, but we would have preferred that the study mentioned that the graph was made with so much missing data.

As for the hazards ratio pertaining to this graph specifically, the study utilized a Cox proportional hazard model with R to get their resulting HR for children who received IIV compared to placebo (while adjusting for age). In our attempts, we asked Deepseek to isolate the part of the code they used from their "Results and Figures" file to calculate this statistic and adjusted for the working directory and file type that the code was pulling from – csv to excel. From the mostly original code, we received an error in 'mutate()' where the R was not able to determine the date format in the Excel file. To fix any redundant errors that had already been fixed in the previous coding for the graph, we uploaded the code from the graph that worked and instructed Deepseek to use that as a model of a working code and to adjust the code to fix any errors based on the working code. The resulting code gave us our HR value for this figure with minimal feedback and fixing from R. For this section of reproduction there was no information missing from the study that did not allow us to reproduce the results especially with a base code that already functioned with our data file.

Finally, to reproduce the HR for direct and indirect effect the process was as follows. To calculate the HR for direct effect, they used a logistic regression model where vaccination was the response variable with post-vax HAI titer and age as the predictors. Then, they used the estimated coefficients of the model to predict the odds ratio of vaccination for each participant. After this, they constructed weights for the vaccinated participants of the study as an inverse of the predicted odds ratio while specifying weights of 1 for all the unvaccinated participants. A proportional hazards model was fitted to the calendar time of inflection and each observation was weighted by the weight derived in the previous step; age and receipt of IIV/placebo were adjusted for. This gave them the direct effect as the HR for IIV. The indirect effect was found by doing a ratio of the total effect and the direct effect, (the total effect being the HR of vaccination). They used bootstrapping with 10000 resamples to estimate the uncertainty in the total, direct, and indirect effect. These analyses were, as mentioned, done using R. For our reproduction of these results we used the code they had in their "Mediation Analysis" file and asked Deepseek to adjust it to pull from our data file and working directory. This direct entry of their code once again resulted in a date formatting error fixed by clarifying the date format along with a coxph function error which was resolved by reloading the survival package. Then, we realized the code in the "Mediation Analysis" file pulled from the information/code in "Mediation functions", and we asked Deepseek to incorporate the file into the code resulting in a longer version of the code. Doing this resulted in a persistent error in charToDate(x): character string is not in a standard unambiguous format. This required additions of debugging code and changing the code to include a universal parser, robust cleaning, and a flexible output. This then led to R showing the "First 6 raw date strings" for us to verify before we were able to make it work. This gave us some values, but they were quite different from the given values in the paper

prompting us to check for error; this led to Deepseek changing the code to follow a more formal mediation analysis framework using the mediation package. This was followed by errors with the mediation package and the boot package. Once these were rectified through Deepseek, we realized that the code was not pulling from any data source directly, so we rectified that by asking Deepseek to modify the code to pull directly from our file with the data. These values were also largely different to the given values, so we had it analyze why that could be. As a way to try and produce more precise results we requested Deepseek to use the method it thinks best applies to this study. When we requested this, Deepseek changed the method to the product method which is more standard in Cox models for calculating HR. Once this change was made, we had some error in drop_na() since there were some columns referenced that did not exist in our data. We fixed this by bettering the diagnostics and including more thorough data checking. Then, due to some issues with the select() function we pushed Deepseek into using a code that avoids any version-specific dplyr functions and instead relied on base R. We then had to remove dependency on HC3 standard errors and switch to a simplified bootstrap approach that does not require hat values due to an error in meatHC(x,type = type, omega = omega) since hatvalues() could not be extracted. We also had to switch to manual CI using the percentile method for this. The original code for this did not work well with our R given that it was equipped with less packages and faced many errors. While there is nothing else the paper could have provided in terms of information, the code was not accessible in terms of reproduction.

Results

For the descriptive table, we were able to reproduce most of the information, but some values differed. One key discrepancy was in the withdrawal before intervention, where we got 0, while they reported 1—possibly due to it occurring after randomization. In doing so, the code referenced A1, N1, PDM 09, and 8H3, along with two variables listed as the fifth and eighth in their information section. However, when running the modified code, our descriptive table did not generate those values, likely because the dataset we used did not contain those specific columns. Since the code could not retrieve that information, this aspect of the table was non-reproducible.

Our biggest difference came from the hazard ratio, which was the key figure in our analysis. While we used their code as a starting point, we had to make adjustments within DeepSeek to get it to work properly. For the HR of Figure 2B, initially, their code didn't function for us, but after modifying it, we obtained a similar value—0.31 compared to their 0.32—suggesting that slight differences in data handling, such as an extra value being included or excluded, could explain the variation. For example, their direct effect hazard ratio (HR) was 0.6, while ours was 0.63, and their indirect effect HR was 0.52, whereas we obtained 0.377, a notable difference. Since the indirect effect was calculated using different methods, this part showed the most variation, while the direct effect differences were likely due to differences in our approach and data processing. Ultimately, we had to modify the code to make it work in DeepSeek, aligning our approach with a method better suited for epidemiological research. While our results weren't drastically different overall, the methodological variations led to significant differences in specific aspects of the analysis.

Figure 1. Reproduced Graph of Figure 2B

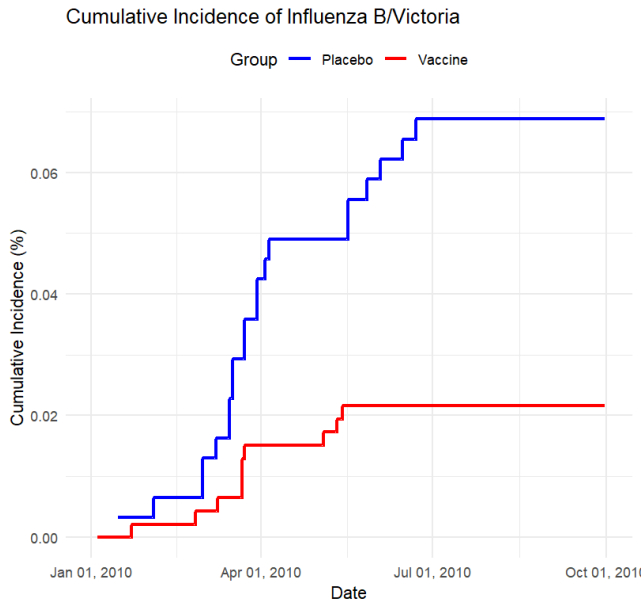

Figure 2. Original Figure 2B Graph

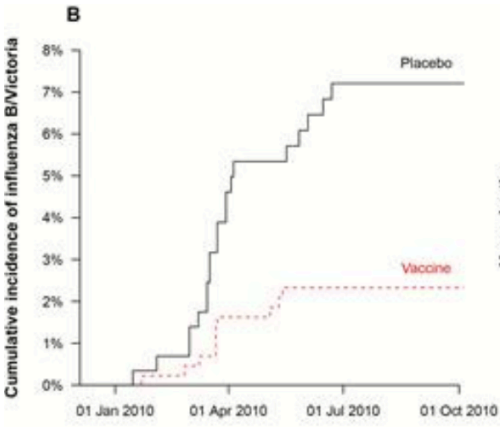

Figure 3. Table of Descriptive Statistics

Parameter	**Value**
Total randomized	796
Withdrew before intervention	0
No valid blood sample	0

Final analysis sample	796
Mean age- Placebo (years)	10
Mean age-IIV (years)	10
B/Victoria in infections	31
B/Victoria in placebo	21/317 (6.6%)
B/Victoria in IIV	10/479 (2.1%)

Discussion

Our conclusion is that this paper was overall kinda reproducible. To make it clear, our paper was not 100% reproducible for the reason that in our reproducing era, we were only able to successfully reproduce a key component of the paper such as the graph. We can not say that we were able to exactly reproduce the other two big factors of the paper – the descriptive table and the hazard ratios – since the numbers we got were not exactly equal to the paper's. When comparing our results and the results demonstrated in the paper, they generally supported the findings of the original study: the influenza B vaccine reduces flu risk through hemagglutination-inhibition (HAI) titers. However, there were differences with hazard ratios for the total, direct, and indirect effect we calculated as well as the descriptive statistics that we were able to produce using their data and the values that were published.

The code provided was able to produce a graph however they were not able to reproduce other key elements of this paper such as the exact hazards ratios that were reported. Since the code did not correctly run for certain parts of the paper, we had to write our own. The parts of the code that ran smoothly for the Kaplan-Meier graph were variables: Sruv_time, flu, group. Challenges in the code for the table included manual filtering, and there was no direct code for

Results_and_figures.R in the original to produce a table so we have to craft/make our own portion of the code for this to work in R. There are other things to take into account such as participants not getting the flu or dropping before the study ended that increases issues. The Cox model does account for these, however if not specified clearly when participants were censored can lead to variations in the data calculated and therefore not fully reproducible. Small differences in this type of work can compound and result in non-reproducible results even if the core analysis work looks straightforward. The exclusion criteria and filtering steps were not fully explained which made it more difficult to match the sample and results more accurately. The exclusion criteria and filtering steps were not completely explained in the paper so the code did not demonstrate how participants were removed/included.

A lack of clear exclusion criteria and filtering logic for this dataset made it difficult to know which participants were removed or why. Although the data set included values that hinted at which participants may have been excluded from this paper there's no clear explanation for why/how they were excluded. This makes it challenging to match the final sample used in the paper.

The main issue arose when calculating the hazard ratio for the direct effect, indirect effect, and total effect. They used the ratio method, whereas the product method is more commonly used in epidemiological literature, especially for handling censored survival data. The ratio method is more sensitive to small changes in data, making it more prone to variability, which significantly affected our results. A big issue arose when we were not able to use the code they had in their paper since it did not work when we tried to run it. This issue resulted in us not having the exact results such as their hazard ratios since we did not have the same code since it was not working and we had to develop our own code to get our results using the help of

deepseek and chatgpt. To have had the code that generated their results from the paper to work in our end, we could have exactly reproduced the data they got in their results dealing with their main components.

Based on our results, the vaccine is quite effective. This is shown by a hazard ratio (HR) of 0.2412 shows this since the value is significantly less than one, indicating that vaccinated individuals are much less likely to get flu compared to those who are not vaccinated. The direct effect evaluates the vaccine's effectiveness without taking antibody levels or other variables into account. With an HR of 0.6389, the level of protection is considered average. However, it can be deduced that this effect is not as strong on its own and that there is some uncertainty about it because the confidence interval crosses 1. The vaccine's capacity to raise flu-fighting antibody levels contributes to its effectiveness. With an HR of 0.3776 for this indirect effect, we can conclude that increasing antibody levels provides a respectable level of protection. It is evident from the constant confidence interval below 1 that these antibodies aid in flu prevention.

To make this study more reproducible the authors should have provided a fully documented and complete data set including variables such as metadata and more organized key columns. Better documentation of all data in the preprocessing steps such as inclusion.exclusion criteria and transformations could have been better. The code base itself needed to have better documentation. Difference in hazard ratio estimates means that the stat info wasn't super clear and described well enough. Having a sensitivity analysis would have clarified the results more clearly and reproducible.

Despite the challenges, our results still pointed in the same direction as the papers: the vaccine worked and the HAI titlers play an important role in that protection.

References

Ashraf, Muhammad Awais, et al. “A Comprehensive Review of Influenza B Virus, Its Biological and Clinical Aspects.” *Frontiers in Microbiology*, U.S. National Library of Medicine, 4 Sept. 2024, pmc.ncbi.nlm.nih.gov/articles/PMC11408344/.

Cowling, Benjamin J, et al. “Influenza Hemagglutination-Inhibition Antibody Titer as a Mediator of Vaccine-Induced Protection for Influenza B.” *Clinical Infectious Diseases : An Official Publication of the Infectious Diseases Society of America*, U.S. National Library of Medicine, 2 May 2019, pmc.ncbi.nlm.nih.gov/articles/PMC6495017/.

Links for code & data

Links for code & data:

Graph:
https://docs.google.com/document/d/1UR4MLTLQcrNWXLUQb8NyKP2PDfxCYW9Ek50hIq2kn6o/edit?usp=sharing

Descriptive Table:
https://docs.google.com/document/d/1Ver4K6L97DU5owAQ8N3zICfwpN1kAwpzxXr0ZtI0-7s/edit?usp=sharing

Causal Analysis - Product Method
https://docs.google.com/document/d/16JTFROInsNTc1M2uK9VQscJnqDpDmFYoczDXOmT6Qwk/edit?addon_store&tab=t.0

Hazard Ratio Code for Figure 2B and Results
https://docs.google.com/document/d/1Psc55HPj0Psx4OZ58_G5hF-9QdYdk_-29Uaq7vwdTjg/edit?tab=t.0