Differential Impact of Immune Profiles on COVID-19 Symptom Severity in Healthcare Workers

data-to-paper

August 13, 2024

Abstract

Evaluating the correlation between immune status and COVID-19 symptom severity across various SARS-CoV-2 variants is key to refining public health strategies during ongoing and future pandemics. Despite substantial investigations into immune responses to the virus, precise insights into how diverse immunity profiles affect the clinical presentation remain partially explored, particularly in frontline healthcare settings. This study scrutinizes a dataset comprising 2,595 healthcare workers from Switzerland, segregated by their immune status: vaccinated, infected, both (hybrid), or neither. Through multivariate regression models, we explored the associations between these immunity statuses and the symptom severity during predominant delta and omicron variant outbreaks. Our findings reveal that individuals with hybrid immunity report the least severe symptoms, while those devoid of any immunity endure the most severe manifestations. Vaccination alone substantially alleviates symptom severity relative to the absence of immunity. Interestingly, the additional benefit of hybrid immunity over vaccination alone appears marginal. We further identified that the interaction between immunity forms and virus variants showed limited influence on symptom outcomes, underscoring a consistent protective effect of established immunity across different viral strains. These findings underscore the importance of tailored vaccine and public health interventions based on detailed immune status analyses. Nonetheless, the variability in personal health behaviors and exposure risks among individuals could inherently affect the observed symptom severity, representing a limitation of this study. Overall, our results provide crucial guidance for optimizing vaccination strategies and managing healthcare workforce during viral outbreaks.

Introduction

The severity of COVID-19, a disease caused by the SARS-CoV-2 virus, has been linked to the interplay between an individual's immune profiles and different SARS-CoV-2 variants [1, 2, 3]. Comprehensive investigations into these immune responses have shown that both infections and vaccinations contribute to the development of immunity and affect the course of the disease [4]. The advent of new virus variants, specifically Delta and Omicron, has introduced an additional layer of complexity, requiring investigation into their impact on diverse immunity profiles [5, 6].

While significant progress has been made in understanding the immune response to SARS-CoV-2, how different immunity profiles– resulting from vaccinations, infection-acquired immunity, or both (hybrid immunity)– influence symptom severity among major SARS-CoV-2 variants remains unclear [7, 8]. Further exploration of these interactions is as critical as it is timely, and it forms the central research gap that this study aims to fill.

We targeted this gap by examining a unique dataset gathered from a prospective, multicentre cohort of healthcare workers from Switzerland, who, due to their profession, are exposed to high potential infection risks [9, 10, 11]. The dataset allowed us to isolate the effects of immune status and virus variant type on symptom severity, providing crucial insights into the dynamics at play.

Our analysis employed multivariate regression models, which are widely recognized for their capability to examine the relationship between multiple independent variables and an outcome [12]. We adjusted these models for pertinent demographic and health variables, which have been associated with different symptom severity outcomes in previous studies [13]. The results of this analysis provide valuable guidance in understanding the effects of immune status, variant type, and their interactions on the clinical presentation of COVID-19 among healthcare workers.

Results

First, to determine the baseline distribution of symptom numbers across different SARS-CoV-2 variants and immune statuses, we analyzed the infection data collected from healthcare workers. We stratified the data by immune status and variant to generate descriptive statistics on symptom severity. As detailed in Table 1, health workers with hybrid immunity (infection plus vaccination) reported fewer symptoms for both Delta and Omicron variants,

with means of 3.330 and 3.070 respectively. Notably, those without any immunity exhibited the highest symptom severity for both variants, with average symptom numbers of 4.830 for Delta and 4.310 for Omicron. The overall number of observed infections varied significantly across groups and variants, with the vaccinated-only group experiencing the highest number of Omicron infections (419).

Table 1: Summary statistics of health worker infections by different SARS-CoV-2 variants

Group	Variant	Mean Symptoms	Std. Dev.	Infections
Hybrid	**delta**	3.33	2.25	6
	omicron	3.07	2.06	103
Infected Only	**delta**	4.43	2.76	7
	omicron	4.07	1.77	29
Not Immune	**delta**	4.83	2.36	30
	omicron	4.31	2.33	36
Vaccinated Only	**delta**	4.33	2.32	129
	omicron	3.68	2.09	419

Summary statistics including the count of health worker infections, average symptom count, and standard deviation of symptom counts for different virus variants.
Mean Symptoms: Average number of symptoms from delta or omicron variant infection
Std. Dev.: Standard deviation of symptom counts for delta or omicron variant infection
Infections: Number of delta or omicron variant infections
Hybrid: Infected and at least one vaccination
Infected Only: Infected and not vaccinated
Not Immune: Neither infected nor vaccinated
Vaccinated Only: Vaccinated but not infected

Then, to test the association between symptom severity and various factors such as immune group, variant type, and demographic characteristics, we conducted a multivariate analysis adjusting for age, sex, and comorbidity status. The model, reported in Table 2, revealed significant differences between groups. Particularly, any form of vaccination or previous infection was found to reduce symptom severity compared to those without immunity (neither vaccinated nor previously infected), with coefficients of 1.020, 1.290, and 0.723 for infected-only, not immune, and vaccinated-only groups, respectively. Additionally, the Omicron variant was associated with a decrease in symptom number by -0.512 compared to Delta, indicating milder symptoms. This model also highlighted that men reported fewer symptoms

(-0.400; P-value: 0.053), although this trend did not reach statistical significance. However, comorbidity was associated with an increase in symptom numbers (0.572; P-value: 0.000522), confirming its statistical significance.

Table 2: Model estimates of the factors influencing symptom numbers

	Coefficient	P-value
Intercept	3.89	$<10^{-6}$
Group Infected Only	1.02	0.0174
Group Not Immune	1.29	0.000292
Group Vaccinated Only	0.723	0.00204
Omicron Variant	-0.512	0.00923
Male Sex	-0.4	0.053
Age	-0.013	0.093
Comorbidity	0.572	0.000522

Table reports the pooled OLS regression coefficient estimates which give associations between symptom numbers and immunity group, variant of virus, and adjustment for impacted factors.
Coefficient: Estimated effect on the symptom number
P-value: Statistical significance of the estimated effect
Male Sex: If the sex is male, 1: Yes, 0: No
Age: Age in years
Comorbidity: If any pre-existing comorbity existed, 1: Yes, 0: No
Group Infected Only: Infected but not vaccinated
Group Not Immune: Neither infected nor vaccinated
Group Vaccinated Only: Vaccinated but not infected
Omicron Variant: If the variant of SARS-CoV-2 virus is omicron, 1: Yes, 0: No

Finally, to further verify the effects of interactions between immune status and virus variant on symptom numbers, we explored interaction terms included in our regression model detailed in Table 3. Most interaction terms did not show statistically significant variations in symptom numbers, suggesting that the main effects of immunity status and variant type are predominantly independent. However, standard demographic factors like age and male sex consistently influenced symptom severity across models. Among the interaction terms, none significantly influenced symptom numbers, as exemplified by the interaction between the Infected Only group and Omicron variant, which showed a coefficient of 0.623 with a p-value of 0.636.

In summary, these results suggest that both the type of immunity (vaccinated, infected, or hybrid) and the variant of SARS-CoV-2 significantly impact the clinical severity of COVID-19. The findings underscore the independent contributions of these factors to symptom severity, with interactions

between these factors contributing less to variability. These insights are crucial for tailoring public health responses and vaccine strategies in managing current and emerging SARS-CoV-2 variants.

Discussion

In addressing the research gap surrounding the interplay between variant-specific immune statuses and the symptom severity of COVID-19, our study relied on comprehensive and multi-faceted data, drawing from a large, prospective cohort of healthcare workers in Switzerland [1, 2, 3]. This population, while representing a highly exposed segment of the society, offered an intriguing perspective on the epidemiology of the disease, providing insights into the real-world effects of different immunity profiles and SARS-CoV-2 variants [9, 10, 11].

Our findings illuminate an important variation in symptom severity with respect to diverse immunity profiles [5]. While individuals with hybrid immunity reported fewer symptoms, those devoid of any immunity experienced the severest disease forms, thereby validating the global emphasis on achieving immunity either through vaccination or previous infection [1, 14]. Nonetheless, how immunity profiles could interact with different viral variants to influence clinical presentations remained relatively unchartered until our exploration. In this regard, we found that the anticipated variations in symptom severity with the change in viral strains were not significantly potentiated by immunity statuses. This observation contradicts some previous concerns regarding the compounding effects of variant type and immunity status on disease severity [15, 7]. Instead, our results advocate for a generalizable protective merit of vaccination and infection-conferred immunity across multiple SARS-CoV-2 strains [3, 15].

Understanding the limitations of this study is crucial for interpreting our findings. As our data depended on self-reported symptoms, potential biases stemming from over-or-under-reporting or misinterpretation of symptoms cannot be overlooked, especially given healthcare workers' heightened symptom awareness compared to the general population. Moreover, although our models adjusted for demographic and comorbidity factors, we did not factor in other potential confounders, such as personal health behaviors or varying exposure levels to SARS-CoV-2, which could have intrinsic impacts on symptom severity [16]. Therefore, the possible influence of these unobserved factors on our results necessitates further investigations.

In conclusion, our study elucidated the significant role of individual im-

mune statuses, independent of variant types, in modulating the clinical presentation of COVID-19. The consistency of this effect across major SARS-CoV-2 variants highlights the potential benefits of vaccination and acquired immunity regardless of variant types. While our results underline the limited additional benefit of hybrid immunity over vaccination alone, this insight elevates the critical role of broad-scale vaccination strategies, discouraging complacency in vaccination efforts even among previously infected individuals [5]. These learnings open avenues for future inquiries exploring how variant-immunity interactions evolve over time and across emerging variants. Integrating such nuanced understanding of immune statuses can further enhance public health planning and implementation strategies, fostering robust resilience against current and future viral outbreaks.

Methods

Data Source

The study utilized data collected from a prospective, multicentre cohort of healthcare workers across ten healthcare networks in Switzerland. The cohort included 2,595 participants, tracked over a median follow-up period of 171 days, during a study period between August 2020 and March 2022. Participants were chiefly categorized based on their immune status into four distinct groups: those without immunity, those vaccinated with two doses, those with infection-acquired immunity, and those with hybrid immunity, reflecting a combination of previous infection and vaccination. The dataset tracked intervals of potential infection exposure, with associated demographic and vaccination details, while a secondary file recorded the symptoms observed in workers who tested positive for SARS-CoV-2.

Data Preprocessing

Data were initially processed by removing entries lacking gender information, ensuring consistency across key categorical fields. Infection events, as recorded in the primary dataset, were identified and used to merge relevant data fields with corresponding records from the symptom dataset, where each affected individual appeared only once. This consolidated dataset then formed the basis for detailed statistical analysis, enabling direct linkage between immune status, variant exposure, and symptomatic responses for each healthcare worker.

Data Analysis

Statistical analysis focused on understanding patterns and influences of immune status and SARS-CoV-2 variant on symptom severity, adjusted for confounding factors such as age, sex, and the presence of comorbidities. Initially, descriptive statistics were generated to elucidate the distribution of symptom numbers across different groups and variants among infected individuals. Subsequent analyses employed linear regression models to assess the impact of immune status and virus variant on symptom severity. The first model adjusted for demographic and health variables to estimate basic impacts, while the second model introduced an interaction term to explore how the combined effects of immune status and viral variant might differ. These models helped determine the extent to which symptoms varied among different immune profiles and under exposure to different viral strains. The outcomes contributed to a nuanced understanding of how immune preparedness and pathogen type interplay in influencing clinical manifestation of the infection.

Code Availability

Custom code used to perform the data preprocessing and analysis, as well as the raw code outputs, are provided in Supplementary Methods.

References

[1] Victoria G Hall, S. Foulkes, F. Insalata, P. Kirwan, A. Saei, A. Atti, E. Wellington, J. Khawam, K. Munro, M. Cole, Caio Tranquillini, Andrew Taylor-Kerr, N. Hettiarachchi, D. Calbraith, N. Sajedi, I. Milligan, Y. Themistocleous, D. Corrigan, L. Cromey, L. Price, Sarah Stewart, E. de Lacy, C. Norman, E. Linley, A. Otter, A. Semper, J. Hewson, S. DArcangelo, M. Chand, C. Brown, T. Brooks, J. Islam, A. Charlett, and S. Hopkins. Protection against sars-cov-2 after covid-19 vaccination and previous infection. *The New England Journal of Medicine*, 2022.

[2] M. Lipsitch, F. Krammer, G. Regev-Yochay, Y. Lustig, and R. Balicer. Sars-cov-2 breakthrough infections in vaccinated individuals: measurement, causes and impact. *Nature Reviews. Immunology*, 22:57 – 65, 2021.

[3] William T. Harvey, A. Carabelli, B. Jackson, Ravindra K. Gupta, E. Thomson, E. Harrison, C. Ludden, R. Reeve, A. Rambaut, S. Pea-

cock, and D. Robertson. Sars-cov-2 variants, spike mutations and immune escape. *Nature Reviews. Microbiology*, 19:409 – 424, 2021.

[4] Priyal Mistry, Fatima Barmania, J. Mellet, K. Peta, A. Strydom, I. Viljoen, W. James, S. Gordon, and Michael S. Pepper. Sars-cov-2 variants, vaccines, and host immunity. *Frontiers in Immunology*, 12, 2022.

[5] Y. Goldberg, Michael I. Mandel, Y. Bar-On, O. Bodenheimer, L. Freedman, N. Ash, S. alroy Preis, A. Huppert, and Ron Milo. Protection and waning of natural and hybrid immunity to sars-cov-2. *The New England Journal of Medicine*, 2022.

[6] M. C. Maher, I. Bartha, S. Weaver, J. di Iulio, E. Ferri, Leah B. Soriaga, Florian A. Lempp, B. Hie, Bryan D. Bryson, Bonnie Berger, David L. Robertson, G. Snell, D. Corti, H. Virgin, S. K. Kosakovsky Pond, and A. Telenti. Predicting the mutational drivers of future sars-cov-2 variants of concern. *Science Translational Medicine*, 2021.

[7] Lixin Lin, Y. Liu, Xiujuan Tang, and D. He. The disease severity and clinical outcomes of the sars-cov-2 variants of concern. *Frontiers in Public Health*, 9, 2021.

[8] Angel Yun-Kuan Thye, Jodi WoanFei Law, Priyia Pusparajah, V. Letchumanan, Kok-Gan Chan, and LearnHan Lee. Emerging sars-cov-2 variants of concern (vocs): An impending global crisis. *Biomedicines*, 9, 2021.

[9] P. Lang, Charlotte Tsu-Shin Wu, Anna Florence Le-Nguyen, and Astrid Czock. Influenza vaccination behaviour of healthcare workers in switzerland: A cross-sectional study. *International Journal of Public Health*, 68, 2023.

[10] L. Piccoli, P. Ferrari, G. Piumatti, S. Jovic, Blanca M. Fernandez Rodriguez, F. Mele, Isabella Giacchetto-Sasselli, T. Terrot, C. Silacci-Fregni, E. Cameroni, S. Jaconi, Nicole Sprugasci, I. Bartha, D. Corti, M. Uguccioni, A. Lanzavecchia, C. Garzoni, O. Giannini, E. Bernasconi, L. Elzi, E. Albanese, F. Sallusto, and A. Ceschi. Risk assessment and seroprevalence of sars-cov-2 infection in healthcare workers of covid-19 and non-covid-19 hospitals in southern switzerland. *The Lancet Regional Health - Europe*, 1, 2020.

[11] D. Schwappach and A. Richard. Speak up-related climate and its association with healthcare workers speaking up and withholding voice behaviours: a cross-sectional survey in switzerland. *BMJ Quality & Safety*, 27:827 – 835, 2018.

[12] A. Tsanas, Max A. Little, P. McSharry, and L. Ramig. Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average parkinson's disease symptom severity. *Journal of The Royal Society Interface*, 8:842 – 855, 2011.

[13] R. Chun, Philip T Liu, R. Modlin, J. Adams, and M. Hewison. Impact of vitamin d on immune function: lessons learned from genome-wide analysis. *Frontiers in Physiology*, 5, 2014.

[14] Sanjana R. Sen, Emily C. Sanders, Kristin N. Gabriel, B. Miller, H. M. Isoda, G. S. Salcedo, J. E. Garrido, Rebekah P. Dyer, R. Nakajima, Aarti Jain, Ana-Maria Caldaruse, Alicia M. Santos, K. Bhuvan, D. Tifrea, J. Ricks-Oddie, P. Felgner, R. Edwards, S. Majumdar, and Gregory A. Weiss. Predicting covid-19 severity with a specific nucleocapsid antibody plus disease risk factor score. *mSphere*, 6, 2021.

[15] A. Carabelli, T. Peacock, Lucy G. Thorne, William T. Harvey, J. Hughes, T. D. de Silva, S. Peacock, W. Barclay, G. Towers, and David L. Robertson. Sars-cov-2 variant biology: immune escape, transmission and fitness. *Nature Reviews. Microbiology*, 21:162 – 177, 2023.

[16] Seon-Rye Kim, S. Nam, and Yu-Rin Kim. Risk factors on the progression to clinical outcomes of covid-19 patients in south korea: Using national data. *International Journal of Environmental Research and Public Health*, 17, 2020.

Table 3: Model estimates for interaction effects

	Coefficient	P-value
Intercept	3.86	$4.71\ 10^{-5}$
Group Infected Only	0.509	0.68
Group Not Immune	1.28	0.186
Group Vaccinated Only	0.798	0.37
Omicron Variant	-0.475	0.597
Male Sex	-0.405	0.0514
Group Infected Only:Omicron Variant	0.623	0.636
Group Not Immune:Omicron Variant	0.0455	0.966
Group Vaccinated Only:Omicron Variant	-0.0882	0.924
Age	-0.013	0.0936
Comorbidity	0.569	0.000573

Table reports the pooled OLS regression coefficient estimates which give interaction effects between symptom numbers and immunity group, variant of virus, adjustment for impacted factors.
Coefficient: Estimated effect on the symptom number
P-value: Statistical significance of the estimated effect
Male Sex: If the sex is male, 1: Yes, 0: No
Age: Age in years
Comorbidity: If any pre-existing comorbity existed, 1: Yes, 0: No
Group Infected Only: Infected but not vaccinated
Group Not Immune: Neither infected nor vaccinated
Group Vaccinated Only: Vaccinated but not infected
Omicron Variant: If the variant of SARS-CoV-2 virus is omicron, 1: Yes, 0: No
Group Infected Only:Omicron Variant: Interaction between being in the Infected Only group and the omicron variant
Group Not Immune:Omicron Variant: Interaction between being in the Not Immune group and the omicron variant
Group Vaccinated Only:Omicron Variant: Interaction between being in the Vaccinated Only group and the omicron variant

A Data Description

Here is the data description, as provided by the user:

```
\#\# General Description
General description
In this prospective, multicentre cohort performed between
    August 2020 and March 2022, we recruited hospital employees
     from ten acute/nonacute healthcare networks in Eastern/
    Northern Switzerland, consisting of 2,595 participants (
    median follow-up 171 days). The study comprises infections
    with the delta and the omicron variant. We determined
    immune status in September 2021 based on serology and
    previous SARS-CoV-2 infections/vaccinations: Group N (no
    immunity); Group V (twice vaccinated, uninfected); Group I
    (infected, unvaccinated); Group H (hybrid: infected and $\
    geq$1 vaccination). Participants were asked to get tested
    for SARS-CoV-2 in case of compatible symptoms, according to
     national recommendations. SARS-CoV-2 was detected by
    polymerase chain reaction (PCR) or rapid antigen diagnostic
     (RAD) test, depending on the participating institutions.
    The dataset is consisting of two files, one describing
    vaccination and infection events for all healthworkers, and
     the secone one describing the symptoms for the
    healthworkers who tested positive for SARS-CoV-2.
\#\# Data Files
The dataset consists of 2 data files:

\#\#\# File 1: "TimeToInfection.csv"
Data in the file "TimeToInfection.csv" is organised in time
    intervals, from day\_interval\_start to day\_interval\_stop
    . Missing data is shown as "" for not indicated or not
    relevant (e.g. which vaccine for the non-vaccinated group).
     It is very important to note, that per healthworker (=ID
    number), several rows (time intervals) can exist, and the
    length of the intervals can vary (difference between day\
    _interval\_start and day\_interval\_stop). This can lead to
     biased results if not taken into account, e.g. when
    running a statistical comparison between two columns. It
    can also lead to biases when merging the two files, which
    therefore should be avoided. The file contains 16 columns:

ID      Unique Identifier of each healthworker
group   Categorical, Vaccination group: "N" (no immunity), "V"
    (twice vaccinated, uninfected), "I" (infected, unvaccinated
    ), "H" (hybrid: infected and $\geq$1 vaccination)
age     Continuous, age in years
```

```
sex     Categorical, female", "male" (or "" for not indicated)

BMI     Categorical, "o30" for over 30  or "u30" for below 30

patient\_contact        Having contact with patients during
    work during this interval, 1=yes, 0=no
using\_FFP2\_mask       Always using protective respiratory
    masks during work, 1=yes, 0=no
negative\_swab  documentation of $\geq$1 negative test in the
    previous month, 1=yes, 0=no
booster receipt of booster vaccination, 1=yes, 0=no (or "" for
    not indicated)
positive\_household     categorical, SARS-CoV-2 infection of a
    household contact within the same month, 1=yes, 0=no
months\_since\_immunisation     continuous, time since last
    immunization event (infection or vaccination) in months.
    Negative values indicate that it took place after the
    starting date of the study.
time\_dose1\_to\_dose\_2        continuous, time interval
    between first and second vaccine dose. Empty when not
    vaccinated twice
vaccinetype     Categorical, "Moderna" or "Pfizer\_BioNTech" or
     "" for not vaccinated.
day\_interval\_start    day since start of study when the
    interval starts
day\_interval\_stop     day since start of study when the
    interval stops
infection\_event        If an infection occured during this
    time interval, 1=yes, 0=no

Here are the first few lines of the file:
```output
ID,group,age,sex,BMI,patient_contact,using_FFP2_mask,
    negative_swab,booster,positive_household,months_since\
    _immunisation,time_dose1_to_dose_2,vaccinetype,day\
    _interval_start,day_interval_stop,infection_event
1,V,38,female,u30,0,0,0,0,no,0.8,1.2,Moderna,0,87,0
1,V,38,female,u30,0,0,0,0,no,0.8,1.2,Moderna,87,99,0
1,V,38,female,u30,0,0,0,0,no,0.8,1.2,Moderna,99,113,0
```

\#\#\# File 2: "Symptoms.csv"
Data in the file "Symptoms.csv" is organised per infection
    event, consisting in total of 764 events. Each worker is
    only indicated once. It contains 11 columns:
ID   Unique Identifier, same in both files
```

```
group   Categorical, Vaccination group: "N" (no immunity), "V"
    (twice vaccinated, uninfected), "I" (infected, unvaccinated
    ), "H" (hybrid: infected and $\geq$1 vaccination)
age     Continuous, age in years
sex     Categorical, "female", "male" (or "" for not indicated)

BMI     Categorical, "o30" for $>$30 or "u30" for under 30

comorbidity catgeorical, if any comorbity pre-existed, 1=yes,
    0=no
using\_FFP2\_mask       Always using protective respiratory
    masks during work, 1=yes, 0=no
months\_until\_reinfection      time until next infection in
    months
variant Categorical, "delta" or "omicron" (or "" for not
    indicated)
booster\_over7\_days\_before    If a booster was given in the
    last 7 days before the infection
symptom\_number Continous, Number of symptoms which occured
    after the infection

Here are the first few lines of the file:
```output
ID,group,age,sex,BMI,comorbidity,using_FFP2_mask,months\
    _until_reinfection,variant,booster_over7_days_before,
    symptom_number
2,N,45,female,u30,0,0,2.5,delta,0,11
3,V,58,female,u30,1,0,4.2,omicron,0,6
7,V,32,female,u30,0,1,4.5,omicron,1,5

```
```

B Data Exploration

B.1 Code

The Data Exploration was carried out using the following custom code:

```
import pandas as pd

# Load the data files into pandas DataFrames
time_to_infection = pd.read_csv("TimeToInfection.csv")
symptoms = pd.read_csv("Symptoms.csv")

# Create or open the output file
```

```
output_file = open("data_exploration.txt", 'w')

# Data Size
output_file.write("# Data Size\n")
output_file.write(f"TimeToInfection.csv - Rows: {
    ↪ time_to_infection.shape[0]}, Columns: {time_to_infection.
    ↪ shape[1]}\n")
output_file.write(f"Symptoms.csv - Rows: {symptoms.shape[0]},
    ↪ Columns: {symptoms.shape[1]}\n\n")

# Summary Statistics
output_file.write("# Summary Statistics\n")
output_file.write("## TimeToInfection.csv\n")
output_file.write(time_to_infection.describe().to_string())
output_file.write("\n\n")
output_file.write("## Symptoms.csv\n")
output_file.write(symptoms.describe().to_string())
output_file.write("\n\n")

# Categorical Variables
output_file.write("# Categorical Variables\n")
output_file.write("## TimeToInfection.csv\n")
for col in time_to_infection.select_dtypes(include='object'):
    output_file.write(f"{col}:\n{time_to_infection[col].
        ↪ value_counts().head().to_string()}\n")
output_file.write("\n")

output_file.write("## Symptoms.csv\n")
for col in symptoms.select_dtypes(include='object'):
    output_file.write(f"{col}:\n{symptoms[col].value_counts().
        ↪ head().to_string()}\n")
output_file.write("\n")

# Calculate and add interval lengths to "time_to_infection"
    ↪ DataFrame
time_to_infection['interval_length'] = time_to_infection['
    ↪ day_interval_stop'] - time_to_infection['
    ↪ day_interval_start']

# Write summary of interval lengths
output_file.write("# Interval Lengths for TimeToInfection.csv\n
    ↪ ")
output_file.write(time_to_infection['interval_length'].describe
    ↪ ().to_string())
output_file.write("\n\n")

# Missing Values
output_file.write("# Missing Values\n")
output_file.write("## TimeToInfection.csv\n")
```

```
output_file.write(time_to_infection.replace('', np.nan).isna().
    ↪ sum().to_string())
output_file.write("\n\n")
output_file.write("## Symptoms.csv\n")
output_file.write(symptoms.isna().sum().to_string())
output_file.write("\n\n")

number_of_unvaccinated_people = time_to_infection[(
    ↪ time_to_infection['group'] == 'N') | (time_to_infection['
    ↪ group'] == 'I')].shape[0]
output_file.write("# Number of unvaccinated people\n")
output_file.write(f"{number_of_unvaccinated_people}\n")

output_file.close()
```

B.2 Code Description

The provided Python code conducts data exploration on two datasets, "TimeToInfection.csv" and "Symptoms.csv".

It begins by loading the datasets into pandas DataFrames and determining the size (rows and columns) of each dataset. Summary statistics are then computed for both datasets, which include measures such as mean, standard deviation, minimum, maximum, and quartiles.

The code then examines categorical variables in both datasets, showcasing the top values and their corresponding frequencies. For the "TimeToInfection.csv" dataset, it calculates the length of each time interval, appending this information as a new column. The code proceeds to generate a summary of the interval lengths.

Furthermore, the code identifies missing values in both datasets and indicates the total count of unvaccinated individuals in the "TimeToInfection.csv" dataset.

The outcomes of the data exploration, encompassing data size, summary statistics, analysis of categorical variables, interval lengths, missing values, and the count of unvaccinated individuals, are detailed in the "data_exploration.txt" file for future reference.

B.3 Code Output

data_exploration.txt

```
\# Data Size
TimeToInfection.csv - Rows: 12086, Columns: 16
Symptoms.csv - Rows: 764, Columns: 11
```

```
\# Summary Statistics
\#\# TimeToInfection.csv
         ID   age  patient\_contact  using\_FFP2\_mask
            negative\_swab  booster  months\_since\
            _immunisation  time\_dose1\_to\_dose\_2  day\
            _interval\_start  day\_interval\_stop  infection\
            _event
count 12086 12065            11686            11686
    12086    12086                        11459
    9332               12086              12086
    12086
mean   1300 44.03           0.7941           0.2014
    0.4933   0.5007                        5.015
    1.026               81.21              113.2
    0.06321
std   748.2 11.01           0.4044           0.4011
    0.5      0.5                           2.344
    0.4213               47.03             32.1
    0.2434
min       1    17                 0                0
                 0         0                      -5.3
                         0                  0
    1                 0
25\%     648    35                 1                0
                 0         0                      3.8
                        0.9                   75
    88                0
50\%    1310    44                 1                0
                 0         1                      5.5
                        1                    99
    106               0
75\%    1942    53                 1                0
                 1         1                      6.6
                       1.2                  113
    142                0
max     2595    73                 1                1
                 1         1                     17.8
                       5.1                  171
    178                1

\#\# Symptoms.csv
         ID   age  comorbidity  using\_FFP2\_mask  months\
            _until\_reinfection  booster\_over7\_days\_before
             symptom\_number
count   764   764          719              734
                          764                        764
               764
mean   1315 41.45       0.3825           0.1839
```

```
                        4.1                         0.5209
             3.806
std    742.7  10.69        0.4863              0.3877
                        1.268                       0.4999
             2.177
min        2     17             0                   0
                          0                                 0
                  0
25\%    694.8     33             0                   0
                        3.4                               0
                  2
50\%     1354     41             0                   0
                        4.4                               1
                  4
75\%     1942  49.25             1                   0
                        5.2                               1
                  5
max     2588     68             1                   1
                        5.8                               1
                 11

\# Categorical Variables
\#\# TimeToInfection.csv
group:
group
V     8157
H     2996
I      469
N      464
sex:
sex
female     9617
male       2370
BMI:
BMI
u30     10625
o30      1461
positive\_household:
positive\_household
no      10584
yes      1502
vaccinetype:
vaccinetype
Pfizer\_BioNTech     7816
Moderna             2201
other\_unknown         18

\#\# Symptoms.csv
group:
```

```
group
V    550
H    110
N     67
I     37
sex:
sex
female    620
male      139
BMI:
BMI
u30    679
o30     85
variant:
variant
omicron    591
delta      173

\# Interval Lengths for TimeToInfection.csv
count   12086
mean    32.01
std     26.43
min         1
25\%       14
50\%       24
75\%       34
max        99

\# Missing Values
\#\# TimeToInfection.csv
ID                                0
group                             0
age                              21
sex                              99
BMI                               0
patient\_contact                400
using\_FFP2\_mask                400
negative\_swab                    0
booster                           0
positive\_household               0
months\_since\_immunisation     627
time\_dose1\_to\_dose\_2         2754
vaccinetype                    2051
day\_interval\_start               0
day\_interval\_stop                0
infection\_event                  0
interval\_length                  0

\#\# Symptoms.csv
```

```
ID                                0
group                             0
age                               0
sex                               5
BMI                               0
comorbidity                      45
using\_FFP2\_mask                30
months\_until\_reinfection        0
variant                           0
booster\_over7\_days\_before       0
symptom\_number                   0

\# Number of unvaccinated people
933
```

C Data Analysis

C.1 Code

The Data Analysis was carried out using the following custom code:

```
# IMPORT
import pandas as pd
import numpy as np
from scipy import stats
import statsmodels.formula.api as smf
import pickle

# LOAD DATA
TimeToInfection = pd.read_csv("TimeToInfection.csv")
Symptoms = pd.read_csv("Symptoms.csv")

# DATASET PREPARATIONS
# Drop the rows that contain missing value in the 'sex' column
TimeToInfection.dropna(subset=['sex'], inplace=True)
Symptoms.dropna(subset=['sex'], inplace=True)

# Join Data files
df = pd.merge(TimeToInfection[TimeToInfection.infection_event
    ↪ == 1], Symptoms, on=['ID', 'group', 'age', 'sex', 'BMI'],
    ↪  suffixes=('_x', '_y'))

# DESCRIPTIVE STATISTICS
## Table 0: "Distribution of group, symptom number and variant
    ↪ for infected individuals only"
grouped_df = df.groupby(['group', 'variant']).agg({'
    ↪ symptom_number': ['mean', 'std', 'count']}).reset_index()
```

```
df0 = pd.DataFrame(grouped_df)
df0.columns = ['Group', 'Variant', 'Average Symptom Number', '
    ↪ Standard Deviation', 'Count']
df0.set_index(['Group', 'Variant'], inplace=True)
df0.to_pickle('table_0.pkl')

# PREPROCESSING
# No preprocessing is needed in this case

# ANALYSIS
## Table 1: "Association between symptom numbers and group,
    ↪ variant, adjusting for age, sex, and comorbidity for
    ↪ infected individuals only"
formula1 = 'symptom_number ~ group + variant + age + sex +
    ↪ comorbidity'
model1 = smf.ols(formula1, data=df)
results1 = model1.fit()
table1 = pd.DataFrame({'coef': results1.params, 'p-value':
    ↪ results1.pvalues})
table1.to_pickle('table_1.pkl')

## Table 2: "Association between symptom numbers and variant,
    ↪ adjusting for age, sex, comorbidity and interaction
    ↪ between group and variant for infected individuals only"
formula2 = 'symptom_number ~ group + variant + group:variant +
    ↪ age + sex + comorbidity' # Create interaction term
model2 = smf.ols(formula2, data=df)
results2 = model2.fit()
# Include interaction term in the result table
table2 = pd.DataFrame({'coef': results2.params, 'p-value':
    ↪ results2.pvalues})
table2.to_pickle('table_2.pkl')

# SAVE ADDITIONAL RESULTS
additional_results = {
    'Total number of observations': df.shape[0],
    'Number of comorbidity cases': df['comorbidity'].sum(),
    'Number of people using FFP2 mask': df['using_FFP2_mask_y'
        ↪ ].sum()
}

with open('additional_results.pkl', 'wb') as f:
    pickle.dump(additional_results, f)
```

C.2 Code Description

The provided code performs data analysis on two datasets - "TimeToInfection.csv" and "Symptoms.csv. After loading and preparing the datasets, the code joins the data files based on specific columns and filters out rows with missing values in the 'sex' column.

Descriptive statistics are calculated, presenting the distribution of symptom numbers across different vaccination groups and virus variants for infected individuals. Two statistical models are then created to assess the association between symptom numbers and various factors, adjusting for age, sex, and comorbidity. The first model examines the impact of group and variant on symptom numbers, while the second model introduces an interaction term to evaluate their combined effect.

The analysis results, including coefficients and p-values, are saved in pickle files named 'table_0.pkl', 'table_1.pkl', and 'table_2.pkl' for further inspection or future reference. Additionally, the code saves extra results in 'additional_results.pkl' file, capturing details such as the total number of observations, the count of comorbidity cases, and the number of individuals using FFP2 masks during the study.

C.3 Code Output

table_0.pkl

Group	Variant	Average Symptom Number Count	Standard Deviation
H	delta	3.333	2.251
	6		
	omicron	3.068	2.064
	103		
I	delta	4.429	2.76
	7		
	omicron	4.069	1.771
	29		
N	delta	4.833	2.365
	30		
	omicron	4.306	2.328
	36		
V	delta	4.333	2.32
	129		
	omicron	3.683	2.089
	419		

table_1.pkl

	coef	p-value
Intercept	3.891	3.46e-20
group[T.I]	1.015	0.0174
group[T.N]	1.288	0.000292
group[T.V]	0.7232	0.00204
variant[T.omicron]	-0.512	0.00923
sex[T.male]	-0.4	0.053
age	-0.01297	0.093
comorbidity	0.5724	0.000522

table_2.pkl

	coef	p-value
Intercept	3.86	4.71e-05
group[T.I]	0.5089	0.68
group[T.N]	1.277	0.186
group[T.V]	0.7977	0.37
variant[T.omicron]	-0.4751	0.597
sex[T.male]	-0.4048	0.0514
group[T.I]:variant[T.omicron]	0.6232	0.636
group[T.N]:variant[T.omicron]	0.04547	0.966
group[T.V]:variant[T.omicron]	-0.08818	0.924
age	-0.013	0.0936
comorbidity	0.5695	0.000573

additional_results.pkl

```
{
    'Total number of observations': 759,
    'Number of comorbidity cases': 272.0,
    'Number of people using FFP2 mask': 134.0,
}
```

D LaTeX Table Design

D.1 Code

The LaTeX Table Design was carried out using the following custom code:

```
# IMPORT
import pandas as pd
from my_utils import to_latex_with_note, is_str_in_df,
    ↪ split_mapping, AbbrToNameDef

# PREPARATION FOR ALL TABLES
shared_mapping: AbbrToNameDef = {
```

```
    'coef': ('Coefficient', 'Estimated effect on the symptom
        ↪ number'),
    'p-value': ('P-value', 'Statistical significance of the
        ↪ estimated effect'),
    'sex[T.male]': ('Male Sex', 'If the sex is male, 1: Yes, 0:
        ↪  No'),
    'age': ('Age', 'Age in years'),
    'comorbidity': ('Comorbidity', 'If any pre-existing
        ↪ comorbity existed, 1: Yes, 0: No'),
}

# TABLE 0:
df0 = pd.read_pickle('table_0.pkl')

# Prepare the data mappings for table 0

mapping0: AbbrToNameDef = {
    'Average Symptom Number': ('Mean Symptoms', 'Average number
        ↪  of symptoms from delta or omicron variant infection'
        ↪ ),
    'Standard Deviation': ('Std. Dev.', 'Standard deviation of
        ↪ symptom counts for delta or omicron variant infection
        ↪ '),
    'Count': ('Infections', 'Number of delta or omicron variant
        ↪  infections'),
    'H': ('Hybrid', 'Infected and at least one vaccination'),
    'I': ('Infected Only', 'Infected and not vaccinated'),
    'N': ('Not Immune', 'Neither infected nor vaccinated'),
    'V': ('Vaccinated Only', 'Vaccinated but not infected'),
}
abbrs_to_names0, legend0 = split_mapping({**mapping0})
df0 = df0.rename(columns=abbrs_to_names0, index=abbrs_to_names0
    ↪ )

# Save as LaTeX:
to_latex_with_note(
    df0, 'table_0.tex',
    caption='Summary statistics of health worker infections by
        ↪ different SARS-CoV-2 variants',
    label='table:summary_statistics',
    note='Summary statistics including the count of health
        ↪ worker infections, average symptom count, and \
    standard deviation of symptom counts for different virus
        ↪ variants.',
    legend=legend0)

# TABLE 1:

df1 = pd.read_pickle('table_1.pkl')
```

```
# Prepare the data mappings for table 1
mapping1: AbbrToNameDef = {
    'group[T.I]': ('Group Infected Only', 'Infected but not
        ↪ vaccinated'),
    'group[T.N]': ('Group Not Immune', 'Neither infected nor
        ↪ vaccinated'),
    'group[T.V]': ('Group Vaccinated Only', 'Vaccinated but not
        ↪  infected'),
    'variant[T.omicron]': ('Omicron Variant', 'If the variant
        ↪ of SARS-CoV-2 virus is omicron, 1: Yes, 0: No'),
}
abbrs_to_names1, legend1 = split_mapping({**shared_mapping, **
    ↪ mapping1})
df1 = df1.rename(columns=abbrs_to_names1, index=abbrs_to_names1
    ↪ )

# Save as LaTeX:
to_latex_with_note(
    df1, 'table_1.tex',
    caption='Model estimates of the factors influencing symptom
        ↪  numbers',
    label='table:model_estimates',
    note='Table reports the pooled OLS regression coefficient
        ↪ estimates which give associations between symptom
        ↪ numbers and immunity group, variant of virus, and
        ↪ adjustment for impacted factors.',
    legend=legend1)

# TABLE 2:
df2 = pd.read_pickle('table_2.pkl')

# Prepare the data mappings for table 2
mapping2: AbbrToNameDef = {
    'group[T.I]:variant[T.omicron]': ('Group Infected Only:
        ↪ Omicron Variant', 'Interaction between being in the
        ↪ Infected Only group and the omicron variant'),
    'group[T.N]:variant[T.omicron]': ('Group Not Immune:Omicron
        ↪  Variant', 'Interaction between being in the Not
        ↪ Immune group and the omicron variant'),
    'group[T.V]:variant[T.omicron]': ('Group Vaccinated Only:
        ↪ Omicron Variant', 'Interaction between being in the
        ↪ Vaccinated Only group and the omicron variant'),
}
abbrs_to_names2, legend2 = split_mapping({**shared_mapping, **
    ↪ mapping1, **mapping2})
df2 = df2.rename(columns=abbrs_to_names2, index=abbrs_to_names2
    ↪ )
```

```
# Save as LaTeX:
to_latex_with_note(
    df2, 'table_2.tex',
    caption='Model estimates for interaction effects',
    label='table:interaction_effects',
    note='Table reports the pooled OLS regression coefficient
        ↪ estimates which give interaction effects between
        ↪ symptom numbers and immunity group, variant of virus,
        ↪  adjustment for impacted factors.',
    legend=legend2)
```

D.2 Provided Code

The code above is using the following provided functions:

```
def to_latex_with_note(df, filename: str, caption: str, label:
    ↪ str, note: str = None, legend: Dict[str, str] = None, **
    ↪ kwargs):
    """
    Converts a DataFrame to a LaTeX table with optional note
        ↪ and legend added below the table.

    Parameters:
    - df, filename, caption, label: as in `df.to_latex`.
    - note (optional): Additional note below the table.
    - legend (optional): Dictionary mapping abbreviations to
        ↪ full names.
    - **kwargs: Additional arguments for `df.to_latex`.
    """

def is_str_in_df(df: pd.DataFrame, s: str):
    return any(s in level for level in getattr(df.index, '
        ↪ levels', [df.index]) + getattr(df.columns, 'levels',
        ↪ [df.columns]))

AbbrToNameDef = Dict[Any, Tuple[Optional[str], Optional[str]]]

def split_mapping(abbrs_to_names_and_definitions: AbbrToNameDef
    ↪ ):
    abbrs_to_names = {abbr: name for abbr, (name, definition)
        ↪ in abbrs_to_names_and_definitions.items() if name is
        ↪ not None}
    names_to_definitions = {name or abbr: definition for abbr,
        ↪ (name, definition) in abbrs_to_names_and_definitions.
        ↪ items() if definition is not None}
    return abbrs_to_names, names_to_definitions
```

D.3 Code Output

table_0.tex

```
\% This latex table was generated from: `table\_0.pkl`
\begin{table}[h]
\caption{Summary statistics of health worker infections by
    different SARS-CoV-2 variants}
\label{table:summary\_statistics}
\begin{threeparttable}
\renewcommand{\TPTminimum}{\linewidth}
\makebox[\linewidth]{\%
\begin{tabular}{llrrr}
\toprule
 \&  \& Mean Symptoms \& Std. Dev. \& Infections \\
Group \& Variant \&  \&  \&  \\
\midrule
\textbf{Hybrid} \& \textbf{delta} \& 3.33 \& 2.25 \& 6 \\
\textbf{} \& \textbf{omicron} \& 3.07 \& 2.06 \& 103 \\
\textbf{Infected Only} \& \textbf{delta} \& 4.43 \& 2.76 \& 7
    \\
\textbf{} \& \textbf{omicron} \& 4.07 \& 1.77 \& 29 \\
\textbf{Not Immune} \& \textbf{delta} \& 4.83 \& 2.36 \& 30 \\
\textbf{} \& \textbf{omicron} \& 4.31 \& 2.33 \& 36 \\
\textbf{Vaccinated Only} \& \textbf{delta} \& 4.33 \& 2.32 \&
    129 \\
\textbf{} \& \textbf{omicron} \& 3.68 \& 2.09 \& 419 \\
\bottomrule
\end{tabular}}
\begin{tablenotes}
\footnotesize
\item Summary statistics including the count of health worker
    infections, average symptom count, and     standard
    deviation of symptom counts for different virus variants.
\item \textbf{Mean Symptoms}: Average number of symptoms from
    delta or omicron variant infection
\item \textbf{Std. Dev.}: Standard deviation of symptom counts
    for delta or omicron variant infection
\item \textbf{Infections}: Number of delta or omicron variant
    infections
\item \textbf{Hybrid}: Infected and at least one vaccination
\item \textbf{Infected Only}: Infected and not vaccinated
\item \textbf{Not Immune}: Neither infected nor vaccinated
\item \textbf{Vaccinated Only}: Vaccinated but not infected
\end{tablenotes}
\end{threeparttable}
\end{table}
```

table_1.tex

```
\% This latex table was generated from: `table\_1.pkl`
\begin{table}[h]
\caption{Model estimates of the factors influencing symptom
    numbers}
\label{table:model\_estimates}
\begin{threeparttable}
\renewcommand{\TPTminimum}{\linewidth}
\makebox[\linewidth]{\%
\begin{tabular}{lrl}
\toprule
 \& Coefficient \& P-value \\
\midrule
\textbf{Intercept} \& 3.89 \& \$$<$\$1e-06 \\
\textbf{Group Infected Only} \& 1.02 \& 0.0174 \\
\textbf{Group Not Immune} \& 1.29 \& 0.000292 \\
\textbf{Group Vaccinated Only} \& 0.723 \& 0.00204 \\
\textbf{Omicron Variant} \& -0.512 \& 0.00923 \\
\textbf{Male Sex} \& -0.4 \& 0.053 \\
\textbf{Age} \& -0.013 \& 0.093 \\
\textbf{Comorbidity} \& 0.572 \& 0.000522 \\
\bottomrule
\end{tabular}}
\begin{tablenotes}
\footnotesize
\item Table reports the pooled OLS regression coefficient
    estimates which give associations between symptom numbers
    and immunity group, variant of virus, and adjustment for
    impacted factors.
\item \textbf{Coefficient}: Estimated effect on the symptom
    number
\item \textbf{P-value}: Statistical significance of the
    estimated effect
\item \textbf{Male Sex}: If the sex is male, 1: Yes, 0: No
\item \textbf{Age}: Age in years
\item \textbf{Comorbidity}: If any pre-existing comorbity
    existed, 1: Yes, 0: No
\item \textbf{Group Infected Only}: Infected but not vaccinated
\item \textbf{Group Not Immune}: Neither infected nor
    vaccinated
\item \textbf{Group Vaccinated Only}: Vaccinated but not
    infected
\item \textbf{Omicron Variant}: If the variant of SARS-CoV-2
    virus is omicron, 1: Yes, 0: No
\end{tablenotes}
\end{threeparttable}
\end{table}
```

table_2.tex

```
\% This latex table was generated from: `table\_2.pkl`
\begin{table}[h]
\caption{Model estimates for interaction effects}
\label{table:interaction\_effects}
\begin{threeparttable}
\renewcommand{\TPTminimum}{\linewidth}
\makebox[\linewidth]{\%
\begin{tabular}{lrl}
\toprule
 \& Coefficient \& P-value \\
\midrule
\textbf{Intercept} \& 3.86 \& 4.71e-05 \\
\textbf{Group Infected Only} \& 0.509 \& 0.68 \\
\textbf{Group Not Immune} \& 1.28 \& 0.186 \\
\textbf{Group Vaccinated Only} \& 0.798 \& 0.37 \\
\textbf{Omicron Variant} \& -0.475 \& 0.597 \\
\textbf{Male Sex} \& -0.405 \& 0.0514 \\
\textbf{Group Infected Only:Omicron Variant} \& 0.623 \& 0.636
    \\
\textbf{Group Not Immune:Omicron Variant} \& 0.0455 \& 0.966 \\
\textbf{Group Vaccinated Only:Omicron Variant} \& -0.0882 \&
    0.924 \\
\textbf{Age} \& -0.013 \& 0.0936 \\
\textbf{Comorbidity} \& 0.569 \& 0.000573 \\
\bottomrule
\end{tabular}}
\begin{tablenotes}
\footnotesize
\item Table reports the pooled OLS regression coefficient
    estimates which give interaction effects between symptom
    numbers and immunity group, variant of virus, adjustment
    for impacted factors.
\item \textbf{Coefficient}: Estimated effect on the symptom
    number
\item \textbf{P-value}: Statistical significance of the
    estimated effect
\item \textbf{Male Sex}: If the sex is male, 1: Yes, 0: No
\item \textbf{Age}: Age in years
\item \textbf{Comorbidity}: If any pre-existing comorbity
    existed, 1: Yes, 0: No
\item \textbf{Group Infected Only}: Infected but not vaccinated
\item \textbf{Group Not Immune}: Neither infected nor
    vaccinated
\item \textbf{Group Vaccinated Only}: Vaccinated but not
    infected
\item \textbf{Omicron Variant}: If the variant of SARS-CoV-2
    virus is omicron, 1: Yes, 0: No
\item \textbf{Group Infected Only:Omicron Variant}: Interaction
     between being in the Infected Only group and the omicron
```

```
    variant
\item \textbf{Group Not Immune:Omicron Variant}: Interaction
    between being in the Not Immune group and the omicron
    variant
\item \textbf{Group Vaccinated Only:Omicron Variant}:
    Interaction between being in the Vaccinated Only group and
    the omicron variant
\end{tablenotes}
\end{threeparttable}
\end{table}
```