UHASSELT
KNOWLEDGE IN ACTION

Maastricht University

Faculty of Sciences
School for Information Technology
Master of Statistics and Data Science

Master's thesis

Modeling the impact of symptom severity and vaccination status on the spread of respiratory diseases

Adem Aragaw Seid

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics and Data Science, specialization Quantitative Epidemiology

SUPERVISOR :

Prof. dr. Andrea TORNERI

2024
2025

Transnational University Limburg is a unique collaboration of two universities in two countries: the University of Hasselt and Maastricht University.

UHASSELT
KNOWLEDGE IN ACTION

www.uhasselt.be
Universiteit Hasselt
Campus Hasselt:
Martelarenlaan 42 | 3500 Hasselt
Campus Diepenbeek:
Agoralaan Gebouw D | 3590 Diepenbeek

Maastricht University

Faculty of Sciences
School for Information Technology
Master of Statistics and Data Science

Master's thesis

Modeling the impact of symptom severity and vaccination status on the spread of respiratory diseases

Adem Aragaw Seid

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics and Data Science, specialization Quantitative Epidemiology

SUPERVISOR :

Prof. dr. Andrea TORNERI

Acknowledgment

I would like to express my sincere gratitude to all those who made this thesis possible.
First and foremost, I extend my heartfelt thanks to my supervisor, Prof.Dr.Andrea Torneri, for his continuous guidance, support, and valuable insights throughout this journey.
I am also grateful to the **VLIR-UOS** scholarship program for providing me with the opportunity and support to pursue my studies. My appreciation also goes to all the professors and academic staff at Hasselt University for their excellent teaching and encouragement.
A special thanks goes to my family for their unwavering love, patience, and support during the entire period of my studies. Their belief in me has been a constant source of motivation.

Thank you all for being a part of this journey.

Contents

List of Tables

List of Figures

Abstract

Background: Social contact patterns are central to the spread of respiratory infectious diseases. While age is a well-established determinant, variations in symptom severity and vaccination status may also shape how individuals interact with others, thereby influencing transmission dynamics.

Objective: This study examines the impact of symptom severity and vaccination status on the number of reported contacts and the likelihood of interaction with individuals in fragile health. It further evaluates how such behavioral differences affect epidemic outcomes through stochastic simulation.

Methods: Contact data from 778 participants (1363 observations) were collected during the 2024/25 winter season via the Infectieradar.be platform. Generalized Linear Mixed Models (Poisson and Negative Binomial) were applied to assess determinants of contact counts, while a Generalized Estimating Equation model was used to evaluate predictors of contact with fragile individuals. These statistical estimates informed an individual-based simulation model.

Results: Participants with moderate or severe symptoms reported fewer contacts compared to those with no or mild symptoms. Vaccinated individuals reported slightly higher contact levels, but vaccination status was not significantly associated with contact with fragile individuals. In the simulations, assumptions about behavioural adjustments such as symptom-related contact reduction or vaccination effects influenced epidemic outcomes, with higher extinction probabilities observed when such adjustments were included compared to scenarios without them.

Conclusion: Symptom severity, age, and vaccination status influence contact patterns in measurable ways. Accounting for these behavioral differences in epidemic models is critical, as neglecting them may bias projections of outbreak dynamics. Combining statistical modeling with stochastic simulation provides a robust framework for evaluating intervention strategies and improving epidemic preparedness.

Keywords: *Infectieradar.be, Social contacts, Symptom severity, Vaccination, Fragile health, GLMM, GEE, Epidemic modeling*

1 Introduction

1.1 Background of the Study

Social contact patterns are a key factor in the transmission of respiratory infectious diseases[1]. Like influenza and COVID-19 continue to cause substantial global health burdens, with studies estimating that 10–20% of symptomatic cases may require medical attention or hospitalization [2]. While age is a well-established determinant, variations in symptom severity and vaccination status may also influence how individuals engage with others, impacting disease dynamics.

Previous research has already begun to explore this direction. For instance, [3] found that influenza-like illness can lead to a measurable reduction in non-household contacts, potentially decreasing transmission. More recently, [4] used CoMix data from 16 European countries to show that individuals perceiving a disease as more severe reduced their number of social contacts, whereas vaccinated individuals tended to report more contacts.

The type, duration, and number of social interactions are very important for the spread of respiratory infectious illnesses. To precisely simulate how diseases spread and put in place effective control measures, it is necessary to understand these interaction patterns.Studies have shown that not everyone interacts with others in the same way. Instead, it varies on their age, work, health, and cultural standards.[1, 5]. For example, school, work, and social activities tend to bring youth and working-age individuals into frequent contact, which can accelerate transmission, whereas older adults typically have fewer social interactions but are more susceptible to severe outcomes when infected.

Fixed demographic characteristics, such as age and sex, are not the only factors influencing how people interact; dynamic conditions, such as current health status and vaccination, can also substantially alter contact behavior. Recent study have revealed that dynamic factors like the severity of symptoms and vaccination status can also have a large effect. People who develop symptoms often reduce their social contacts, either because they feel unwell or as a result of public health measures such as isolation or quarantine, which aim to lower the risk of transmission [6, 7]. People who do not develop symptoms, on the other hand, may continue their usual activities, which can facilitate transmission without being detected. Such variation in behavior complicates both the modeling of disease spread and the monitoring of outbreaks.

Vaccination generally provides substantial protection against infection and disease. However, in the case of leaky vaccines where protection is partial rather than complete the number of contacts remains an important factor in transmission. If vaccinated individuals have different contact patterns compared to unvaccinated ones, this behavioral difference can introduce bias when estimating the effectiveness of vaccine-based interventions [8, 9].

During the COVID-19 epidemic, it became evident that reducing social contacts through measures such as lockdowns, school closures, and remote working policies could substantially decrease transmission events [10, 11]. These things worked for a while, but we don't know how

they will affect behavior in the long run. Infectieradar.be is a participatory monitoring platform in Belgium that collects repeated self reported information on symptoms, social contacts, and vaccination status. In this study, we use these data to investigate two key questions: first, whether individuals adjust their contact behavior based on symptom onset or vaccination status; and second, whether these factors influence the probability of visiting frail individuals those more vulnerable to severe symptoms.

Addressing these questions helps fill an important gap in understanding how behavioral changes interact with epidemiological risk, thereby improving the accuracy of disease spread models and informing targeted public health strategies. In addition,our study examines how assumptions about contact behavior particularly those related to symptom onset and vaccination status can influence the outputs of epidemiological models. While it is not possible to fully capture or predict actual human behavior, our analysis allows us to explore how differences in contact patterns could shape epidemic dynamics. We provide insight people understand how diseases spread in a more complicated way by informing epidemiological model real-world data to determine age-specific contact rates and seeing how these rates fluctuate with vaccination status and symptoms. The results of this study are of primary interest to public health officers and policy makers, who can use these insights to improve preparedness, and to develop policies aimed at containing and mitigating epidemic threats.

The World Health Organization (WHO) usually puts symptoms into several groups based on how bad they are, such mild, moderate, and severe. However, this study used a simpler binary classification. We put the symptoms into two groups based on how likely they were to affect how people interacted. The first group, called "None or Mild," was made up of those who said they had no symptoms or just mild ones, such sneezing, a runny or plugged nose, or watery eyes. These symptoms usually don't get in the way of normal everyday activities and are less likely to impair social interactions.

The second group, "Moderate or Severe," included symptoms that showed a more serious disease, such as cough, sore throat, headache, chills, nausea, muscle or joint pain (moderate), and fever, shortness of breath, chest discomfort, disorientation, and vomiting (severe). This two-part method shows that we need clearer differences in how people behave and how easy it is to understand the model, while still following the WHO's standards for how severe a disease is based on clinical patterns [12].

Despite increasing evidence that symptom severity and vaccination status can influence social contact behavior, there is limited quantitative understanding of how these factors jointly shape age-specific contact patterns and the probability of interacting with frail individuals. Few studies have combined participatory surveillance data with statistical modeling to estimate such behavior-driven differences, and even fewer have incorporated these empirical estimates into epidemic simulations. Addressing this gap is crucial for refining epidemic models and for informing more realistic public health preparedness plans.

1.2 Study Design

The data for this study were collected through Infectieradar.be, the Belgian arm of the international Influenzanet network of participatory surveillance platforms. Infectieradar.be invites volunteers from the general population to register online and provide repeated self-reported information on their health status, symptoms, social contacts, and vaccination history. Participants receive short surveys at regular intervals, allowing near real-time monitoring of respiratory illness trends and contact patterns in the community. In this study, a contact was defined as any face-to-face interaction within 3 meters, reported by participants for the 24-hour period from 5:00 AM on the survey day to 5:00 AM the previous day.

For this investigation, we specifically analyzed data gathered during the Belgian winter respiratory season of 2024/25. A total of 778 participants provided detailed information about their daily social contacts, symptom experiences, and vaccination records. This rich, longitudinal dataset offers a unique opportunity to explore how individual characteristics and health status influence contact behavior in a real-world setting. To collect contact information, a rotating 13-group strategy was adopted, ensuring that each participant reported contact data approximately every 12 weeks while distributing the data collection evenly over time.

Infectieradar.be collects symptom information on a weekly basis, while social contact data are collected less frequently approximately every 13 weeks through the rotating 13 group strategy. For this study, we merged the contact data,symptom data, the intake dataset containing demographic characteristics. Records were matched using participant ID and survey date. For each contact survey, we paired the corresponding symptom status reported for the same survey period, ensuring that health status reflected the day the contact data were collected.

1.3 Objective of the Study

The main goal of this study is to find out how the presence of symptoms and the vaccination status affect how people interact with others during the winter respiratory season. The study's main goal is to use important indicators such symptom level, age group, sex, and vaccination status to model the number of reported daily interactions.

1.4 Variable Description

The final analysis dataset was created by linking weekly symptom survey responses with the nearest social contact diary entries, supplemented with demographic and vaccination information from intake and vaccination records. The questionnaire used for collecting social contact information is provided in the Appendix for reference. Table 1 provides an overview of all variables used in the analysis.

Table 1: Description of Variables Used in the Analysis

Variable	Type	Description
`contact_number`	Count	Number of reported social contacts in the 24-hour reference period.
`symptom_severity`	Binary	No/Mild (1) vs. Moderate/Severe (0) symptoms.
`AgeG`	Categorical	Age group (18–59, 60+).
`SEX`	Categorical	Sex of participant (Male/Female).
`vac_status_binary`	Binary	Vaccinated (1) or not vaccinated (0).
`contact_with_fragile`	Binary	Reported contact with a fragile individual (1) or not (0).
`participantId`	Identifier	Unique participant identifier (random effect in models).
`week`	Time	Survey week of participation.

2 Statistical Modeling

We used a full statistical framework based on Generalized Linear Mixed Models(GLMM) and Generalized Estimating Equations(GEE) to properly analyze the longitudinal and partially repeated measures data collected in this study [13, 14].

2.1 Poisson GLMM

For outcomes measured as non-negative integer counts, the Poisson GLMM is a commonly used modeling approach [15]. Since our outcome variable, the number of social contacts, is a count, this framework was a natural starting point for the analysis. The Poisson distribution says that the mean and variance of the count variable are the same. In this framework, we describe the number of social contact that person i reported at time j, which we call Y_{ij}, follows a Poisson distribution with a conditional mean of μ_{ij}:

$$Y_{ij} \sim Poisson(\mu_{ij}), \quad \log(\mu_{ij}) = \beta_0 + \beta_1 Severity_{ij} + \beta_2 AgeGroup_i + \beta_3 Sex_i + \beta_4 VaccStatus_i + u_i$$

In this case, β_0 is the intercept, while β_1 through β_4 are fixed-effect coefficients that show how symptom severity, age group, sex, and vaccination status affect the outcome. We added the participant-specific random intercept $u_i \sim \mathcal{N}(0, \sigma^2)$ to account for the fact that repeated measures can cause correlation between individuals. When using a Generalized Linear Mixed Model (GLMM) with a log link and a single random intercept $b \sim \mathcal{N}(0, \sigma^2)$, the marginal mean of the response variable is not just $\exp(\beta_0 + X\beta)$, but also $\exp(\beta_0 + \frac{1}{2}\sigma^2 + X\beta)$. This happens

when you take the average of the random effect. We can find $E[e^b] = \exp\left(\frac{1}{2}\sigma^2\right)$ by using the moment-generating function of the normal distribution at $t = 1$ [13]. So, the random effect raises the intercept by a consistent amount, which is $\frac{1}{2}\sigma^2$. On the log scale, fixed effects like symptom severity and vaccination status are usually understood. However, this constant offset changes how they affect the marginal mean. This shows how useful GLMMs are for looking at these impacts because they take into account both the differences in contact counts between individuals and within individuals.

2.2 Negative Binomial GLMM

Initial model diagnostics showed that the variance in contact counts was much greater than the mean, which went against the Poisson distribution's assumption of equal dispersion [16]. We used the Negative Binomial GLMM to fix this. It adds another dispersion parameter, θ, so that the variance might be greater than the mean. This extra flexibility is especially helpful for modeling real-world epidemiological count data, which often shows a lot of variation across people.

The Negative Binomial model used the same linear predictor structure as the Poisson GLMM, but it changed the conditional distribution:

$$Y_{ij} \sim NegativeBinomial(\mu_{ij}, \theta),$$
$$\log(\mu_{ij}) = \beta_0 + \beta_1 Severityij + \beta_2 AgeGroupi + \beta_3 Sexi + \beta_4 VaccStatusi + ui$$

We used the data to directly estimate the dispersion parameter θ to adjust for overdispersion in the count outcome. The Negative Binomial model allows for more variation than the Poisson model because it adds an extra parameter to account for extra volatility beyond the mean. The Negative Binomial model fit the observed contact data better than the other models, according to model fit indices and residual diagnostics. This supports its choice for the final analysis [17].

2.3 GEE for Contact with Fragile Individuals

We also looked at how likely it was for people to come into contact with someone who are Vulnerable individuals, in addition to modeling contact rates. This outcome is binary and shows if a participant said they had at least one encounter with a fragile person on the day of the survey. We used Generalized Estimating Equations (GEE) with a logistic link function to deal with the fact that the same people made repeated observations that were related to each other [18, 19].

Let C_{ij} be a binary variable that equals 1 if person i says they had contact with a vulnerable person at time j, and 0 otherwise. The GEE model makes the following assumptions:

$C_{ij} \sim Bernoulli(\pi_{ij}), \quad logit(\pi_{ij}) = \beta_0 + \beta_1 Severity_{ij} + beta_2 AgeGroup_i + \beta_3 Sex_i + \beta_4 VaccStatus_i$

The expected chance π_{ij} of coming into contact with a fragile person is written as:

$$\pi_{ij} = \frac{1}{1 + \exp\left(-(\beta_0 + \beta_1 Severity_{ij} + \beta_2 AgeGroup_i + \beta_3 Sex_i + \beta_4 VaccStatus_i)\right)}$$

We assume that the working correlation structure was exchangeable, which means that all pairs of observations within a subject are equally correlated. We employed strong sandwich standard errors to make sure that our conclusions were still valid even if the correlation structure wasn't set up correctly. This method gives population-averaged estimates of how factors affect the chances of meeting vulnerable people in social situations.

2.4 Handling Missing Data

Missing values, including in key covariates, were present in the dataset. To avoid the loss of statistical power and potential bias introduced by listwise deletion, we applied multiple imputation, which allowed us to retain all available information while appropriately accounting for uncertainty due to missingness. This method works on the idea that data is missing at random (MAR) and makes several possible datasets by filling in missing values with information that is already known. After that, we looked at each of the imputed datasets one at a time and then put the results together to deal with the uncertainty induced by the missing data [20]. This method let us keep more observations, which made our regression analysis more accurate and dependable.

2.5 Data description

Table 2 presents a summary of the characteristics of the study population included in the analysis. A total of 778 unique participants were observed through the Infectieradar.be platform. Among them, 561 participants (72.1%) had repeated measurements over time, indicating multiple weekly responses. This longitudinal structure is important for modeling individual variation and temporal trends in reported symptoms and contact behavior.

The majority of participants were aged 60 years and above (63.1%), while 36.1% were in the 18–59 age group. Female participants made up 57.8% of the sample, with male participants accounting for 41.9%. A small proportion of respondents had undefined sex due to missing or ambiguous data.Vaccination coverage was relatively high, with 61.4% of participants reporting either COVID-19 or influenza vaccination. However, vaccination status data were missing for 34.6% of participants. On average, participants reported 12.5 social contacts per observation, with a wide range from 0 to 155 contacts.

Table 2: Descriptive statistics of study participants (collapsed at individual level)

Variable	**Summary**
Total participants	778
Participants with repeated measures	561 (72.1%)
Participants with only one measurement	217 (27.9%)
Mean number of contacts	12.5 (Range: 0 – 155)
Symptom severity (unique-level)	Moderate/Severe: 170 (21.9%), No/Mild: 608 (78.1%)
Age group	18–59: 281 (36.1%), 60+: 491 (63.1%), Missing: 6 (0.8%)
Sex	Female: 450 (57.8%), Male: 326 (41.9%), Undefined: 2 (0.3%)
Vaccination status	Vaccinated: 478 (61.4%), Not vaccinated: 31 (4.0%), Missing: 269 (34.6%)
Contact with Fragile person	Yes: 703 (90.4%), No: 75 (9.6%)

2.6 Simulation Study

An individual-based stochastic SIR-type model was developed to explore the impact of symptom-driven behavioural changes and vaccination status on epidemic dynamics. In this framework, each individual could be in one of three states: susceptible (S), infected (I), or recovered (R). Age-specific contact rates determined daily interactions, and transmission occurred when a susceptible individual contacted an infected individual, with probability q. The transmission probability was calibrated such that the basic reproduction number was $R_0 = 1.5$ in a fully susceptible population.

The simulated population consisted of 1000 individuals, equally distributed across four age groups (0–12, 13–17, 18–59, and 60+ years). The epidemic was seeded with one initial infection and simulated for a maximum of 365 days, or until no infectious individuals remained. The stochastic design meant that outcomes varied across runs due to the random nature of contacts and transmissions.

Behavioural parameters were derived from the statistical models, including contact reduction during symptomatic periods and vaccination-related effects. For each scenario, 100 independent simulations were performed. Key outcomes included the final epidemic size, extinction probability, peak incidence, and timing of the epidemic peak.

Several scenarios were defined to evaluate how behavioural and vaccination-related factors influence epidemic dynamics. In the baseline scenario, no behavioural adjustments were introduced and contact rates remained unchanged regardless of symptoms or vaccination status. In the statistical model scenario, contact reductions during symptomatic periods and vaccination-related effects were incorporated using estimates from the GLMM analyses. To capture possible risk-compensating behaviour, a “double vaccination” scenario assumed that vaccinated individ-

uals increased their contacts twice. In another set of scenarios, vaccination effects were fixed adapted from the statistical model, while symptom-related contact reductions were varied: a low-reduction case assumed contact rates scaled up to 1.5(no reduction/increase), whereas a high-reduction case assumed contact rates is 0.1 (near isolation). Each of these scenarios was simulated under three vaccination coverage levels (30%, 60%, and 90%), with vaccine effectiveness fixed at 50%. This design allowed us to disentangle the relative impact of vaccination coverage, behavioural adaptation, and symptom-related contact reductions on epidemic outcomes.

3 Results

3.1 Exploratory Data analysis

Figure 1 indicate the overall incidence of flu-like symptoms during the 2024/25 Belgian winter respiratory season started high at around 250 per 1,000 in week 2024W38, fluctuated with peaks near 225 and 200 per 1,000 in weeks 2024W49 and 2025W04, and then declined sharply to below 150 per 1,000 by week 2025W10, with a slight rise by week 2025W12.

$$Incidence_{week} = \left(\frac{Symptomatic_{week}}{ActiveReports_{week}} \right) \times 1000$$

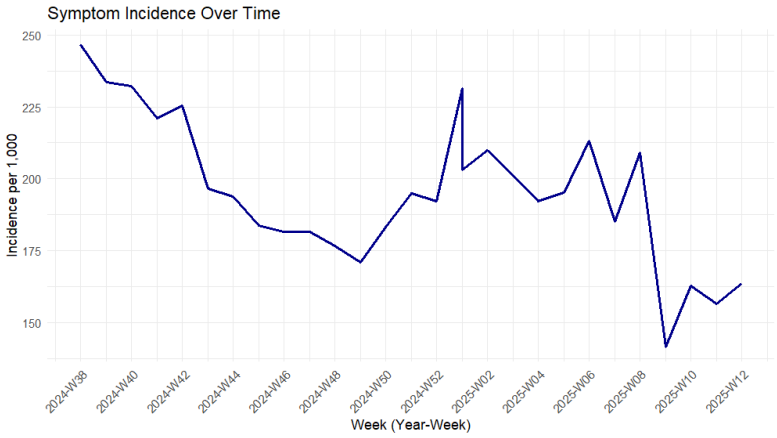

Figure 1: Overall weekly symptom trend (2024–2025)

Figure 2 indicate Women have a higher incidence of symptom due to differences in reporting, exposure levels, or health-seeking behaviors.Individuals in education, full-time work, and self employment show higher and more volatile symptom trends, potentially due to increased exposure and groups like the retired or part-time employed display lower and more stable incidence, possibly reflecting reduced contact rates or consistent routines (Figure 9).

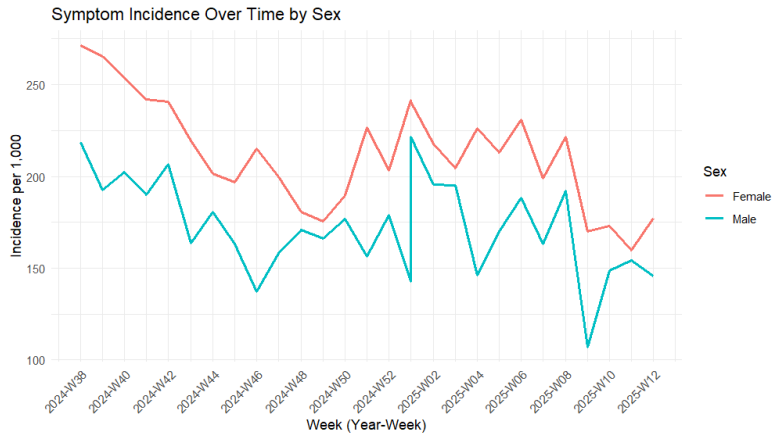

Figure 2: Symptom trend by gender

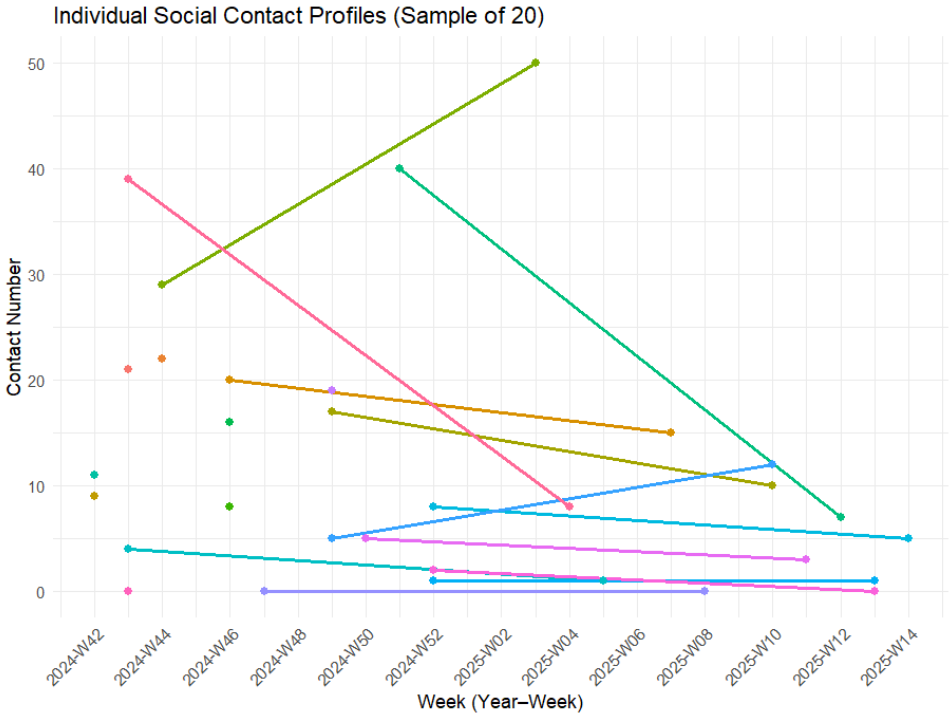

Figure 3: Individual Profile of 20 random samples

In Figure 3, individual contact trajectories for a random sample of 20 participants (one diary wave roughly every 12 weeks) show substantial heterogeneity both at baseline and over time. Several participants start at markedly different contact levels, and their paths subsequently diverge some increasing, others decreasing or fluctuating. This pattern naturally motivates a mixed-effects specification with subject-specific random intercepts to capture stable differences in average contact propensity, and random slopes in time to allow participant specific trends across week.

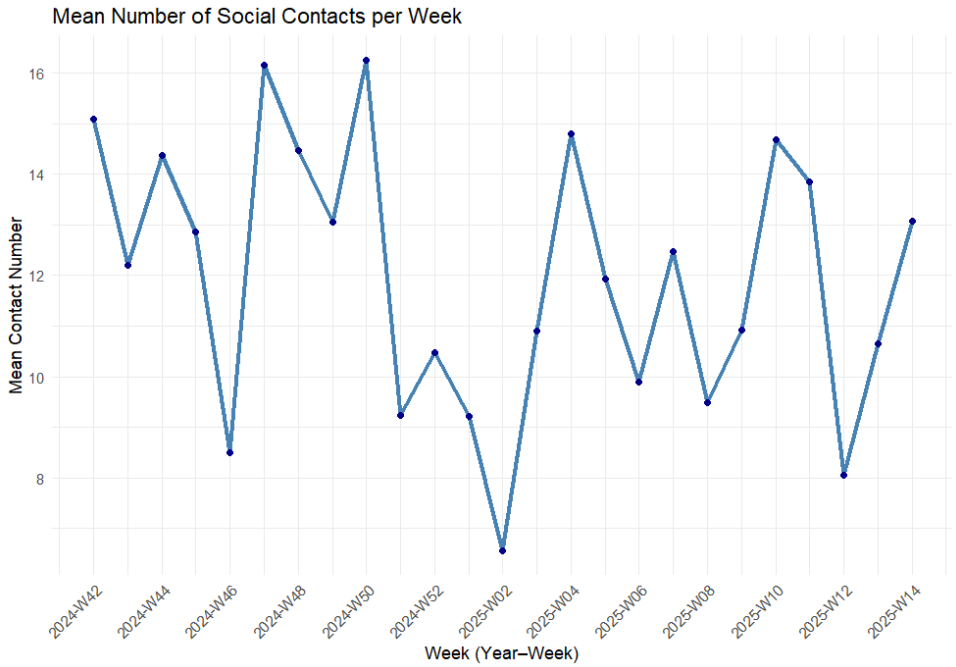

Figure 4: Mean structure(average number of Contact) over time

The weekly average number of social contacts that participants reported during the study period is shown in Figure 4. With values typically falling between 8 and 16 contacts per week, the plot shows clear variations in the average contact number over time and the number of individual filled in specific week presented in Table 6. It is possible to spot periodic peaks that signify times when social interaction is higher. Notwithstanding this fluctuation, the mean contact trend does not exhibit a consistent upward or downward trajectory, indicating that the average social mixing behavior was fairly constant throughout the study weeks.

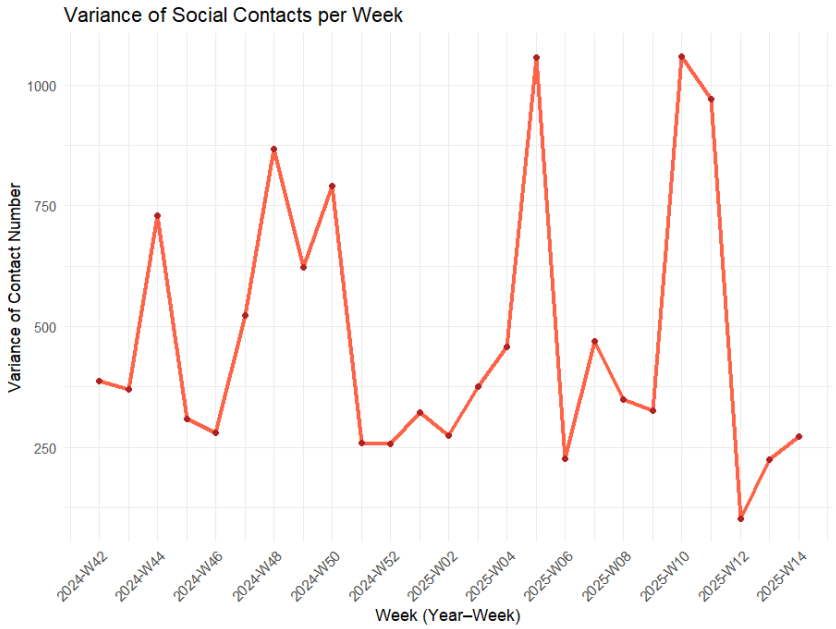

Figure 5: Variance structure of Contact

Figure 5 displays the variance of social contact numbers across weeks. The variance fluctuates considerably over time, indicating notable heterogeneity in individual contact behaviors also the number of individual filled in each week presented in Table 6. Some weeks, such as week 2025-W06 and 2025-W10, show extremely high variability . Other weeks, like 2025-W13, show relatively low variance, indicating more homogeneous contact behavior among participants. These patterns reinforce the presence of overdispersion in contact data and justify the use of models like the negative binomial distribution, which can accommodate such variability better than the Poisson model. Additionally, spikes in variance may coincide with holidays, policy changes, or public health messaging, warranting further contextual investigation.

3.2 Modeling Results

Both random intercept only and random intercept and slope GLMM specifications were fitted, but the model with only a random intercept was preferred and used as the initial parameterisation in the simulation, as the variance of the random slope for week was almost zero (Table 7) and the data exhibited overdispersion (Table 3), for which a negative binomial distribution was employed.

Table 3: Negative Binomial GLMM results comparing Complete Case and Multiple Imputation (MI) analyses.

Covariate	**Complete Case**			**Multiple Imputation**		
	Estimate	Std. Error	*p*-value	Estimate	Std. Error	*p*-value
Intercept	2.7881	0.1827	<0.001	2.6660	0.1833	<0.001
No/Mild symptoms	0.0536	0.1005	0.590	0.0817	0.0863	0.344
Age 60+	-0.4573	0.0836	<0.001	-0.4602	0.0718	<0.001
Male	-0.1170	0.0772	0.130	-0.0868	0.0696	0.212
Vaccinated	-0.0652	0.1612	0.690	-0.1215	0.1715	0.479
Random effect (ID)	Variance = 0.160, SD = 0.400			Variance = 0.368, SD = 0.607		
Dispersion (nbinom1)	17.5			14.1		

The results in Table 3 show that the patterns are the same in both the complete case and the multiple imputation (MI) datasets. People aged 60 and older reported significantly fewer contacts than people aged 18 to 59 ($p < 0.001$) in both approaches. There were no statistically significant differences in the number of contacts between participants with no or mild symptoms ($n = 608$) and those with moderate or severe symptoms ($n = 170$).

The much smaller sample size in the moderate/severe group means the analysis may lack sufficient statistical power to detect small differences, so the absence of significance should be interpreted with care. This could mean that mild symptoms do not lead to substantial changes in behaviour, or that individuals remain socially active even when experiencing sever symptoms. The effects of sex and vaccination status were also not statistically significant, but males had less contact in both models. The fact that the complete case and MI analyses agree with each other makes these results even stronger. This means that the interpretation wasn't changed much by filling in missing data.

Directly to express as mean contact, the predicted mean contact count for the reference group (female, unvaccinated, 18–59 years, moderate/severe symptoms) was approximately 14.4 (exp(2.666)). Relative to this baseline, individuals aged 60+ reported about 37% fewer contacts (mean ≈ 9), while males and vaccinated individuals reported slightly fewer contacts (means ≈ 14 and 13, respectively). Those with no or mild symptoms reported a slightly higher mean (16), though the difference was not statistically significant.

Table 4: GEE logistic regression results for Fragile status (Complete Case vs Multiple Imputation).

Covariate	Complete Case			Multiple Imputation		
	Estimate	Std. Error	p-value	Estimate	Std. Error	p-value
Intercept	2.799	0.637	<0.001	2.318	0.547	<0.001
No/Mild symptoms	-0.185	0.319	0.560	0.138	0.248	0.580
Age group: 60+	-0.030	0.229	0.900	-0.034	0.188	0.850
Sex: Male	0.157	0.222	0.480	0.232	0.186	0.210
Vaccinated	-0.640	0.607	0.290	-0.346	0.536	0.520

In the multiple imputation (MI) GEE model, none of the covariates were statistically significant predictors of reporting contact with fragile persons. The coefficient for symptom severity ($\beta = 0.138$, $p = 0.58$) suggests that individuals with no or mild symptoms had slightly higher log-odds of contacting a fragile person compared to those with moderate or severe symptoms, but the effect was negligible. Similarly, the effect of age ($\beta = -0.034$, $p = 0.85$) and sex ($\beta = 0.232$, $p = 0.21$) were close to zero, indicating little to no difference in log-odds across age groups or between males and females. Vaccination status also did not meaningfully influence contact with fragile individuals ($\beta = -0.346$, $p = 0.52$).

When expressed in terms of probabilities, the estimated likelihood of reporting a fragile contact was consistently high across all groups: about 92% for individuals with no or mild symptoms versus 91% for those with moderate or severe symptoms; approximately 91–92% for males and females; and around 89% for vaccinated individuals compared with 91% for unvaccinated ones. These findings indicate that the probability of fragile contacts remained stable and did not differ substantially across demographic or clinical characteristics.

3.3 Simulation Result

We informed the simulation-based model using the estimates obtained from the statistical analysis, and supplemented it with additional publicly available estimates for the categories not observed in our dataset. For children aged 0–12 years, the mean number of contacts was set to 10, while adolescents aged 13–17 years were assumed to have an average of 14 contacts, both based on the SoCRATES social contact online tool(https://lwillem.shinyapps.io/socrates_rshiny/).Estimates for adults were derived from our generalized linear mixed model (GLMM): adults aged 18–59 reported on average 16 contacts, whereas older adults aged 60 and above reported about 10. These contact rates demonstrate how people of different ages generally interact with each other, and they were used as input parameters in the individual-based simulation model to depict how the virus actually spreads.

Figures 6 and 7 present the simulated final epidemic sizes across multiple intervention scenarios, comparing a *Baseline* (blue), a *Statistical Model-based adaptive contact reduction* (green), and a

Vaccination intensification scenario (*Vac_2x*, red). The scenarios vary by vaccination coverage (30%, 60%, 90%) and combinations of symptom-driven contact reduction and vaccination-based contact reduction.

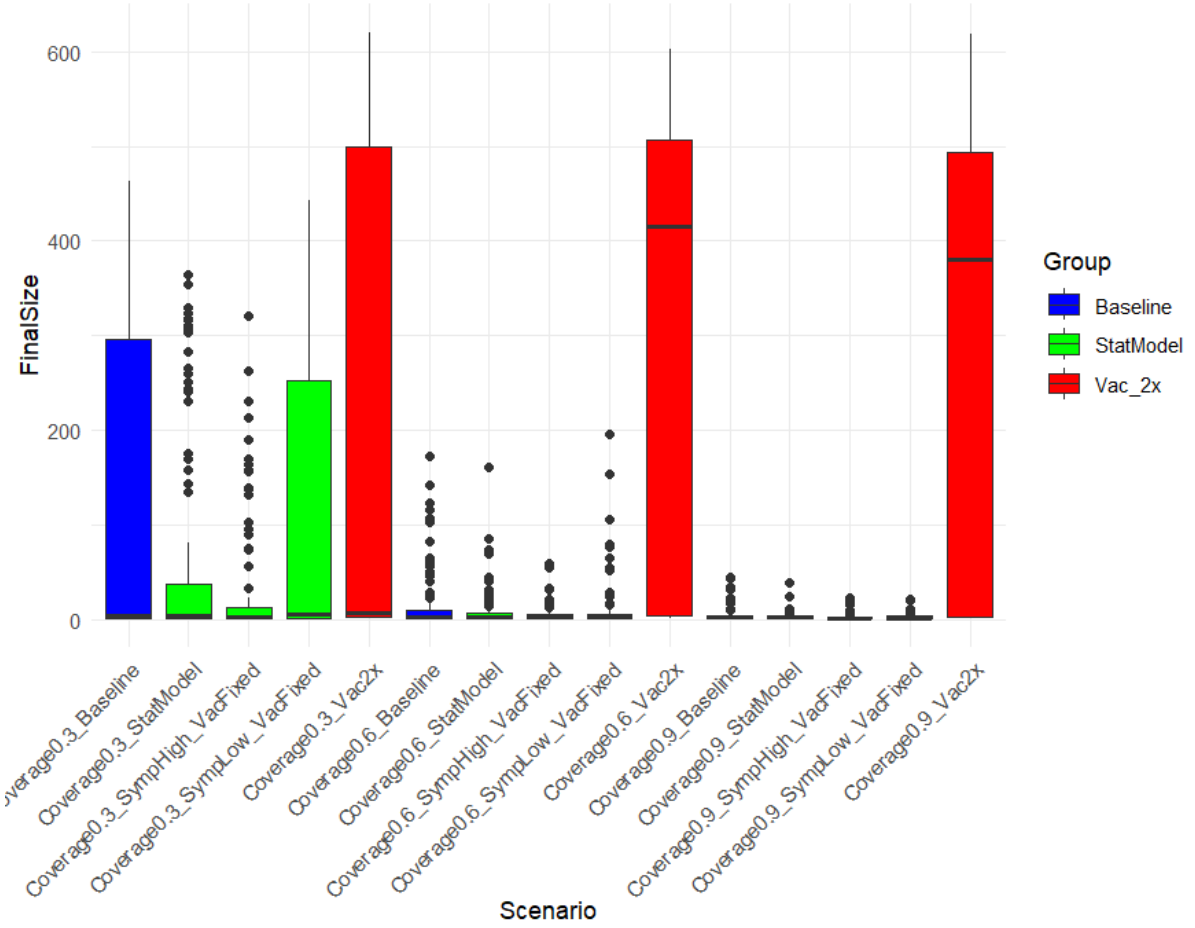

Figure 6: Over all Epidemic Final Size

Figure 6, which includes **all outbreaks** (including those that quickly die out), shows more variability and a larger number of small epidemic sizes across scenarios with higher contact reductions. Both the *Baseline* and *StatModel* approaches at higher coverage levels (60% and 90%) produce a significant number of early extinctions, as indicated by the large proportion of small final sizes near zero. By contrast, the *Vac_2x* scenario shows consistently high epidemic sizes with minimal early extinctions, again pointing to the importance of integrating behavioral interventions with vaccination programs.

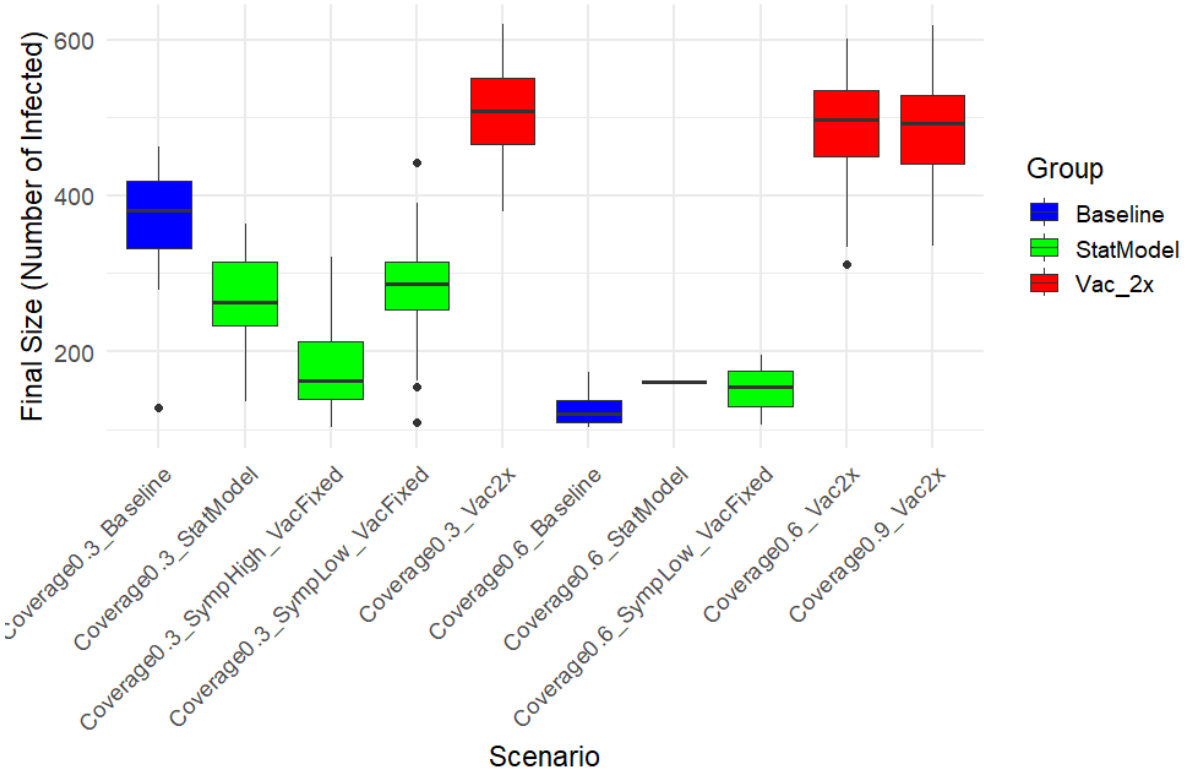

Figure 7: Final Epidemic Size (Non-Extinct Outbreaks Only)

In Figure 7, which includes only non-extinct outbreaks (final epidemic size exceeding 10% of the population), the *StatModel* consistently yields smaller epidemic sizes than the *Baseline* for all coverage levels, indicating that targeted reductions in contact rates informed by statistical modelling can slow transmission. For the scenario contact rate is double for vaccinated group results in substantially larger epidemic sizes, likely due to the absence of strong contact reductions despite higher vaccination efforts, highlighting that vaccination alone without significant behavioral modifications may be insufficient to suppress large outbreaks, especially in scenarios with partial vaccine effectiveness.

Overall, these results emphasize that while increasing vaccination coverage is beneficial, combining vaccination with targeted contact reductions especially for symptomatic individuals yields the most pronounced reductions in epidemic size and increases the probability of outbreak extinction. This outcome reflects the dynamics of a leaky vaccine: although susceptibility is reduced by half, increased contact frequency among vaccinated individuals offsets this benefit, sustaining or even amplifying transmission.

Table 5 presents extinction probabilities across vaccination coverage levels (30%, 60%, 90%) and behavioral scenarios, where "extinct" is defined as the epidemic prevalence dropping to $\leq$10% of the population ($\leq$100 individuals) within 100 days, based on 100 simulations per scenario.

Table 5: Extinction rates (The value in the table under the Condition of Each Column)

Coverage	Baseline	Statistical Model	Double (Vac)	Symp Contact rate(VacFixed)
30%	0.63	0.78	0.46	high= 0.41, low = 1.00
60%	0.95	0.96	0.46	high= 0.61, low= 1.00
90%	1.00	1.00	0.55	high = 1.00, low = 1.00

Extinction rates for different vaccination coverages and contact rate adjustment scenarios are presented in Table 5. As expected, higher vaccination coverage consistently increases the probability of outbreak extinction across all scenarios. The baseline scenario, where no contact rate adjustments are applied, shows lower extinction rates compared to the statistical model scenario, which incorporates contact reductions from the fitted model. In the "Double" scenario, where vaccinated individuals have twice as many contacts as unvaccinated individuals, extinction rates are substantially lower, especially at intermediate coverage (e.g., 0.46 at 60% coverage), indicating the potential negative effect of increased contact behavior among vaccinated individuals.

When vaccination effects on contact rates are fixed to those from the statistical model, symptom-related contact reductions strongly influence extinction probabilities. Strong reductions (contact rate = 0.1(low), near isolation) lead to extinction being almost certain across all coverage levels. In contrast,if we allow higher contact for symptomatic (contact rate = 1.5(high)) result in much lower extinction probabilities, showing that limited behavioural change when symptomatic can sustain transmission even at higher vaccination coverage.

These results highlight that both vaccination coverage and behavioral adaptations play a critical role in determining epidemic extinction.

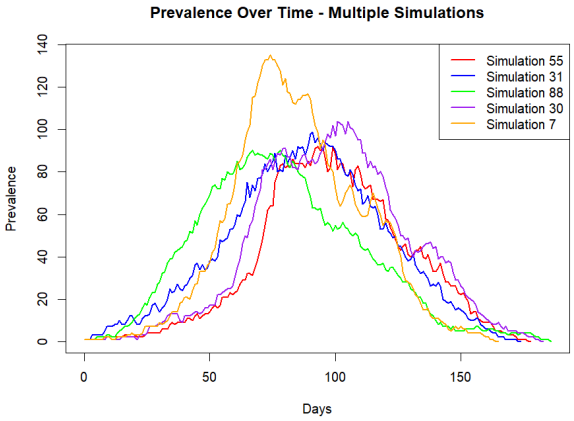

Figure 8: prevalence across 5 simulations

The evolution of infections over time for five stochastic simulations under the Statistical Model scenario is shown in Figure 8. Despite identical parameters, the epidemic curves differ in peak size and timing, illustrating the role of chance in early transmission. All runs display a single

wave with peaks between days 60 and 90, but maximum prevalence varies (about 90–135 cases). This highlights the importance of stochastic simulation for capturing variability, while overall outcomes are summarized in the boxplots.

4 Discussion

This study combined statistical modeling and individual-based simulations to investigate how social contact behaviour, symptom onset, and vaccination status shape the spread of respiratory infections. At the statistical modeling stage, generalized linear mixed models (GLMM) and generalized estimating equations (GEE) were applied to evaluate factors of number of social contact and the probability of interacting with fragile individuals. The simulation stage then used these data-driven parameters to explore epidemic outcomes under alternative behavioural and vaccination scenarios.

At the descriptive level, the mean structure of social contacts revealed moderate week-to-week fluctuations with identifiable peaks and troughs. These variations may reflect seasonal or behavioural changes, such as holidays, policy shifts, or school and work schedules, consistent with previous evidence showing that contact activity is not constant across time [21]. The individual profile plots further demonstrated heterogeneity in contact behaviour, with some individuals consistently reporting higher numbers of contacts than others. Such variability motivates the use of random intercept and slope models, since both baseline levels and changes over time differ across individuals.

Symptom reporting showed clear gender differences: females reported higher symptom incidence than males. A mix of biological and behavioural factors may explain this finding. For example, fluctuations in sex hormones during the menstrual cycle are known to influence immune response [22], and prior studies have suggested that women are generally more likely to perceive and report symptoms compared to men [23]. These observations underscore the importance of considering gender in both behavioural surveys and epidemic models. In contrast, the GLMM analysis found no strong statistical evidence that the number of contacts was significantly reduced among individuals with moderate or severe symptoms compared to those with mild or no symptoms.

However, the direction of the effect was consistent with the hypothesis that symptomatic individuals reduce mixing, a behavioural adjustment highlighted in other social contact studies [24]. The limited power of this analysis, particularly due to the smaller sample of participants with severe symptoms, may partly explain the lack of significance. Age also emerged as a consistent determinant, with older adults (60+) reporting fewer contacts, consistent with findings from pre-pandemic social contact data too such as SoCRATES [25].For example, in the SOCRATES contact rate data tool, the mean contact rate for older people is 8, while our findings show a mean contact rate of 9, which are almost the same.

The GEE analysis on contact with fragile person further indicated that demographic and symptom-related determinants had only weak or non-significant effects. Multiple imputation was essential in this context, as complete-case analysis alone would have reduced power and potentially biased estimates, given the non-random pattern of missingness [26]. While multiple imputation improved the robustness of results, the lack of strong associations suggests that contact with fragile individuals may be driven by contextual or situational factors not captured in our dataset in addition to the sample (the number of individual in each category of the factor).

Simulation results reinforced the importance of behavioural assumptions made in epidemiological models. In baseline scenarios without behavioural adjustment, outbreaks were larger and extinction rates were lower. Incorporating contact reductions due to symptom onset and vaccination (as informed by the statistical model) increased extinction probabilities, demonstrating that failing to account for adaptive behaviours leads to biased epidemic projections. For instance, at 30% vaccination coverage, extinction rates increased from 0.63 under the baseline to 0.78 under the statistical model assumptions. Notably, scenarios assuming that vaccinated individuals doubled their contact rates produced both larger final epidemic sizes and substantially lower extinction probabilities (e.g., 0.46 at 60% coverage). This counterintuitive result illustrates the concept of a "leaky vaccine": although vaccination reduces susceptibility, its benefits can be offset if behavioural risk compensation occurs, where vaccinated individuals engage in higher levels of social mixing [27]. In such cases, the epidemiological protection of vaccines may be undermined by behavioural responses.

Taken together, these findings emphasize that both biological and behavioural mechanisms must be integrated into epidemic models. Models assuming homogeneous mixing or static behaviour risk overestimating or underestimating epidemic potential. Our study shows that realistic behavioural corrections particularly reductions in contact during symptomatic periods and vaccination-related behavioural changes are essential to capture epidemic dynamics.

However, several limitations must be acknowledged. First, the contact data were collected during a specific period of the winter season and may not capture seasonal variability in mixing patterns, which are known to affect respiratory disease transmission. Second, the study population was self-selected through the Infectieradar.be platform and may not be fully representative of the general population in terms of age, health status, or social behaviour, potentially limiting the generalizability of our findings. Third, symptom reports and contact numbers were self-reported and thus subject to recall bias and reporting bias, which could differ systematically between groups (e.g., across gender or vaccination status). Finally, the simulations simplified several aspects of epidemic dynamics by assuming fixed vaccine effectiveness, no waning immunity, and homogeneous transmission probability within age groups. These assumptions may not reflect real-world heterogeneity and could affect the projected epidemic trajectories.

Overall, this study demonstrates the value of combining empirical data, statistical models, and simulations to understand epidemic processes. Gender differences in symptom reporting, behavioral adjustments during symptoms, and risk compensation among vaccinated individuals were explored; however, we did not achieve sufficient statistical power to establish significance,

and the true adjustments could differ. These findings show that vaccination programs should include clear messages to discourage people from increasing their social interactions after getting vaccinated. They also suggest that epidemic models need to account for differences in how people behave to provide accurate and useful predictions for policymakers.Overall, our findings demonstrate the importance of integrating empirical contact data into epidemic models. Symptom onset and vaccination status influence behavior in small but important ways, and accounting for these dynamics can substantially improve estimates of epidemic potential and intervention impact.

5 Conclusion and Future Work

This study combined social contact data, statistical modeling, and individual-based simulations to assess how symptom onset and vaccination status may influence transmission dynamics of respiratory infections. Using generalized linear mixed models and generalized estimating equations, we found that contact behaviour was shaped by age and symptom reporting. Although women reported symptoms more frequently than men at the descriptive level, gender was not a significant predictor in the statistical models. Older adults consistently reported fewer contacts, in line with prior evidence. Reductions in contacts during symptomatic periods were modest and not statistically significant, but the direction of the effect aligned with behavioural expectations described in earlier studies.

The simulation analysis illustrated how different behavioural assumptions lead to different epidemic outcomes. Scenarios assuming no behavioural change produced larger epidemic sizes and lower extinction probabilities compared with scenarios incorporating contact reductions derived from the statistical models. Increasing vaccination coverage generally led to higher extinction probabilities, while a theoretical scenario where vaccinated individuals doubled their contact rates showed how behavioural compensation could counteract vaccine benefits. Although this doubling assumption is unlikely in practice, it provides insight into how sensitive epidemic outcomes are to behavioural changes following vaccination. Taken together, these findings highlight the importance of exploring a range of plausible behavioural assumptions when using simulation models to inform epidemic preparedness.

In conclusion, behavioural heterogeneity, particularly changes in social mixing during symptomatic periods and following vaccination, plays a central role in epidemic dynamics. Future epidemic models should explicitly incorporate such adaptive processes, as failing to do so risks overlooking important variability in epidemic trajectories. For public health, the findings emphasize the value of communication strategies that discourage risk-compensating behaviours after vaccination and encourage timely contact reduction when symptoms occur.

A natural next step is to extend the simulations to reflect a broader set of behavioural and epidemiological mechanisms. This could include waning immunity, reinfections, seasonality, and heterogeneity in vaccine effectiveness. Another avenue is to use models such as generalized

estimating equations to simulate not only overall contact counts but also the probability of interaction with fragile individuals, thereby providing a more nuanced picture of transmission risk. Future work should also validate the results against data collected in other regions or time periods to assess generalizability. Developing more adaptive simulation frameworks that integrate empirical findings with flexible behavioural assumptions will improve the policy relevance of epidemic forecasts.

Ethical Considerations

The data used in this study were not publicly available but were accessed through formal permission granted by the data provider for academic research purposes. All data were fully anonymized prior to access, and no personally identifiable information was included. The dataset was stored securely and handled in compliance with data protection and confidentiality standards. As the study involved secondary analysis of de-identified data and did not involve direct interaction with participants, no additional ethical approval was required. The research was conducted with full respect for participant privacy and in accordance with ethical research practices.

Stakeholder Awareness

The data for this study were obtained through the participatory surveillance platform Infectieradar.be, which engages citizens in real-time reporting of symptoms and social contacts. Such participatory systems not only provide timely data but also encourage community involvement in public health monitoring. The results of this study carry important implications for stakeholders: by showing how symptom severity and vaccination status shape contact behaviour, they highlight opportunities for more adaptive and targeted epidemic control strategies. For example, public health authorities can use these insights to refine risk communication, encouraging individuals to reduce contacts when symptomatic and discouraging compensatory increases in mixing after vaccination. Moreover, the findings underscore the value of integrating behavioural data into epidemic preparedness planning, allowing policymakers to anticipate realistic patterns of disease spread rather than relying on static or homogeneous assumptions. Ultimately, strengthening stakeholder awareness of these behavioural mechanisms can support the design of more effective, evidence-based interventions during future respiratory epidemics.

References

[1] Joël Mossong et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLOS Medicine*, 5(3):e74, 2008.

[2] World Health Organization. Influenza (seasonal). https://www.who.int/news-room/fact-sheets/detail/influenza-(seasonal), 2023. https://www.who.int/news-room/fact-sheets/detail/influenza-(seasonal).

[3] Kim Van Kerckhove, Niel Hens, W. John Edmunds, and Ken T.D. Eames. The impact of illness on social networks: implications for transmission and control of influenza. *American Journal of Epidemiology*, 178(11):1655–1662, 2013.

[4] S. Wambua, L. Willem, P. Beutels, P. Coletti, N. Goeyvaerts, N. Hens, et al. Individual-level determinants of contact behaviour during the covid-19 pandemic in 16 european countries. *BMC Public Health*, 23(1):1881, 2023.

[5] Kathy Leung, Mark Jit, Eric HY Lau, and Joseph T Wu. Social contact patterns relevant to the spread of respiratory infectious diseases in hong kong. *Scientific reports*, 7(1):7974, 2017.

[6] Y Ibuka et al. Social contacts, vaccination decisions, and influenza in japan. *Journal of Epidemiology and Community Health*, 64(6):502–507, 2010.

[7] Timo Smieszek et al. Collecting close-contact social mixing data with contact diaries: reporting errors and biases. *Epidemiology and Infection*, 140(4):744–752, 2012.

[8] Chad R Wells et al. The impact of vaccination on covid-19 outbreaks in the united states. *The Lancet*, 397(10287):1303–1312, 2021.

[9] Sebastian Funk, Erez Gilad, Chris Watkins, and Vincent AA Jansen. Modelling the influence of human behaviour on the spread of infectious diseases: a review. *Journal of the Royal Society Interface*, 7(50):1247–1256, 2010.

[10] Kiesha Prem et al. The effect of control strategies to reduce social mixing on outcomes of the covid-19 epidemic in wuhan, china: a modelling study. *The Lancet Public Health*, 5(5):e261–e270, 2020.

[11] Christopher I Jarvis et al. Quantifying the impact of physical distance measures on the transmission of covid-19 in the uk. *BMC medicine*, 18(1):1–10, 2020.

[12] World Health Organization. Global influenza programme: Surveillance standards for influenza, 2017.

[13] Geert Verbeke and Geert Molenberghs. Generalized linear mixed models–overview. In Marc A. Scott, Jeffrey N. Rouder, and Richard E. Schumacker, editors, *The Sage Handbook of Multilevel Modeling*, pages 127–140. SAGE Publications, 2013.

[14] Charles E. McCulloch, Shayle R. Searle, and John M. Neuhaus. *Generalized, Linear, and Mixed Models*. Wiley, 2004.

[15] Peter McCullagh and John A. Nelder. *Generalized Linear Models*. Chapman and Hall, 1989.

[16] William Gardner, Edward P. Mulvey, and Eric C. Shaw. Regression analyses of counts and rates: Poisson, overdispersed poisson, and negative binomial models. *Psychological Bulletin*, 118(3):392–404, 1995.

[17] Dominique Lord and Fred L. Mannering. Poisson regression, overdispersion, and quasi-likelihood methods. *Accident Analysis & Prevention*, 37(1):163–170, 2005.

[18] Kung-Yee Liang and Scott L. Zeger. Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22, 1986.

[19] Andreas Ziegler. *Generalized Estimating Equations*. Springer, 2011.

[20] Roderick J.A. Little and Donald B. Rubin. *Statistical Analysis with Missing Data*. Wiley, 3rd edition, 2019.

[21] Petra Klepac, Adam J. Kucharski, Andrew J.K. Conlan, Stephen M. Kissler, Maria L. Tang, Hannah Fry, and Julia R. Gog. Contacts in context: large-scale setting-specific social mixing matrices from the bbc pandemic project. *medRxiv*, 2020. Preprint.

[22] Sabra L Klein and Katie L Flanagan. Sex influences immune responses to viruses, and efficacy of prophylaxis and treatments for viral diseases. *BioEssays*, 38(7):682–690, 2016.

[23] Xi He, Eric H.Y. Lau, Peng Wu, Xilong Deng, Jian Wang, Xinxin Hao, Yiu Chung Lau, Jessica Y. Wong, Yu Guan, Xinghua Tan, Xiaohui Mo, Yanqing Chen, Baolin Liao, Weilie Chen, Fengyu Hu, Qing Zhang, Ming Zhong, Yanrong Wu, Lingzhai Zhao, Fang Zhang, Benjamin J. Cowling, Fang Li, and Gabriel M. Leung. Temporal dynamics in viral shedding and transmissibility of covid-19. *Nature Medicine*, 26:672–675, 2020.

[24] Kathy Leung, Mark Jit, Eric H.Y. Lau, and Joseph T. Wu. Individual-level variations in human contact patterns reveal the importance of demographic and temporal factors in infectious disease transmission. *Nature Communications*, 8(1):1–10, 2017.

[25] Dina Mistry, Maria Litvinova, Ana Pastore y Piontti, et al. Inferring high-resolution human mixing patterns for disease modeling. *Nature Communications*, 12(1):323, 2021.

[26] Ian R. White, Patrick Royston, and Angela M. Wood. Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine*, 30(4):377–399, 2011.

[27] Charles F Manski and Francesca Molinari. Estimating the covid-19 infection rate: Anatomy of an inference problem. *Journal of Econometrics*, 220(1):181–192, 2021.

Appendix

Symptom Trend by Occupation

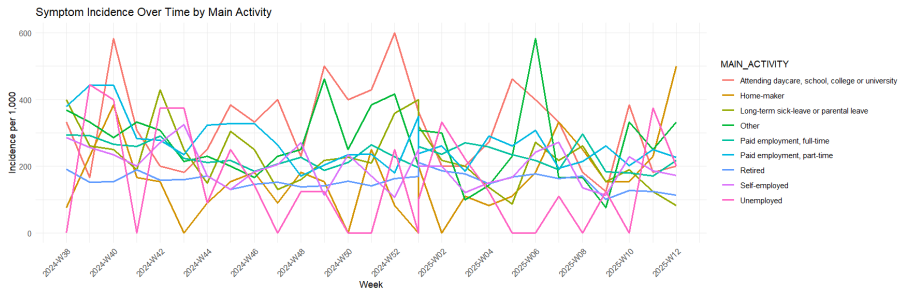

Figure 9: Symptom trend by occupational status

Symptom Trend by Flu Vaccination Status

Figure 10: Symptom trend by flu vaccination status

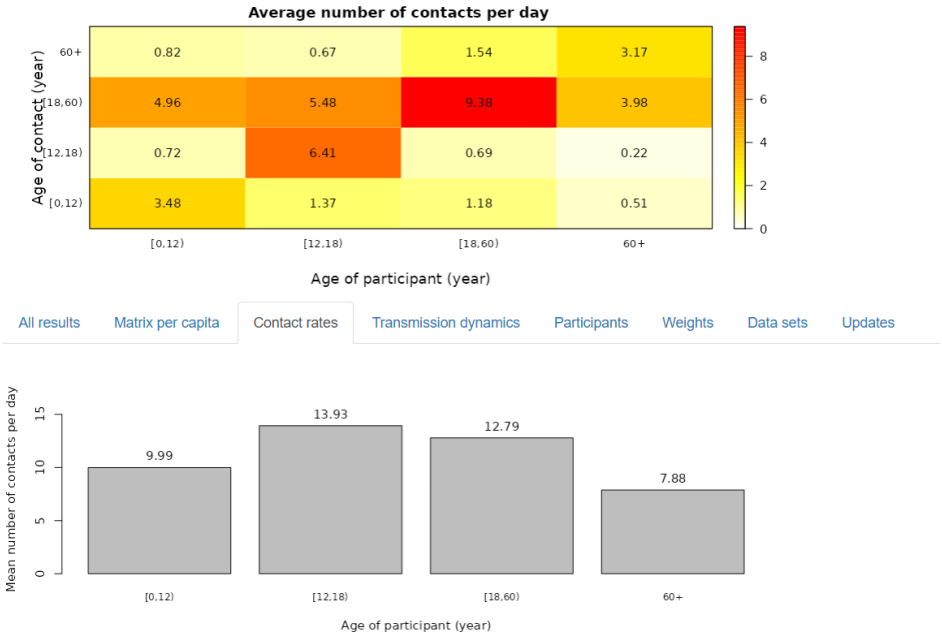

Figure 11: Social Contact Rates (SOCRATES) Data Tool :Belgium2006(Mossong 2008)

Table 6: Mean contacts and number of observations by week

Week Date	Mean Contact	Number of Observations
2024-10-14	15.10	94
2024-10-21	12.20	96
2024-10-28	14.37	78
2024-11-04	12.87	61
2024-11-11	8.49	67
2024-11-18	16.16	57
2024-11-25	14.47	62
2024-12-02	13.05	56
2024-12-09	16.25	63
2024-12-16	9.23	39
2024-12-23	10.47	57
2024-12-30	9.23	44
2025-01-06	6.57	53
2025-01-13	10.91	32
2025-01-20	14.80	61
2025-01-27	11.94	48
2025-02-03	9.89	45
2025-02-10	12.47	45
2025-02-17	9.49	53
2025-02-24	10.92	48
2025-03-03	14.69	45
2025-03-10	13.84	51
2025-03-17	8.05	21
2025-03-24	10.66	50
2025-03-31	13.08	37

Table 7: Poisson GLMM results (Complete Case vs Multiple Imputation)

Covariate	**Complete Case (CC)**			**Multiple Imputation (MI)**		
	Estimate	SE	p-value	Estimate	SE	p-value
Intercept	2.402	0.237	<0.001	2.345	0.150	<0.001
No/Mild symptoms	0.175	0.046	<0.001	-0.075	0.036	0.035
Age group: 60+	-0.704	0.122	<0.001	-0.675	0.099	<0.001
Sex: Male	-0.185	0.115	0.108	-0.144	0.105	0.170
Sex: Undefined	1.536	1.216	0.207	-0.352	0.976	0.718
Vaccinated	-0.199	0.243	0.414	-0.043	0.135	0.751

Table 8: Negative Binomial GLMM with Random intercept and slope

	Complete Case			Multiple Imputation		
Term	Estimate	Std. Error	p-value	Estimate	Std. Error	p-value
(Intercept)	2.369	0.226	< 0.001	2.304	0.220	< 0.001
symptom_severity_binaryNo/Mild symptoms	0.057	0.127	0.653	0.091	0.101	0.371
AgeG60+	-0.599	0.107	< 0.001	-0.560	0.088	< 0.001
SEXMale	-0.174	0.101	0.083	-0.113	0.084	0.181
SEXUndefined	1.271	0.948	0.180	0.051	0.828	0.951
vac_status_binaryVaccinated	-0.078	0.213	0.713	-0.144	0.209	0.491
Random intercept variance (participantId)	1.385	–	–	1.333	–	–
Random slope variance (week)	0.000912	–	–	0.000760	–	–

R-code

```
cont <- readRDS("C:/Users/ademu/Desktop/Uhasselt/SYSS/Thesis/updated
    dataset/MT_CON.rds")
cont2 <- readRDS("C:/Users/ademu/Desktop/Uhasselt/SYSS/Thesis/updated
    dataset/MT_CON2.rds")
intake <- readRDS("C:/Users/ademu/Desktop/Uhasselt/SYSS/Thesis/updated
    dataset/MT_intake.rds")
vac <- readRDS("C:/Users/ademu/Desktop/Uhasselt/SYSS/Thesis/updated
    dataset/MT_VAC.rds")
symp <- readRDS("C:/Users/ademu/Desktop/Uhasselt/SYSS/Thesis/updated
    dataset/MT_weekly.rds")
# Prepare contact counts
contact_count <- cont2 %>%
  mutate(ADT = as.Date(ADTM), week = isoweek(ADT), value_num = as.
      numeric(value)) %>%
  filter(!is.na(value_num)) %>%
  group_by(participantId, week) %>%
  summarise(contact_number = sum(value_num, na.rm = TRUE), .groups = "
      drop")

# Base contact dataset
cont_base <- cont %>%
  mutate(ADT = as.Date(ADTM), week = isoweek(ADT), participantId =
      participantID) %>%
  select(participantId, ADT, week, Contacts, Fragile_NONE)

# Merge and clean
cont_with_contacts <- cont_base %>%
  left_join(contact_count, by = c("participantId", "week")) %>%
  mutate(contact_number = if_else(Contacts == "No", 0, contact_number))

symp_clean <- symp %>%
  mutate(ADT = as.Date(ADTM), week = isoweek(ADT)) %>%
  group_by(participantID, week) %>% slice(1) %>% ungroup()

intake_clean <- intake %>% group_by(participantID) %>% slice(1) %>%
    ungroup()
vac_clean    <- vac %>% group_by(participantID) %>% slice(1) %>%
    ungroup()

# Merge datasets
merged_data <- cont_with_contacts %>%
  left_join(symp_clean, by = c("participantId" = "participantID", "week
      ")) %>%
  left_join(intake_clean, by = c("participantId" = "participantID"))
      %>%
  left_join(vac_clean, by = c("participantId" = "participantID"))
```

```
# Symptom severity
names(merged_data)[names(merged_data) == "No symptoms"] <- "no_symptoms
    "
mild_symptoms <- c("Sneezing", "Runny or blocked nose", "Watery
    bloodshot eyes")
modsev_symptoms <- c("Cough", "Sore throat", "Headache", "Chills", "
    Nausea",
                     "Muscle/joint pain", "Shortness of breath", "Chest
                         pain",
                     "Fever", "Confusion", "Vomiting")

merged_data <- merged_data %>%
  rowwise() %>%
  mutate(
    mild_count = sum(c_across(all_of(mild_symptoms)) == "True", na.rm =
        TRUE),
    modsev_count = sum(c_across(all_of(modsev_symptoms)) == "True", na.
       rm = TRUE),
    symptom_severity_binary = case_when(
      no_symptoms == "True" & mild_count == 0 & modsev_count == 0 ~ "No
         /Mild symptoms",
      modsev_count > 0 ~ "Moderate/Severe symptoms",
      mild_count > 0 ~ "No/Mild symptoms",
      TRUE ~ NA_character_
    )
  ) %>%
  ungroup()

# Recode
merged_data <- merged_data %>%
  mutate(
    vac_status_binary = case_when(
      FLU_VACCINE_2425 == "Yes" | VACCINATION_STATUS_COV2425 == "Yes" ~
          "Vaccinated",
      FLU_VACCINE_2425 %in% c("No", "I don't know") &
        VACCINATION_STATUS_COV2425 %in% c("No plan to receive
            vaccination", "No will not get vaccinated", "I don't know")
            ~ "Not vaccinated",
      TRUE ~ NA_character_
    ),
    AgeG = case_when(
      AGE >= 18 & AGE <= 59 ~ "18-59",
      AGE >= 60 ~ "60+",
      TRUE ~ NA_character_
    ),
    Fragilestatus = ifelse(Fragile_NONE == "True", 1, 0)
  )
```

```
# Final model data
model_data <- merged_data %>%
  select(participantId, contact_number, symptom_severity_binary, AgeG,
     SEX, vac_status_binary, Fragilestatus) %>%
  mutate(participantId = as.factor(participantId))

# Impute missing
imputed_data <- mice(model_data, m = 5, method = "pmm", seed = 123)
completed_data <- complete(imputed_data, 1) %>%
  mutate(across(c(symptom_severity_binary, AgeG, vac_status_binary, SEX
     , participantId, Fragilestatus), as.factor))

# Poisson model
poisson_glmm <- glmer(
  contact_number ~ symptom_severity_binary + AgeG + SEX + vac_status_
     binary + (1 | participantId),
  data = completed_data,
  family = poisson()
)
summary(poisson_glmm)

# NegBin model
nb_glmm <- glmmTMB(
  contact_number ~ symptom_severity_binary + AgeG + SEX + vac_status_
     binary + (1 | participantId),
  data = completed_data,
  family = nbinom2()
)
summary(nb_glmm)

# GEE
gee_data <- completed_data %>%
  filter(SEX %in% c("Male", "Female")) %>%
  mutate(Fragilestatus = as.numeric(as.character(Fragilestatus)))

gee_fragile <- geeglm(
  Fragilestatus ~ symptom_severity_binary + AgeG + SEX + vac_status_
     binary,
  data = gee_data,
  id = participantId,
  family = binomial(link = "logit"),
  corstr = "exchangeable"
)
summary(gee_fragile)
###
Code For simulation\\
```

```
######################
# Set shared simulation parameters


n <- 1000
age.distr <- c(0.25, 0.25, 0.25, 0.25)
ct.rates <- c(10, 14, 16, 10)
nSeeds <- 1
R0 <- 1.5
inf.param <- R0 / (sum(ct.rates * age.distr))
pathogen <- "FLU" #or "COVID-19"
bc <- 1
veff <- 0.5
vaccov <- 0.3
t.stop <- 100
nSim <- 100
nSeed <- 1062021
set.seed(nSeed)
#############
# Load simulation function
source("SimFunctions.R")

# Shared parameters

coverages<-c(0.3,0.6,0.9)
bc_stat<-rep(0.9,4)
vac_stat<-rep(0.9,4)

# 4. Vaccination coverage & effectiveness scenarios
for (cov in coverages) {
  # Baseline coverage
  scenarios[[paste0("Coverage", cov, "_Baseline")]] <- list(bc = rep(1,
      4), vac = rep(1, 4), vaccov = cov)

  # Statistical model coverage
  scenarios[[paste0("Coverage", cov, "_StatModel")]] <- list(bc = bc_
     stat, vac = vac_stat, vaccov = cov)

  # Additional: vaccinated twice as many contacts as unvaccinated
  scenarios[[paste0("Coverage", cov, "_Vac2x")]] <- list(bc = rep(0.9,
     4), vac = rep(1.9, 4), vaccov = cov)

  # Low and high symptom contact reduction, vaccination fixed
  scenarios[[paste0("Coverage", cov, "_SympLow_VacFixed")]] <- list(
    bc = rep(1.5, 4),   # Low reduction due to symptoms
    vac = vac_stat,     # Fixed from statistical model
    vaccov = cov
  )
```

```
  scenarios[[paste0("Coverage", cov, "_SympHigh_VacFixed")]] <- list(
    bc = rep(0.1, 4),   # High reduction due to symptoms
    vac = vac_stat,     # Fixed from statistical model
    vaccov = cov
  )

  }

# Run all scenarios
results <- list()
extinction_rates <- data.frame(Scenario = character(), ExtinctionRate =
    numeric())

for (sc_name in names(scenarios)) {
  cat("Running:", sc_name, "\n")
  scen <- scenarios[[sc_name]]

  epi.list <- replicate(nSim, sim.epi(
    n = n,
    ct.rates = ct.rates,
    age.distr = age.distr,
    nSeeds = nSeeds,
    inf.param = inf.param,
    pathogen = pathogen,
    contact.reduction.bc = scen$bc,
    contact.reduction.vac = scen$vac,
    t.stop = t.stop,
    bc = 1,
    veff = veff,
    vaccov = scen$vaccov
  ), simplify = FALSE)

  # Final sizes
  final_sizes <- sapply(epi.list, function(x) x$FinalSize$FinalSize1)
  results[[sc_name]] <- final_sizes

  # Extinction rate ( 10 % infected)
  ext_rate <- mean(final_sizes <= 0.1 * n)
  extinction_rates <- rbind(extinction_rates,
                            data.frame(Scenario = sc_name,
                               ExtinctionRate = round(ext_rate, 3)))
}

# Combine results for plotting
plot_data <- data.frame(
  FinalSize = unlist(results),
  Scenario = rep(names(results), each = nSim)
```

```
)

# Plot
boxplot(FinalSize ~ Scenario, data = plot_data,
        las = 2, col = rainbow(length(results)),
        main = "Final Epidemic Sizes by Scenario",
        ylab = "Final size (number infected)")

par(mar = c(12, 5, 4, 2) + 0.1)  # increase bottom margin for labels

boxplot(FinalSize ~ Scenario, data = plot_data,
        las = 2,                      # rotate labels to vertical
        cex.axis = 0.7,               # smaller axis text
        col = rainbow(length(results)),
        main = "Final Epidemic Sizes by Scenario",
        ylab = "Final size (number infected)")

# Extinction table
print(extinction_rates)

# Save extinction table for Overleaf
library(xtable)
xtable(extinction_rates, caption = "Extinction rates for all simulation
    scenarios")


>´
library(ggplot2)plot_data<- subset(plot_data,
                   !(Scenario %in% c("Baseline", "StatModel","VacFixed_
                      SympVarHigh","VacFixed_SympVarLow")))
# Assign grouping category for coloring
plot_data$Group <- ifelse(grepl("Baseline", plot_data$Scenario), "
   Baseline",
                          ifelse(grepl("Vac2x", plot_data$Scenario), "
                             Vac_2x",
                                 ifelse(grepl("Vac0.95", plot_data$
                                    Scenario), "Vac_0.95",
                                        ifelse(grepl("BC0.9", plot_data
                                           $Scenario), "BC_0.9", "
                                           StatModel"))))

# Custom colors for groups
group_colors <- c("Baseline" = "blue",
                  "BC_0.9" = "orange",
                  "Vac_0.95" = "black",
                  "Vac_2x" = "red",
                  "StatModel" = "green")
```

```
ggplot(plot_data, aes(x = Scenario, y = FinalSize, fill = Group)) +
  geom_boxplot() +
  scale_fill_manual(values = group_colors) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))


>´
###################NoNE# Filter only non-extinct outbreaks
non_extinct_df <- subset(plot_data, FinalSize > 0.10 * n)

# Assign grouping category for coloring
non_extinct_df$Group <- ifelse(grepl("Baseline", non_extinct_df$
   Scenario), "Baseline",
                               ifelse(grepl("Vac2x", non_extinct_df$
                                  Scenario), "Vac_2x",
                                      ifelse(grepl("Vac0.95", non_
                                         extinct_df$Scenario), "Vac_
                                         0.95",
                                             ifelse(grepl("BC0.9", non_
                                                extinct_df$Scenario), "
                                                BC_0.9", "StatModel")))
                                                )

# Custom colors for groups
group_colors <- c("Baseline" = "blue",
                  "BC_0.9" = "orange",
                  "Vac_0.95" = "black",
                  "Vac_2x" = "red",
                  "StatModel" = "green")

# Plot
library(ggplot2)
ggplot(non_extinct_df, aes(x = Scenario, y = FinalSize, fill = Group))
   +
  geom_boxplot() +
  scale_fill_manual(values = group_colors) +
  theme_minimal(base_size = 14) +
  labs(title = "",
       y = "Final Size (Number of Infected)",
       x = "Scenario") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))


>´
# Filter non-extinct outbreaks and remove "Baseline" and "StatModel"
   alonenon_extinct_df <- subset(plot_data,
                          FinalSize > 0.10 * n &
```

```
                          !(Scenario %in% c("Baseline", "StatModel")))

# Assign grouping category for coloring
non_extinct_df$Group <- ifelse(grepl("Baseline", non_extinct_df$
   Scenario), "Baseline",
                               ifelse(grepl("Vac2x", non_extinct_df$
                                  Scenario), "Vac_2x",
                                      ifelse(grepl("Vac0.95", non_
                                         extinct_df$Scenario), "Vac_
                                         0.95",
                                             ifelse(grepl("BC0.9", non_
                                                extinct_df$Scenario), "
                                                BC_0.9", "StatModel")))
                                                )

# Custom colors for groups
group_colors <- c("Baseline" = "blue",
                  "BC_0.9" = "orange",
                  "Vac_0.95" = "black",
                  "Vac_2x" = "red",
                  "StatModel" = "green")

# Plot
ggplot(non_extinct_df, aes(x = Scenario, y = FinalSize, fill = Group))
   +
  geom_boxplot() +
  scale_fill_manual(values = group_colors) +
  theme_minimal(base_size = 14) +
  labs(title = "",
       y = "Final Size (Number of Infected)",
       x = "Scenario") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))


>´
        The source For the above Code(simFunction)\\        #
           ##########################################################################

# Set of function for the co-infection script
#
   ##########################################################################


#function used to define the epidemiological progression
incubation.period<-function(pathogen){
  if (pathogen=="COVID-19"){
    return(rlnorm(1,meanlog = 1.54, sdlog = 0.47)) # Zhang et al. 2020
       DOI: 10.1016/S1473-3099(20)30230-9
```

```
  }
  if (pathogen=="FLU"){
    return(rlnorm(1,meanlog = log(1.4), sdlog = log(1.51))) # Lessler
       et al. 2009 DOI: 10.1016/S1473-3099(09)70069-6
  }
}

infection.period.length<-function(pathogen){
  if (pathogen=="COVID-19"){
    return(12) # 10 days of infecious period + 2 days of latent period
  }
  if (pathogen=="FLU"){
    return(8)  # Carrat et al. 2008
  }
}

infectiousness.measure<-function(pathogen,t){
  if (pathogen=="COVID-19"){
    if (t<2){ #Latent period - https://doi.org/10.1016/j.ijid
       .2020.06.036
      return(0)
    }else{
      return(1/10)
    }


>´
  }  if (pathogen=="FLU"){
    if (t<1.6){ #Latent period - https://doi.org/10.1016/j.epidem
       .2012.06.001
      return(0)
    }else{
      return(1/6.4)
    }


>´
  }
}


>´
prob.symptinf<-function(pathogen){  if (pathogen=="COVID-19"){
    return(0.69) #
  }
  if (pathogen=="FLU"){
    return(0.67)  # Carrat et al. 2008
  }
```

```
}

#Computing Rt - started but haven't look at this anymore
# Rt
comp.RT<-function(status.matrix,individual,Rt){
  infectees<-which(status.matrix$infector==individual) #people infected
      by individual
  Rt.temp<-0
  if (length(infectees)>0){
    for (i1 in 1:length(infectees)){
      if (status.matrix$time.of.infection[infectees[i1]]>status.matrix$
         time.of.infection[individual]){
        Rt.temp<-Rt.temp+1
      }
    }
  }
  Rt<-rbind(Rt,c(status.matrix$time.of.infection[individual],Rt.temp))
  return(Rt)
}


>´
#Creating a synthetic population#input: age.distr -> frequency/total
   pop per class
pop.matrix<-function(n, age.distr){
  age.group<-sample(c("C","T","A","E"),n, prob = age.distr, replace =
     TRUE)
   status.matrix<- data.frame(infected          = rep(0,n),      # 1
                              time.of.infection  = NA,           # 2
                              infector           = NA,           # 3
                              severity           = 0,            # 4 1
                                 Symptomatic 2 Asymptomatic
                              TimeSymptomOnset   = Inf,          # 5
                              Vaccinated         = 0,            # 6 no
                                 immunity, 1 vaccinated
                              AgeGroup           = age.group,    # 7 "C"
                                  : 0-11 / "T" : 12-17 / "A" : 18 -65 /
                                  "E": 65+
                              Recovery           = Inf)          # 8
   return(status.matrix)
}

sim.epi<-function(n,ct.rates, age.distr, nSeeds, inf.param, pathogen,
   contact.reduction.bc,contact.reduction.vac,t.stop,bc, veff, vaccov){

  status.matrix <-pop.matrix(n, age.distr)
```

```
#type of the events that can happen in this simulator
events<-data.frame("NextCtc"=Inf, "BehavChange"=Inf, "Recovery"=Inf)

infectives<-rep(0,n) # vector that indicates who is infectious at the
    current time: 1 infectious 0 non infectious
current.time<-0
index.contact<-rep(0,n) # vector that selects the individuals that
   have to propose a new social contact - 1 yes 0 no
time.events<-matrix(NA,1,3)

#transmission parameter dataframe: each line is an individual, the
   first colum is the ID, the second and third the transmission
   parameter given  household or global contacts for pathogen 1, and
   fourth and fifth for pathogen 2
transmission.parameters<-data.frame("id"=1:n,"q"=rep(inf.param,n),"
   contact_rate"=rep(NA,n))   #matrix containing the proposed time of
    the next contact (first column) and the contact individual (
   second column)

# Vaccination - we assume individual are vaccinated at the start of
   the outbreak, and vaccine effectiveness is constant
nvacc<-round(n*vaccov)
if (nvacc>0){
status.matrix$Vaccinated[sample(1:n,nvacc)]<-1
}


# setting contact rate according to the age group
for (i in 1:n){
  if (status.matrix$Vaccinated[i]==0){
    transmission.parameters$contact_rate[i]<-ct.rates[which(c("C","T"
        ,"A","E")==status.matrix$AgeGroup[i])]*inf.param
  }else{
    transmission.parameters$contact_rate[i]<-ct.rates[which(c("C","T"
        ,"A","E")==status.matrix$AgeGroup[i])] *inf.param* contact.
        reduction.vac[which(c("C","T","A","E")==status.matrix$AgeGroup
        [i])]
  }
}

#fixed epi parameters
rho<-prob.symptinf(pathogen)


#Keep track at the moment individuals make contacts
contact.time<-data.frame("id"=1:n,"pr.ctc"=rep(NA,n))

# first infected: randomly chosen in the population
```

```
first.cases<-sample(n,nSeeds)

# Day at which individuals change their contact behavior
bc.day<-rep(Inf,n)


for (j in first.cases){
  first<-j
  status.matrix$infected[first] <- 1
  status.matrix$time.of.infection[first] <- 0
  status.matrix$Recovery[first]<-current.time+infection.period.length
     (pathogen = pathogen)
  if (runif(1)<rho){ #if symptomatic #index cases are always
     symptomatic individuals
  status.matrix$TimeSymptomOnset[first]<-current.time+incubation.
     period(pathogen=pathogen)
  if(runif(1)<bc){
    bc.day[first]<-status.matrix$TimeSymptomOnset[first]
  }
  status.matrix$severity[first]<-1
  time.events<-rbind(time.events,c(current.time,1.1,first))
  }
  else{
  status.matrix$severity[first]<-2
  time.events<-rbind(time.events,c(current.time,1.2,first))
  }
  infectives[first]<-1
  contact.time$pr.ctc[first]<-rexp(1,transmission.parameters$contact_
     rate[first])+current.time # I generate the next interarrival
     time for individual i
}

proposed.individual<-0
temp.contact.time<-0
indiv.prop.ctc<-0
recovered<-0
Rt1<-matrix(data = NA, nrow = 1, ncol = 2)


while(sum(infectives)>0 & current.time<t.stop){ #while there are
   still infectives, we are within the t.stop
  #Phase 1: individuals that has to, propose a new social contact
  for (i in which(index.contact==1) ){ # for all the individuals that
      has to propose a global contact
    contact.time$pr.ctc[i]<-rexp(1,transmission.parameters$contact_
       rate[i])+current.time# I generate the next interarrival time
       for individual i
    index.contact[i]<-0
```

```
}

#Phase 2: identify the next event: select the minimum time among
   the events that can occur
ifelse(length(which(is.na(contact.time$pr.ctc)==FALSE))>0,events$
   NextCtc<-min(contact.time$pr.ctc, na.rm = T),events$NextCtc<-Inf
   ) # among all the proposed social contact between houeholds we
   select the minimum
ifelse(length(which(!is.infinite(bc.day)))>0,events$BehavChange<-
   min(bc.day),events$BehavChange<-Inf ) #minimum quarantine
   pathogen 1
ifelse(length(which(is.na(status.matrix$Recovery)==FALSE))>0,events
   $Recovery<-min(status.matrix$Recovery, na.rm = T),events$
   Recovery<-Inf) # among all the proposed social contact between
   houeholds we select the minimum

next.evts<-colnames(events)[which(min(events)==events)] # if more
   than one event is occuring at a specific time, just sample one
if (length(next.evts)>1){
  next.evts<-sample(colnames(events)[which(min(events)==events)],1)
}

if (next.evts=="NextCtc"){

  # identify the infector
  current.time<-events$NextCtc
  if (length(which(contact.time$pr.ctc == events$NextCtc))>1){ #
     when two contacts happen at the same time select one of the
     two
    infector<-sample(which(contact.time$pr.ctc == events$NextCtc)
       ,1)
  }else{
    infector<-which(contact.time$pr.ctc == events$NextCtc)
  }
    infectee.pool<- setdiff(1:n,infector) #individuals not in the
       same class
    infectee<-sample(infectee.pool,1) #pick a random individual not
        in the class (and not a teacher)
    index.contact[infector]<-1
    contact.time$pr.ctc[infector]<-NA

  #Loop to check whether an infection event takes place

  if (status.matrix$infected[infector]==1 & status.matrix$infected[
     infectee]==0){
    ifelse(status.matrix$Vaccinated[infectee]==1, tempVE<-veff,
       tempVE<-0)
```

```
      acc.rate<-(1-tempVE)*infectiousness.measure(pathogen = pathogen
          , t=(current.time- status.matrix$time.of.infection[infector
          ]))
      if (runif(1)<acc.rate){ # bernoulli experiment to characterize
          infection
        status.matrix$infected[infectee] <- 1
        status.matrix$time.of.infection[infectee] <- current.time
        status.matrix$infector[infectee] <- infector
        status.matrix$Recovery[infectee]<-current.time+infection.
            period.length(pathogen = pathogen)
        if (runif(1)<rho){ #if symptomatic #index cases are always
            symptomatic individuals
          status.matrix$TimeSymptomOnset[infectee]<-current.time+
              incubation.period(pathogen=pathogen)
          status.matrix$severity[infectee]<-1
          if(runif(1)<bc){
            bc.day[infectee]<-status.matrix$TimeSymptomOnset[infectee
                ]
          }
          time.events<-rbind(time.events,c(current.time,1.1,infectee)
              )
        }else{
          status.matrix$severity[infectee]<-2
          time.events<-rbind(time.events,c(current.time,1.2,infectee)
              )
        }
        infectives[infectee]<-1
        contact.time$pr.ctc[infectee]<-rexp(1,transmission.parameters
            $contact_rate[infectee])+current.time # I generate the
            next interarrival time for individual i
      }
      }

    }

  if (next.evts=="BehavChange"){ #next event is changing the contact
      rate
    current.time<-events$BehavChange
    bc.individuals<-which(bc.day==current.time)
    for (k in bc.individuals){
      transmission.parameters$contact_rate[k]<-transmission.
          parameters$contact_rate[k]*contact.reduction.bc[which(c("C",
          "T","A","E")==status.matrix$AgeGroup[k])]
      contact.time$pr.ctc[k]<-rexp(1,transmission.parameters$contact_
          rate[k])+current.time # I generate the next interarrival
          time for individual i
    }
    bc.day[bc.individuals]<-Inf
```

```
  }

  if (next.evts=="Recovery"){ #recovery from infection
    current.time<-events$Recovery
    temp.recovered<-which(status.matrix$Recovery==events$Recovery)
    for (recovered in temp.recovered){
          Rt1<-comp.RT(status.matrix = status.matrix,individual =
             recovered,Rt=Rt1)
          status.matrix$infected[recovered]<--1
          status.matrix$Recovery[recovered]<-Inf
          infectives[recovered]<-0
          contact.time$pr.ctc[recovered]<-NA
          index.contact[recovered]<-0
          time.events<-rbind(time.events,c(current.time,-1,recovered)
             )
    }
  }

}

#When also the other pathogen is present.
time.events<-time.events[-1,]
timev.name<-c("time","event","who")
dimnames(time.events)<-list(NULL,timev.name)

# compute some summary measures that will be given as output

C1<-nSeeds
Y1<-nSeeds
last.day<-round(max(time.events[,1], na.rm = T))

for (i in 1:last.day){
  temp.time<-setdiff(which(time.events[,1]>i),which(time.events[,1]>i
     +1))
  temp.inf.1<-c(which(time.events[temp.time,2]==1.1),which(time.
     events[temp.time,2]==1.2))
  temp.time.1<-setdiff(1:length(time.events[,1]),which(time.events
     [,1]>i+1))
  C1<- c(C1,length((which(time.events[temp.time.1,2]==1.1)))+length((
     which(time.events[temp.time.1,2]==1.2)))-length((which(time.
     events[temp.time.1,2]==-1))))
  Y1<- c(Y1,length(temp.inf.1))
}

Fs1<-length(which(time.events[,2]==1.1))+length(which(time.events
   [,2]==1.2))
```

```
  epi.details<-data.frame("Days"=0:last.day, "Incidence1"=Y1, "
     Prevalence1"=C1)
  FinalSize<-data.frame("FinalSize1"=Fs1)
  PeakIncidence<-data.frame("PeakIncidence"=max(epi.details$Incidence1)
     ,"TimePeakIncidence1"=which(epi.details$Incidence1==max(epi.
     details$Incidence1))[1])
  PeakPrevalence<-data.frame("PeakPrevalence1"=max(epi.details$
     Prevalence1),"TimePeakPrevalence1"=which(epi.details$Prevalence1==
     max(epi.details$Prevalence1))[1])
  return(list(time.events=time.events, status.matrix=status.matrix,epi.
     details=epi.details, FinalSize=FinalSize, PeakIncidence=
     PeakIncidence, PeakPrevalence=PeakPrevalence,Rt1=Rt1))
}
```

Survey Questions about the Contact

1. Social contact between yesterday 5 a.m. and today 5 a.m.

- Yes
- No
- I don't want to say

2. Settings of contacts

Please select all the settings that apply:

- Home
- Work
- School
- Leisure
- Other

3. Number of contacts per age category and gender

Indicate the number of contacts at:

- Home

- Work
- School
- Leisure
- Other activities

Table template:

Age range	Female	Male
0–3		
3–6		
7–12		
13–18		
19–29		
30–39		
40–49		
50–59		
60–69		
70–79		
80–89		
90+		

4. Visits to facilities with fragile people

Between yesterday 5 a.m. and today 5 a.m., did you visit an institute with (many) fragile people?

- No
- Yes, a care home
- Yes, a facility for assisted living
- Yes, a healthcare center other than a hospital (e.g., general practitioner, physiotherapist, vaccination clinic)
- Yes, a hospital
- Yes, the palliative care unit of a hospital (hospice)
- Other: _____