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

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

UHASSELT KNOWLEDGE IN ACTION

Maastricht University

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

Master's thesis

Unravelling non-household contact patterns during the COVID-19 pandemic in Belgium

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

SUPERVISOR :
Prof. dr. Niel HENS

2024
2025

UHASSELT

KNOWLEDGE IN ACTION

Maastricht University

Faculty of Sciences

School for Information Technology

Master of Statistics and Data Science

Master's thesis

Unravelling non-household contact patterns during the COVID-19 pandemic in Belgium

Stijn Lapere

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

SUPERVISOR :

Prof. dr. Niel HENS

Acknowledgement

I would like to take the opportunity to thank everyone who supported me during this master's thesis project, and throughout the past years.

First and foremost, I would like to thank my supervisor Prof. Dr. Niel Hens for suggesting this topic, his guidance throughout the process and valuable feedback. I also want to express my gratitude to Dr. James Wambua for giving access to the data and Mr. Neilshan Loedy, who will hopefully have obtained his doctoral degree in Science by the time this thesis is defended, for sharing his code and answering all of my questions. Without their help, I would not have been able to develop myself academically to the point I am right now. During the constructive discussions in our meetings, Prof. Dr. Niel Hens gave me tools how to search for solutions to problems I was tackling with and allowed me to take the initiative in making decisions about the research instead of just giving solutions or strict guidelines, which is how a researcher also has to deal with a problem. His expertise and critical thinking made sure I could explore new insightful directions to try to meaningfully interpret the obtained results. I am also grateful for the suggestions made by Prof. Dr. Andrea Torneri and Dr. Liesbeth Bruckers during the intermediate presentations.

Furthermore, I would like to express my gratitude towards my parents for their warmful support. Writing this thesis was far from easy and I had to overcome multiple obstacles. However, as I am writing this section, I realise that this bumpy track has made me to the person I am right now. This was not possible without my parents and brother, but all my friends also deserve some credits. In order to avoid forgetting to mention someone, I just want to thank everyone for their much appreciated distraction and encouragement. If it was going for lunch or a coffee together, having a beer at the pub, or a different activity, it all contributed in a certain way to this thesis.

Unfortunately, submitting this master's thesis also means the end of my student period. It all started at the bachelor's program of mathematics at the KULAK in Kortrijk as a new adventure. 8 years and (hopefully) 3 master's degrees later, after a journey across all of Flanders, ending in Hasselt with a stopover in Leuven, this adventure has come to an end. I will definitely miss all of the experiences, but these will be memories that I will remember for the rest of my life. Together with my engagement as a student representative at KU Leuven, this study period has greatly enriched myself.

Stijn, June 2025

Abstract

Background: As a longitudinal study, the CoMix study was employed to monitor social behavior during the COVID-19 pandemic in, amongst others, Belgium. Dynamics of transmission are driven by human behavior, especially for this airborne disease.

Objectives: This study aimed to analyse how non-household contacts changed across age groups and time between December '20 and March '22. More specifically, the drivers that influenced the presence and reported number of non-household contacts based on the CoMix study in Belgium were explored by building statistical models. Furthermore, implications for transmission of COVID-19 in the population were made.

Methodology: Both negative binomial (NBI) and generalised Poisson (GPO) generalised additive models for location, scale and shape (GAMLSS) model per age category were built to examine the factors that influenced the average number of reported non-household contacts. Due to the excess of zeros in the reported number of non-household contacts, Hurdle models were employed to also explore the drivers of reporting contacts outside the household. Based on the contact locations, contact patterns of the participants were found based on an agglomerative hierarchical clustering algorithm.

Results: Although the lockdown restrictions emposed by the government were gradually relaxed throughout the study period, the average number of reported non-household contacts remained constant during this period. The GAMLSS and Hurdle models revealed the presence of under-reporting due to survey fatigue effect where fewer non-household contacts were reported as respondents participated in more survey rounds. Children reported on average more non-household contacts than adults and elderly. Whereas wearing a face mask and being vaccinated were consistently positively associated with the presence and number of non-household contacts, males and participants living in larger households had lower odds of reporting contacts outside their household. Service employees reported on average more non-household contacts, in contrast to the low number of contacts reported by participants who are not in labor force. More non-household contacts were reported with increasing income level. Most of the non-household contacts made by children and adults took place at school and work, respectively, while for elderly these type of contacts were mostly occurring at home.

Conclusions: Based on data from the CoMix study, several factors were associated with the presence and number of reported non-household contacts. These drivers may have an impact on the spread/transmission of COVID-19 in the population. The average number of reported non-household contacts did not considerably increase after relaxing the lockdown restrictions indicating the longer-term impact of the pandemic on social contact behavior. However, the under-reporting due to participant survey fatigue has to be taken into account as well.

Key Words: CoMix study, COVID-19, Social contact data, GAMLSS, Hurdle models

Contents

1 Introduction

At early 2020, the COVID-19 outbreak became a global pandemic which changed the daily lifes of many people. Since the SARS-CoV-2 virus, causing the airborne disease COVID-19, is transmitted via contacts in physical proximity, non-pharmaceutical measures (NPIs) were imposed to reduce the transmission rate. In Belgium, the number of contacts per person was reduced by e.g. closing schools, requiring employees to work remotely and limiting the number of contacts allowed to be made outside the household. Béraud et al. (2015) highlighted the potential high impact of home quarantine or school closures on the contact and transmission rate. Moreover, it became mandatory to keep a safe distance from each other and wear face masks in public places. Coletti et al. (2020) reported that the proportion of participants in their study reporting to wear a face mask increased from 18% at the end of April 2020 to 75% at the end of July, since more non-household contacts were made in the latter period and face mask wearing was obligatory.

From the beginning of May 2020 onwards, lockdown measures were lifted stepwise but tightened again starting from the end of July 2020 due to the Alpha variant. In fact, the number of hospital admissions was higher in early November 2020 than before the summer, leading to an even stricter lockdown including an evening curfew. In January 2021, a vaccination campaign started in Belgium. The measures were relaxed from April 2021 onwards as the vaccination coverages were increasing. At the end of July 2021, 60% of the total population of Belgium was vaccinated, whereas this was 75% by the end of 2021 (Kremer et al., 2023). In July 2021, the COVID safe ticket system was introduced as a certificate proving vaccination or recent recovery used to have access to public events.

Data on social contact behavior in the population is of pivotal importance in understanding the dynamics of virus transmission, since COVID-19 is a close-contact transmitted disease. Willem et al. (2021) and Kremer et al. (2023) noted the effectiveness of the contact tracing system (CTS) to reduce onward transmission. In order to investigate the effect of NPIs on the contact behavior, contact surveys were employed in many countries. In 21 European countries, the CoMix study was rolled out to collect contact data, resulting into information on the impact of COVID-19 and NPIs on behavioral changes (Wong et al., 2023). Previous research already found the reduction of the number of contacts over time in the first periods of the pandemic compared to pre-pandemic levels (Coletti et al., 2020; Gimma et al., 2022; Tizzani et al., 2023; Wong et al., 2023; Veneti et al., 2024). However, the number of social contacts remained below pre-pandemic levels after relaxing the NPIs (Loedy et al., 2023; Backer et al., 2024; Reichmuth et al., 2024) indicating the long-lasting impacts of the pandemic on individuals' behavior. This also has an impact on the reproduction number (Jarvis et al., 2024).

Prior studies also focused on the effect of risk perceptions (Wambua et al., 2022; Wambua et al., 2023), frailty (Loedy et al., 2025) and pregnancy (Wong et al., 2022) on the contact patterns. More recent research also examined the effect of socio-economic factors including education, income level and occupation on the contact behavior of individuals (Thomas et al., 2021; Reichmuth et al., 2023; Lucchini et al., 2024; Soussand et al., 2025). However, most of the studies considered the total number of contacts and limited interest went towards investigating the effects on the number of non-household contacts. Since NPIs mainly affect contacts outside the household (Dobreva et al., 2022; Phuong et al., 2025), the focus of this thesis will go towards non-household contacts. Previous research on non-household contact patterns during the COVID-19 pandemic includes Feehan and Mahmud (2021), Bridgen et al. (2022), Backer et al. (2023) and Walde et al. (2023).

The goal of this thesis is to investigate which factors influenced the presence and reported number of non-household contacts based on the CoMix study in Belgium between December '20 and March '22 by considering both the Generalised Additive Model for Location, Scale and Shape (GAMLSS) framework as well as zero-inflation and Hurdle models to account for the excess of zero reported non-household contacts. Since the age group that was the main contributor to transmission changed at different time periods (Angeli et al., 2025), the evolution of the number of non-household contacts was examined across age groups and time, together with the effect of different characteristics on the reported number of contacts made outside the household. Moreover, similar to Kretzschmar and Mikolajczyk (2009), the distribution of non-household contacts across different locations is investigated via agglomerative hierarchical clustering since these contact patterns may play an important role in how transmission in the population exactly takes place. First, more details about the data collection procedure and the methodology of the analyses performed in this thesis are given, whereafter the results of the analyses are presented. This is followed by a discussion of the results, where implications towards the transmission of COVID-19 in the populations will be made. This thesis will conclude by emphasising the societal relevance of this study and reflecting on its ethical aspects and stakeholder awareness.

2 Methods and materials

2.1 Data collection

The CoMix study is a longitudinal survey used to keep track of the public behavior during the COVID-19 pandemic. The study started in March 2020 with data collections in Belgium, the Netherlands and the United Kingdom (Verelst et al., 2021). Between December 2020 and October 2021, an additional 17 countries took part in the CoMix study (Wong et al., 2022). A map of the different participating European countries can be found in Verelst et al. (2021). In each study country, quota sampling was used to recruit a nationally representative sample based on, among others, age, gender and geographical region to reflect the distribution of the total population (Wong et al., 2023). The design of the CoMix study is based on the POLYMOD survey, which recorded the daily social contacts of participants in 8 European countries (Mossong et al., 2008). In the CoMix study, participants self-reported their social contacts made between 5 am on the day before filling in the questionnaire up to 5 am on the day of the survey, where a contact is defined as an in-person conversation of at least a few words or physical contact (Verelst et al., 2021).

During the first 8 waves of the CoMix survey in Belgium, only adults were included in the data collection. Children were omitted to make sure that ethical clearance was obtained as fast as possible (Loedy et al., 2023). From wave 9 onwards, children also took part in the study. In order to accomplish this, the design of the study was changed and the questionnaires for children were filled in by one of its parents. Since waves 9 to 11 can be seen as a transition period where questions changed, our attention will go towards the data collected between wave 12 and wave 43. Since participants dropped out during this period, the group was continuously supplemented with new people to make sure the sample size requirements were met during all waves of data collection (Loedy et al., 2023). Note that participants could voluntarily decide to join or leave the study at any time. Moreover, informed consent was collected and the data was pseudo-anonymised (Coletti et al., 2020).

Next to the number of contacts a participant made, also the place of the contact was recorded, whether or not the contacted person was part of the same household and if the participant was wearing a face mask during the contact. Information about the age and gender of the contacted person was also asked for, as well as whether the contact was made during a holiday (period) and during a weekday or weekend. With respect to the participant, the age and gender was collected, together with its income category, education of the main earner of the household, occupation (if the participant's age was eligible to work) and social group. Note that the education of the main earner in the household will be considered in the sequel instead of the education of the participant themselves since there were a lot of missing values for the education status of the participant and it can be assumed that the social behavior may be influenced by the views of the highest earner within the household, which in turn depend on that person's educational background. Furthermore, the size of the household and the area the participant is living in was recorded, as well as its vaccination status, symptomatic status and whether the participant had an elevated risk or not. Note that the CoMix study collected a rich amount of information, of which only a selection of components have been mentioned above. More details on the CoMix study, including the protocol, can be found in Jarvis et al. (2020), Coletti et al. (2020), Verelst et al. (2021), Wong et al. (2023) and Jarvis et al. (2024). Details on the variables considered in the analyses of this thesis can be found in Table 1.

Table 1: Description of variables used in the analyses.

Variable	Description	Values
Wave	Wave of participation	12, 13, . . . , 43
Wave count	Number of times participant already participated	1, 2, . . . , 7, 8+
Social group	Social group of the participant	Group 1&2, Group 3&4, Group 5&6, Group 7&8
Vaccination status	Whether or not the participant had at least one injection of the vaccine	Yes, No
Elevated risk	Chronic liver disease, neurological disease, diabetes, weakened immune system, asplenia or malfunctioning spleen, morbid obesity (BMI $\geq$ 40), pregnant women	Yes, No
Face mask	Whether or not the participant used a face mask during the reported contact	Yes, No
Symptomatic status	Whether or not the participant had fever or chills, cough, shortness of breath, extreme tiredness, muscle or body aches or headache, runny nose, or sore throat during the 7 days before participation (Jarvis et al., 2024)	Yes, No
Area	Region of residency of participant	Brussels Hoofdstedelijk gewest, Vlaams Gewest, Waals Gewest
Holiday	Belgium nationally recognised non-working day, when most business and institutions are closed (includes both school holidays as well as one-day national holidays)	Yes, No
Weekday/weekend	When contact was reported	Weekday, Weekend
Gender	Gender of the participant	Male, Female
Household size	Number of people who live at the same address and share the same kitchen with the participant	1, 2, 3, 4+
Age category	Age group participant was located in during its first participation	Children (age 0-17), Adults (age 18-65), Elderly (age 66-...)
Day number	Number of years the participation took place after the start of the study (22 December '20)	Real number
Contacts age	Age category of contact	Children (age 0-17), Adults (age 18-65), Elderly (age 66-...)
Education main earner	Highest education level of the main earner in the household	Low, Medium, High
Employment status	Current employment status of the participant	Employed, Not in labor force, Student
Income level	Income level of the participant	Very Low, Low, Medium, High, Very high
Occupation	Occupation of the participant	6 categories (see Table 6)

For every participant, the age at the first wave of participation was considered as the participant's age. The age category factor variable of the participant (see Table 1) was created based on the exact age of the participant (or age interval if the exact age was not recorded). Creation of the age category factor variable of the contacted person was solely based on the reported age interval of the contacted person. Since household sizes up to 12 were present in the dataset, a factor variable for the household size was made which merged all household sizes of at least four to the category 4+. Participants whose social group was 'not allocated' were removed from the dataset, as this group only accounted for 0.2% of the sample. In addition, some assumptions were made which limited the amount of missing data. First, it was assumed that as soon as the participant reported to be vaccinated, this person remained vaccinated for the other participation rounds as well (so the number of doses of vaccine the participant received was not taken into account). It was moreover assumed that children were not vaccinated as COVID-19 vaccination was not yet recommended for them. For adults and elderly, the most recent vaccination status reported by each participant, if available, was used to fill in missing data for this variable. This approach was also employed for elevated risk status and face mask usage. For missingness in these variables for children, we assumed that they generally did not have elevated risk factors and were not subject to mandatory face mask policies during the study period.

2.2 Methodology

2.2.1 Exploratory analysis

As exploratory analysis, summary statistics for the CoMix dataset were given, together with the average number of non-household contacts per age category and wave with corresponding 95% confidence intervals. The longitudinal trend of the average number of non-household contacts was compared with the stringency index as computed by Hale et al. (2021). This government response tracker summarised the stringency of government policies such as school closings and stay-at-home requirements during the COVID-19 pandemic on a score between 0 (no interventions) and 100 (strict interventions). Furthermore, histograms showing the proportion of survey waves in which participants reported zero non-household contacts were constructed and examined. These were created for the overall dataset, as well as stratified by each variable considered in this thesis, to explore patterns in the reporting of non-household contacts.

Since over the recent years, more interest is going towards investigating the relationship between socio-economic factors and social behavior (see, amongst others, Gimma et al., 2022; Tizzani & Gauvin, 2024; and Di Domenico et al., 2025), respondents' characteristics such as their income and education level, employment status and occupation category were considered as well in this thesis. The average number of non-household contacts per category of each SES-related variable was computed and accompanied with a 95%

CI. Finally, as the perception of the COVID-19 pandemic severity can change throughout the study period, the social behavior of participants with symptoms can be different as well. Therefore, the longitudinal trend of the average number of non-household contacts of participants with and without reporting symptoms was depicted with a 95% CI. Moreover, the average number of reported non-household contacts stratified by symptomatic status and other variables, such as vaccination status, was explored.

2.2.2 Number of non-household contacts via GAMLSS models

In order to investigate which factors influenced the number of non-household contacts based on the data from the CoMix study, the Generalised Additive Models for Location, Scale and Shape (GAMLSS) framework, introduced by Rigby and Stasinopoulos (2005), was considered. The models were fitted by using the `R` package `gamlss` (Rigby and Stasinopoulos, 2005). All analyses were performed using `R` Statistical Software (v4.4.1; R Core Team, 2023). More details on the implementation of GAMLSS in `R` with multiple data examples can be found in Stasinopoulos and Rigby (2007). GAMLSS extends the generalised linear mixed models (GLMM) and generalised additive mixed models (GAMM) to distributions of the response which are not part of the exponential family. Moreover, next to the mean, also other parameters of the distribution of the response variable such as the variance can be modeled via linear, non-linear and/or (smooth) non-parametric functions of the explanatory variables as well as random effects. Since the number of non-household contacts exhibited overdispersion and there were a considerable number of responses with zero non-household contacts, the GAMLSS framework was considered to model these contacts. Note that the presence of overdispersion was assessed by comparing the deviances and AIC-values of the final NBI GAMLSS model with the Poisson GAMLSS model with exactly the same structure as the former model.

Fitting GAMLSS models to the whole dataset resulted into convergence issues. Therefore, subgroup analyses per age category (children, adults, elderly) were performed to investigate whether the factors that drove the number of non-household contacts were different across the three age categories. Since no children reported living alone, the reference household size category in the model for children was 2. Moreover, the household size categories 3 and 4+ were merged together for the elderly age group, since the number of elderly participants reporting to be living with more than 2 people in their household was very small. Note that, since it was assumed that no children had elevated risks or were vaccinated and the symptomatic status of children was never reported, these variables were not considered in the model for children. Most of the adults and elderly participants reported their gender, such that this variable could be included in their corresponding models for the number of non-household contacts. However, gender was not considered in the model of the children because none of the children reported their gender. Gimma et al. (2022) found an influence of employment status on the number of contacts. Therefore, the model of the adults also accounted for both the employment status and education of the highest earner (see Table 1 and Table 6 for more details about the categories of these variables). Note that the education of the participant itself was often not reported such that the education of the highest earner was considered instead. The income level and occupation of the participant were not included in the model of the adults, as 21.0% and 27.4% of these variables were missing values, respectively. These socio-economic factors were reconsidered in the clustering analysis of contact patterns based on the location of the contact.

Following the recommendations given by Stasinopoulos, Rigby and Akantziliotou (2008), the model building started by first considering a simple model with only main effects for the mean parameter. Thereafter, the interaction terms which decreased the AIC-value of the model the most were added one-by-one until no significant decrease of AIC-value was occuring anymore. Since numerous pairwise interaction effects could be considered, only meaningful pairwise interaction effects were considered to be added to the model for the mean. Once a final model for the mean was fitted, attention went towards building a model for the variance parameter where a similar model building procedure was performed as for the mean parameter.

In this thesis, two GAMLSS distributions were considered to model the number of non-household contacts. Negative binomial and generalised Poisson regression models were employed due to the overdispersion present in the data and the considerable responses with zero non-household contacts. The corresponding probability functions are given below. All significance results were stated at a 5% level of significance and were accompanied with their corresponding 95% confidence intervals.

Negative Binomial distribution (NBI)
The probability function of the negative binomial distribution is given by

$$p(y|\mu,\sigma) = \frac{\Gamma(y+\frac{1}{\sigma})}{\Gamma(\frac{1}{\sigma})\Gamma(y+1)}\left(\frac{\sigma\mu}{1+\sigma\mu}\right)^{y}\left(\frac{1}{1+\sigma\mu}\right)^{1/\sigma} \tag{1}$$

for $y = 0, 1, 2, \ldots$ the number of non-household contacts, $\mu > 0$ and $\sigma > 0$ representing the dispersion parameter. This formulation of the probability distribution was introduced by Anscombe (1950) where $y = r, \sigma = 1/k$ and $\mu = m$. With this parametrisation, $\mathrm{E}(Y) = \mu$ and $\mathrm{Var}(Y) = \mu(1 + \sigma\mu)$.

Next to parametric effects, the mean and variance structure of the final age-group specific models contained a smoothing penalised varying coefficient introduced by Hastie and Tibshirani (1993) on the number of years the participation took place after the start of the study considered in this thesis (22 December '20) to make sure that its effect could vary depending on the combination of vaccination and symptomatic status, together with a random intercept for each participant. Cubic splines for the day number were considered instead of employing a penalised varying coefficient term, but the latter additive effect was superior in the models for adults and elderly based on its lower corresponding AIC-value. In the model for children, a smoothing penalised varying coefficient term for the day number was included which did not change smoothly according to another variable instead since both the vaccination status and symptomatic status of children were not included in the model.

Generalised Poisson (GPO)

Since, next to the presence of overdispersion in the data, there is also an excess of zeros, the Generalised Poisson GAMLSS model (GPO GAMLSS) was considered as well. Based on real time and simulated data, Yadav et al. (2021) found that the latter model consistently fitted overdispersed data with an excess of zeros better compared to negative binomial or zero-inflation models.

The probability function of the Generalised Poisson distribution is given by

$$p(y|\mu, \sigma) = \left(\frac{\mu}{1+\sigma\mu}\right)^y \frac{(1+\sigma y)^{y-1}}{y!} \exp\left(-\mu\right)$$

for $y = 0, 1, 2, \ldots$ the number of non-household contacts and $\mu > 0$, see Rigby et al. (2019). Note that the dispersion parameter σ is not restricted to be positive. A negative value of σ can be used to account for underdispersion, which rarely occurs in practice. With this parametrisation, $\mathrm{E}(Y) = \mu$ and $\mathrm{Var}(Y) = \mu(1 + \sigma\mu)^2$.

For all three age categories, the same variables and interaction effects as for the NBI GAMLSS models were considered in the model building process. Note that in the models for the variance parameter of all three age categories, no random effect for participant was included due to convergence issues.

As model diagnostic tool, randomised residuals of the NBI and GPO GAMLSS models were employed. Via the `plot.gamlss` function in `R`, the normalised randomised quantile residuals were checked by plotting the residuals against the fitted values and the index, considering the kernel density estimate of the residuals and its normal QQ-plot (Stasinopoulos et al., 2008). Alongside examining the diagnostic plots, the mean, variance, coefficient of skewness and coefficient of kurtosis of the quantile residuals were studied as well. If the GAMLSS model behaves well, i.e. the randomised quantile residuals are approximately normally distributed, the four coefficients mentioned above should be close to zero, one, zero and three, respectively. Since a discrete distribution family (negative binomial or generalised Poisson) for the response variable was considered to model the data, the quantile residuals were

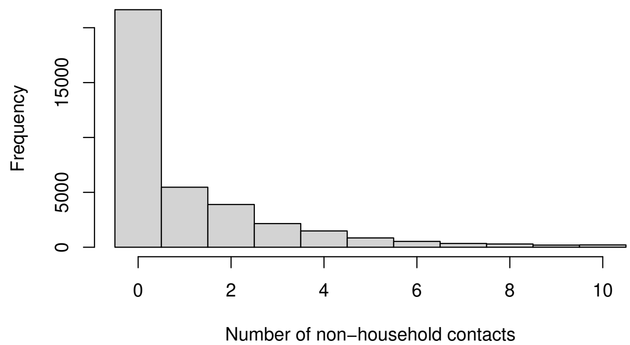


Figure 1: Histogram of number of non-household contacts. Note that this histogram zooms in to the range of contacts between 0 and 10 to highlight the large number of zero non-household contacts. However, up to 744 non-household contacts in one wave have been reported by participants, although most reported number of non-household contacts were between 0 and 10.

randomised. Therefore, based on recommendations from Stasinopoulos et al. (2008), the function `rqres.plot` was also used to create 40 realisations of the normalised randomised quantile residuals from the fitted GAMLSS model and consequently construct a QQ-plot of their median. Although not considered in this thesis, also other types of model diagnostic tools could be considered. One example is the worm plot of the residuals, introduced by van Buuren and Fredriks (2001) from which specific regions of a (dominant) explanatory variable where the model does not fit the data adequately could be identified. Moreover, the Q-statistics to test normality of the residuals within a region of an explanatory variable could be employed (see Royston and Wright (2000) for more details).

2.2.3 Non-household contacts via zero-inflation and Hurdle models

Next to the GAMLSS framework, both zero-inflation and Hurdle models were considered as well to investigate which factors influenced the presence and number of non-household contacts in our dataset. The histogram of the number of non-household contacts depicted in Figure 1 shows the large number of participations with zero reported non-household contacts. Moreover, as noted by Quilty et al. (2024), the overdispersion of contact rates causes an increase in the dispersion of the reproduction number. This motivates the choice of employing these models.

Both zero-inflation Poisson GAMLSS models (ZIP GAMLSS) and zero-inflation negative binomial counterparts (ZINBI GAMLSS) were considered in this thesis. Suppose that the number of non-household contacts (Y) is 0 with probability σ and follows a Poisson distribution with parameter μ ($Y \sim Po(\mu)$) with probability $(1 - \sigma)$. Then Y has a zero-inflation Poisson distribution with probability distribution given by, adapted from Lambert (1992) where $y = k, \sigma = p$ and $\mu = \lambda$,

$$p(y|\mu, \sigma) = \begin{cases} \sigma + (1-\sigma)e^{-\mu} & \text{if } y = 0 \\ (1-\sigma)\frac{\mu^y}{y!}e^{-\mu} & \text{if } y = 1, 2, 3, \ldots, \end{cases}$$

see also Ridout et al. (1998), whereas Greene (1994) introduced the zero-inflation negative binomial distribution where

$$p(y|\mu, \sigma, \nu) = \begin{cases} \nu + (1-\nu)p(y|\mu, \sigma) & \text{if } y = 0 \\ (1-\nu)p(y|\mu, \sigma) & \text{if } y = 1, 2, 3, \ldots \end{cases} \tag{2}$$

with $p(y|\mu, \sigma)$ given by (1) in order to account for overdispersion as well. Zero-inflation generalised Poisson models could also be defined, see Gupta et al. (1996), although these were not considered in this thesis. The zero-inflation models can be extended to also include covariates. Thomas et al. (2018) compared the performance of the different GAMLSS models for a dataset with an excess of zeros.

In addition to zero-inflation models, Hurdle models were considered to account for both overdispersion and the excess of zeros. These models are defined in two steps. In a first step, the probability of having at least 1 non-household contact is modeled via logistic regression. The second step only considers the participations with a non-zero number of reported non-household contacts and models the number of non-household contacts. Hurdle models were introduced by Cragg (1971) to analyse the demand for durable goods and were extended to count data with the truncated negative binomial distribution for the second step by Welsh et al. (1996). As described in Feng (2021), the structure of a Hurdle model is given by

$$P(Y_i = y_i) = \begin{cases} p_i & y_i = 0 \\ (1-p_i)\dfrac{p(y_i, \mu_i)}{1 - p(y_i = 0, \mu_i)} & y_i > 0, \end{cases} \tag{3}$$

where Y_i denotes the number of non-household contacts for participant-wave combination i and $p(y_i, \mu_i)$ is a probability mass function for a count distribution. Similarly as described earlier for zero-inflation models, all Hurdle models can be generalised to include covariates as well via $\text{logit}(p_i) = \boldsymbol{x_i}^\top \boldsymbol{\alpha}$ and $\log(\mu_i) = \boldsymbol{z_i^T} \boldsymbol{\beta}$ with covariate vectors $\boldsymbol{x_i}$ and $\boldsymbol{z_i}$ and vectors of regression coefficients $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. More details about both steps of the Hurdle models considered here will be discussed below.

The outcome variable of the first step of the Hurdle model was a binary variable (0/1) which attained the value 1 if a participant reported at least 1 non-household contact in that wave and 0 otherwise. A generalised linear mixed model (GLMM) was employed per age category with the function `glmer` in the `R` library `lme4`. The binomial distribution with logit link function was considered. For every age category, the initial model was a model with all main effects as in the GAMLSS modeling framework, together with a random intercept for the participant. In order to check whether there was overdispersion present in the data, the function `dispersion_glmer` in the library `blmeco` was used. Since, for all three age categories, the dispersion factor was between 0.93 and 0.98, the binomial distribution was continued to being used throughout the model building process. This process consisted of including interaction effects that significantly improved the model fit the most assessed by the Likelihood Ratio Test (LRT) until no significant improvements were occurring anymore. For feasibility, only meaningful pairwise interaction effects were considered to be added to the model. Thereafter, main effects that did not contribute significantly to the model (i.e., whose removal did not result in a significant LRT) were excluded in order to simplify the model. In the sequel, we will refer to this first step as the Hurdle 1 model.

As a second step, given that non-household contacts were reported, the number of non-household contacts was modeled via a zero-truncated probability distribution. Since, for all three age categories, there was clear overdispersion present in the data, the truncated negative binomial distribution was considered via the function `glmmTMB` in the package `glmmTMB` instead of the Poisson distribution. The Hurdle negative-binomial model is then (via (3)) given by

$$P(Y_i = y_i) = \begin{cases} p_i & y_i = 0 \\ \dfrac{1-p_i}{1-\left(\frac{1}{1+\sigma\mu_i}\right)^{1/\sigma}} \dfrac{\Gamma\left(y_i + \frac{1}{\sigma}\right)}{\Gamma\left(\frac{1}{\sigma}\right)\Gamma(y_i+1)} \left(\dfrac{\sigma\mu_i}{1+\sigma\mu_i}\right)^{y_i} \left(\dfrac{1}{1+\sigma\mu_i}\right)^{1/\sigma} & y_i > 0 \end{cases} \tag{4}$$

with dispersion parameter σ. The same variables and model building strategy as in the first step of the Hurdle model were considered. This model will be referred to as the Hurdle 2 model.

In order to assess the fit of the Hurdle models, the `DHARMa` package in `R` (Hartig, 2017) was used to obtain simulation-based residual diagnostics for hierarchical regression models. For every age category, standardised quantile residuals based on the final model were simulated by first simulating new response data from the fitted model for each observation and thereafter calculate the empirical cumulative distribution function for the simulated observations. The residual is defined as the value of the empirical distribution function at the value of the observed data and will be a number between 0 and 1 representing the proportion of simulated data that is lower than the observed value. Based on the residuals, two plots were created. On the one hand, a QQ-plot of the residuals was made to detect overall deviations from the expected distribution with tests for correct distribution (Kolmogorov-Smirnov test), dispersion and the presence of more simulation outliers than expected. On the other hand, a plot of the residuals against the predicted value was produced where simulation outliers were highlighted. Note that the probability of an outlier depends on the number of simulations. For the Hurdle models in this thesis, 1 000 simulations were considered to obtain stable results. Furthermore, a simulation-based dispersion test was performed.

By comparing (2) with (3) and (4), it is clear that whereas Hurdle models assume that there is only one process to obtain zero non-household contacts, zero-inflation models include two different processes that can give rise to a zero count. These two processes include (1) participants who never reported non-household contacts and (2) participants who could have reported non-household contacts but did not always do. In Hurdle models, only the first process could produce zero counts since a zero-truncated probability distribution function was used when the ”hurdle” was crossed. See Feng (2021) for more details about the differences in handling zero-inflation and the generating processes for zeros.

2.2.4 Clustering of contact profiles

Next to modeling the presence and number of non-household contacts, contact patterns of the participants across different locations were also investigated in order to explain some observations from the regression analyses. By analysing the reported number of non-household contacts in six different locations (home, work, school, leisure, transport, other), contact profiles were defined for the whole dataset and per age category. This was done based on an agglomerative hierarchical clustering algorithm in order to group the participants into clusters where those within each cluster had more similar contact patterns than participants assigned to other clusters. The agglomerative hierarchical clustering algorithm is a bottom-up algorithm where each participant is initially considered as a single-element cluster. At every step of the algorithm, the two clusters that are the most similar are combined into a bigger cluster. This process continues until there is only one large cluster. Sometimes, contacts with a person were recorded at multiple locations. Similar to the approach of Kretzschmar and Mikolajczyk (2009), the contact was only counted once with the hierarchy home $>$ work $>$ school $>$ leisure $>$ other place $>$ transport.

First, the average number of non-household contacts per location over all waves was calculated for all participants. These six count variables were thereafter scaled to make sure that clustering was not driven by locations with larger counts. Afterwards, the Euclidean distance between every pair of participants was calculated via

$$d(x,y) = \sqrt{\sum_{i=1}^{6}(x_i - y_i)^2}$$

where x and y are two vectors of length 6 representing the non-household average number of contacts at all six locations for two participants. This distance matrix was employed to measure how similar two clusters were by performing the agglomerative hierarchical clustering analysis using the function `dist` from the `R` package `proxy`. As agglomeration method, Ward's D2 measure of dissimilarity was considered. This measure was introduced by Ward (1963) and minimises the total within-cluster variance. Therefore, at each step of the algorithm, the two clusters with minimum between-cluster distance were merged. This measure of dissimilarity was employed via the function `hclust` from the `R` package `stats` since the option "ward.D2" implements the clustering criterion introduced in Ward (1963) (Murtagh and Legendre, 2014).

In order to choose a number of clusters, three methods were considered. First, the dendrogram was studied as an exploratory tool to visualise the hierarchical tree structure. Secondly, an elbow plot was used where the number of clusters can be chosen such that the total within sum of squares will not decrease a considerable amount when considering more clusters. Thirdly, the average silhouette plot was inspected to determine how well each participant lied within its cluster (Kaufman and Rousseeuw, 2005). The plot was made by computing the average silhouette for different cluster sizes, based on the function `silhouette` in the `R` package `cluster`. The optimal cluster size was the number for which the average silhouette was maximised. Once a number of clusters was chosen, the `R`-function `cutree` was employed to cluster the data into the desired number of groups. For each cluster, the average number of non-household contacts made per location was calculated and visualised. This hierarchical clustering algorithm was considered for the whole dataset and to investigate and compare the contact patterns per age category. Furthermore, the contact patterns in different income and occupation categories were examined via the same clustering methodology. Since parents filled in the questionnaires for their child, but reported their own income and occupation instead of that of their child, the children were not taken into consideration during the clustering analyses of the contact patterns per income and occupation category. Finally, also contact patterns of symptomatic and non-symptomatic participants were compared.

Alongside clustering of contact patterns in age categories and socio-economic groups, also participants with at least 1 non-household contact were clustered based on demographic characteristics to partly explain the observations made in the Hurdle 2 models. Since the demographic variables such as gender and household size of the participant were categorical variables instead of numeric, the Euclidean distance could not be used anymore. Alternatively, the Gower metric was considered to compute all pairwise 'distances' between two participants via the function `daisy` in the `R` package `cluster` (Kaufman and Rousseeuw, 2005). This similarity measure was introduced by Gower (1971) and can calculate the dissimilarity between two participants based on both numerical and categorical variables. The Gower metric first computes the distance for every variable separately and thereafter calculates the overall Gower distance as the average of the individual distances. For numerical variables, the range-normalised Manhattan distance was computed via

$$\mathrm{d}_j = \frac{|x_j - y_j|}{\mathrm{range}_j} \tag{5}$$

for the j-th variable of participants x and y where range_j is the difference between the maximum and minimum values of the numerical variable to make sure that the distance is between 0 and 1. For categorical variables, Dice distance was calculated as

$$\mathrm{d}_j = \begin{cases} 0 & \text{if } x_j = y_j \\ 1 & \text{if } x_j \neq y_j. \end{cases}$$

In order to reflect the ordering of the education of the main earner and household size, both categorical variables were recoded as ordinal variables with equidistant numerical values, assigning a difference of 1 between consecutive categories except for the household size for children since there were a considerable amount of participants in the 4+ group who lived in a household of size 5. Therefore, this variable with categories (2, 3, 4+) was encoded by (1, 2, 3.5) to account for these larger households. Based on the assigned numerical values for the ordinal variables, the distance between two participants for these variables was calculated via (5).

After computing the Gower dissimilarity matrix, hierarchical clustering was performed, although not with Ward's D2 agglomeration method. Since this method minimises the total within-cluster variance and is based on a (squared) Euclidean distance, it cannot be employed with the non-Euclidean Gower distance. The average agglomeration method does not assume Euclidean geometry and was therefore considered instead. The number of clusters was determined by considering both the dendrogram and the average silhouette plot. Thereafter, the proportion of each variable observed within each cluster was computed and visualised, together with the average number of non-household contacts per location in every cluster.

3 Results

3.1 Exploratory analysis

During the study period considered in this thesis, 4 208 individuals participated and reported 39 028 responses and 166 208 contacts, of which 118 506 were non-household contacts. The highest number of non-household contacts reported by an individual was 744.

Table 2 shows the distribution of the variables considered in the models per age category. More than half of the participants was an adult and over 50% of the respondents were living in Vlaams Gewest. Moreover, almost 3 out of 4 contacts took place during weekdays. 80% of participations were done by individuals living in a household smaller than 4.

Table 2: Summary of dataset by age group of participant. Percentages may not exactly add up to 100% due to rounding and a small proportion of missing values in some variables.

Age category of participant	Children	Adults	Elderly	**Total**
Total number of participants	1007 (23.93%)	2593 (61.62%)	608 (14.45%)	**4208 (100%)**
Total number of participations	7750 (19.86%)	23099 (59.19%)	8179 (20.96%)	**39028 (100%)**
Total number of non-household contacts	37409 (31.57%)	66466 (56.09%)	14631 (12.35%)	**118506 (100%)**
Contacts age				
Children	27706 (23.38%)	9556 (8.06%)	1947 (1.64%)	**39209 (33.09%)**
Adults	8414 (7.10%)	45174 (38.12%)	8710 (7.35%)	**62298 (52.57%)**
Elderly	1030 (0.87%)	10460 (8.83%)	3694 (3.12%)	**15184 (12.81%)**
Don't know/Prefer not to answer	259 (0.22%)	1276 (1.08%)	280 (0.24%)	**1815 (1.53%)**
Weekday/Weekend				
Weekday	5944 (15.23%)	17545 (44.95%)	5487 (14.06%)	**28976 (74.24%)**
Weekend	1806 (4.63%)	5554 (14.23%)	2692 (6.90%)	**10052 (25.76%)**
Area				
Brussels Hoofdstedelijk Gewest	750 (1.92%)	1968 (5.04%)	571 (1.46%)	**3289 (8.43%)**
Vlaams Gewest	4263 (10.92%)	13805 (35.37%)	5142 (13.18%)	**23210 (59.47%)**
Waals Gewest	2737 (7.01%)	7326 (18.77%)	2466 (6.32%)	**12529 (32.10%)**
Holiday				
Yes	2455 (6.29%)	7189 (18.42%)	2491 (6.38%)	**12135 (31.09%)**
No	5295 (13.57%)	15910 (40.77%)	5688 (14.57%)	**26893 (68.91%)**
Household size				
1	0 (0.00%)	6962 (17.84%)	2436 (6.24%)	**9398 (24.08%)**
2	716 (1.83%)	8831 (22.63%)	5348 (13.70%)	**14895 (38.17%)**
3	2627 (6.73%)	3958 (10.14%)	302 (0.77%)	**6887 (17.65%)**
4+	4407 (11.29%)	3348 (8.58%)	93 (0.24%)	**7848 (20.11%)**
Elevated risk				
Yes	0 (0.00%)	6327 (16.21%)	3767 (9.65%)	**10094 (25.86%)**
No	7750 (19.86%)	16713 (42.82%)	4411 (11.30%)	**28874 (73.98%)**
Face mask usage				
Yes	3035 (7.78%)	14921 (38.23%)	5255 (13.46%)	**23211 (59.47%)**
No	4715 (12.08%)	8178 (20.95%)	2924 (7.49%)	**15817 (40.53%)**
Symptomatic status				
Yes	0 (0.00%)	16623 (42.59%)	7106 (18.21%)	**23729 (60.80%)**
No	0 (0.00%)	6431 (16.48%)	1073 (2.75%)	**7504 (19.23%)**
NA	7750 (19.86%)	45 (0.12%)	0 (0.00%)	**7795 (19.97%)**
Vaccination status				
Yes	0 (0.00%)	14428 (36.97%)	6517 (16.70%)	**20945 (53.67%)**
No	7750 (19.86%)	8626 (22.10%)	1661 (4.26%)	**18037 (46.22%)**
Gender				
Female	0 (0.00%)	12451 (31.90%)	3065 (7.85%)	**15516 (39.76%)**
Male	0 (0.00%)	10601 (27.16%)	5114 (13.10%)	**15715 (40.27%)**
NA	7750 (19.86%)	47 (0.12%)	0 (0.00%)	**7797 (19.98%)**
Social group				
Group 1&2	2813 (7.21%)	7204 (18.46%)	937 (2.40%)	**10954 (28.07%)**
Group 3&4	2143 (5.49%)	5722 (14.66%)	3213 (8.23%)	**11078 (28.38%)**
Group 5&6	1509 (3.87%)	4444 (11.39%)	2276 (5.83%)	**8229 (21.08%)**
Group 7&8	1285 (3.29%)	5729 (14.68%)	1753 (4.49%)	**8767 (22.46%)**
Waves of participation				
1	1007 (2.58%)	2593 (6.64%)	608 (1.56%)	**4208 (10.78%)**
2	688 (1.76%)	1708 (4.38%)	456 (1.17%)	**2852 (7.31%)**
3	574 (1.47%)	1409 (3.61%)	411 (1.05%)	**2394 (6.13%)**
4	499 (1.28%)	1269 (3.25%)	401 (1.03%)	**2169 (5.56%)**
5	440 (1.13%)	1178 (3.02%)	392 (1.00%)	**2010 (5.15%)**
6	408 (1.05%)	1106 (2.83%)	384 (0.98%)	**1898 (4.86%)**
7	376 (0.96%)	1051 (2.69%)	372 (0.95%)	**1799 (4.61%)**
8+	3758 (9.63%)	12785 (32.76%)	5155 (13.21%)	**21698 (55.60%)**

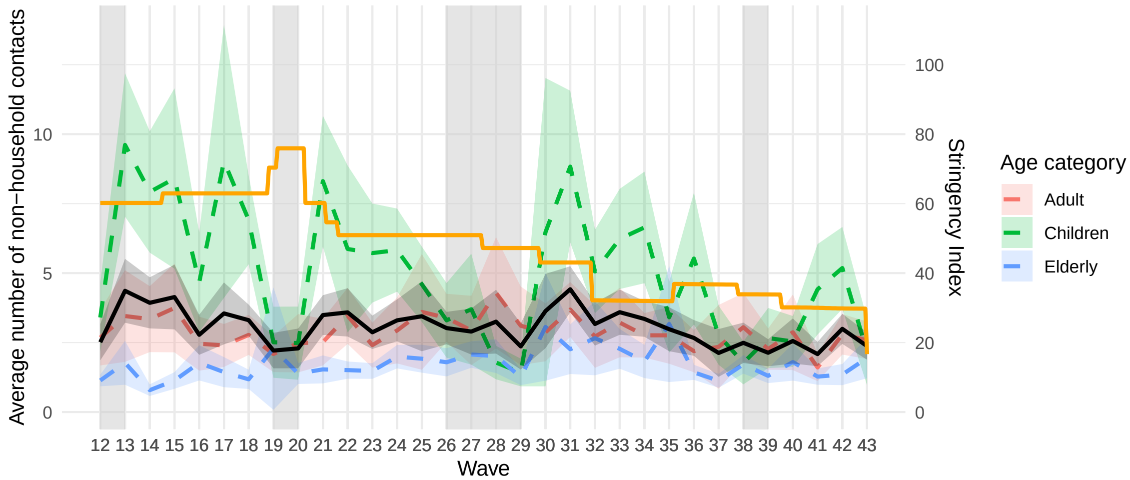


Figure 2: Average number of non-household contacts per age category and wave, together with the corresponding 95% CI. The black solid line represents the average number of non-household contacts per wave over all three age categories. The grey shaded regions represent the Christmas, Easter and summer periods. The stringency index at a specific moment is indicated in orange.

The average number of reported non-household contacts was 3.04 with the median being 0 due to the large number of zero reported non-household contacts (see also Figure 1). The number of reported non-household contacts decreased with the number of participations of the respondent with an average of 4.74 (CI [4.18; 5.30]) for first participations and 2.51 (CI [2.33; 2.69]) for individuals already participating for at least 8 times. As Figure 12 indicates that the latter group dominated at later periods of the study period, this will influence the average number of reported non-household contacts.

Figure 2 depicts the longitudinal trend of the average number of reported non-household contacts per age category together with the evolution of the stringency index. This index increased to almost 80 during the Easter period of 2021 and thereafter decreased again to ultimately reach a value of 30 at the end of the study period. However, the average number of reported non-household contacts was quite stable throughout the study period. The clear drop in average number of reported non-household contacts of children during the holiday periods could be attributed to the closure of schools during these periods. On average, children reported more contacts outside their household (4.83, CI [4.46; 5.20]) compared to adults (2.88, CI [2.68; 3.08]) and elderly (1.79, CI [1.63; 1.95]).

Based on Figure 3, a considerable amount of participants could be divided into two groups according to the proportion of waves in which they reported zero non-household contacts.

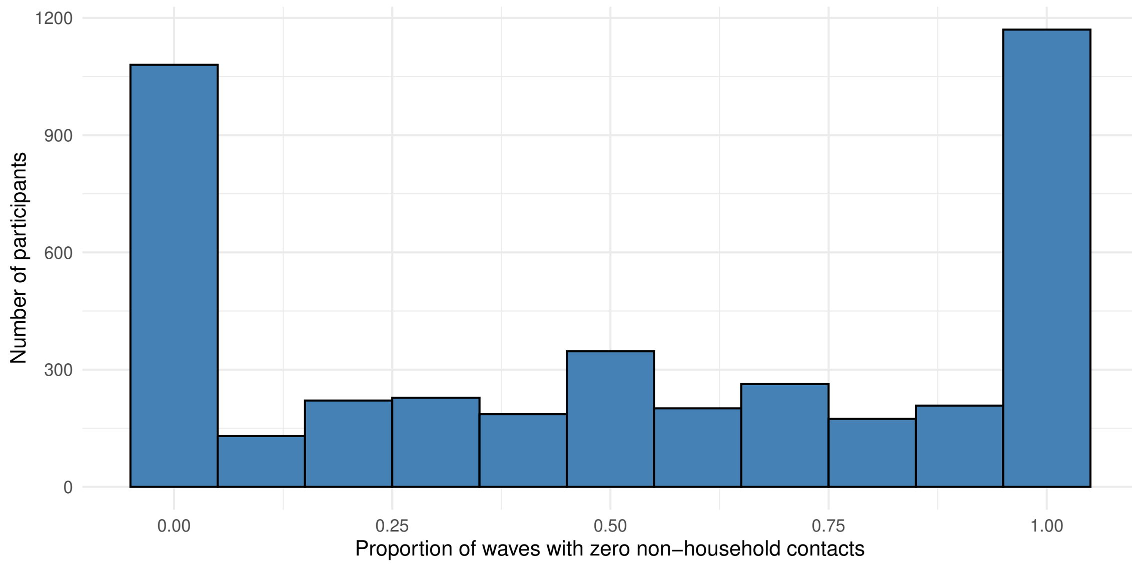


Figure 3: Histogram of the proportion of waves per participant where zero non-household contacts were reported.

One group almost never reported such contacts, while the other group almost always did. If non-household contacts were made, participants with over 90% of waves reporting zero non-household contacts had an average of 3.71 non-household contacts (median: 2), whereas participants with fewer than 10% of waves reporting zero non-household contacts reported on average 9.39 contacts outside their household (median: 3). A similar pattern was observed for the categories of every variable, although the proportion of participants reporting zero non-household contacts in more than 90% of the waves was considerably lower in Vlaams Gewest (23.3%) compared to Waals Gewest (38.7%) and Brussels Hoofdstedelijk Gewest (35.3%).

Table 3 indicates that participants working as service employees reported the highest average number of non-household contacts whereas respondents not in labor force had the lowest number. The differences in contact patterns between occupation categories was further explored in the clustering analysis. Moreover, there was an increasing trend in the average number of reported non-household contacts as the income level was higher. Finally, note that for both symptomatic and non-symptomatic participants, there was no clear increasing trend over time in the average number of reported non-household contacts. On average over all waves, non-symptomatic participants reported more contacts outside their household (3.31 vs 2.37). No deviations from this effect of symptomatic status were observed in the comparison of the average number of reported non-household contacts stratified by every other variable.

Table 3: Average number of reported non-household contacts (95% CI) per category of different socio-economic variables.

Variable	Mean (CI)	Variable	Mean (CI)
Occupation		**Income**	
Managers & Professionals	3.22 [2.85; 3.59]	Very low	1.72 [1.49; 1.95]
Office employees	3.10 [2.79; 3.41]	Low	2.45 [2.08; 2.82]
Service employees	6.13 [5.26; 7.00]	Middle	2.91 [2.64; 3.18]
Manual workers	3.28 [2.81; 3.75]	High	3.54 [3.23; 3.85]
Self-employed/Small business	3.05 [2.14; 3.96]	Very high	4.06 [3.62; 4.50]
Not in labor force	1.86 [1.72; 2.00]		
Employment status		**Education main earner**	
Not in labor force	1.57 [1.48; 1.66]	Low	2.14 [1.89; 2.39]
Employed	3.73 [3.41; 4.05]	Medium	3.35 [3.08; 3.62]
Student	3.02 [2.31; 3.73]	High	3.08 [2.89; 3.27]

3.2 Number of non-household contacts via GAMLSS models

Both NBI GAMLSS and GPO GAMLSS models were considered for every age category to investigate which factors influenced the number of non-household contacts. Table 4 presents the AIC values for the NBI GAMLSS and GPO GAMLSS models across all three age categories. The former models showed substantially lower AIC values compared to the GPO GAMLSS models in all age groups, suggesting a better fit when a NBI GAMLSS model was considered. However, as already touched upon in the previous section, this difference was largely attributable to the fact that in the GPO GAMLSS models it was not possible to include a random effect in the model for the variance parameter. This had a considerable impact on the overall model fit. If one only considered the models for the mean parameter without building a model for the variance parameter, the GPO distribution consistently provided a better fit across all three age categories compared to the NBI distribution.

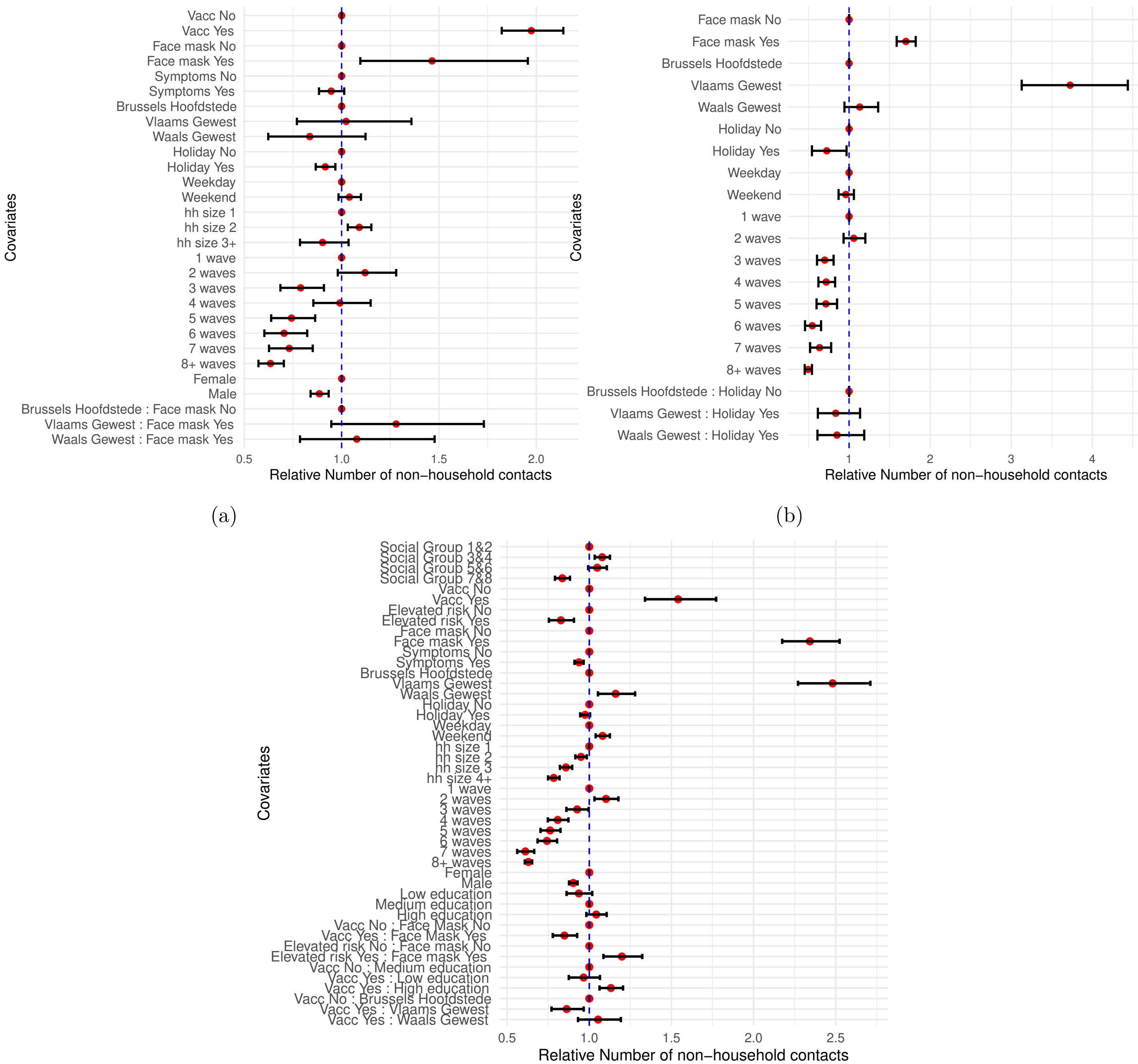
Table 4: AIC values for NBI GAMLSS and GPO GAMLSS models for all three age categories.

Age category	NBI	GPO
Children	25965.0	26871.4
Adults	63222.0	65952.4
Elderly	22814.9	23581.1

Figure 4 and Figure 5 show the results of, respectively, the NBI and GPO GAMLSS models for the average number of non-household contacts in the different age categories. A summary of the coefficients for the mean and variance structure of the NBI and GPO GAMLSS models can be found in Tables 7-18 in the Appendix. Based on the model diagnostic plots depicted in Figures 13-15, one can conclude that no serious violations to the model fit were present, although some outlying observations could be detected based on the residual plots from the GPO GAMLSS model for elderly. Moreover, Table 19 shows that the mean, variance, coefficient of skewness and coefficient of kurtosis of the median randomised quantile residuals were close to their expected values, suggesting approximate normality. These observations on the residual diagnostics therefore indicate that the models adequately captured the structure of the data with no major patterns suggesting some misspecifications.

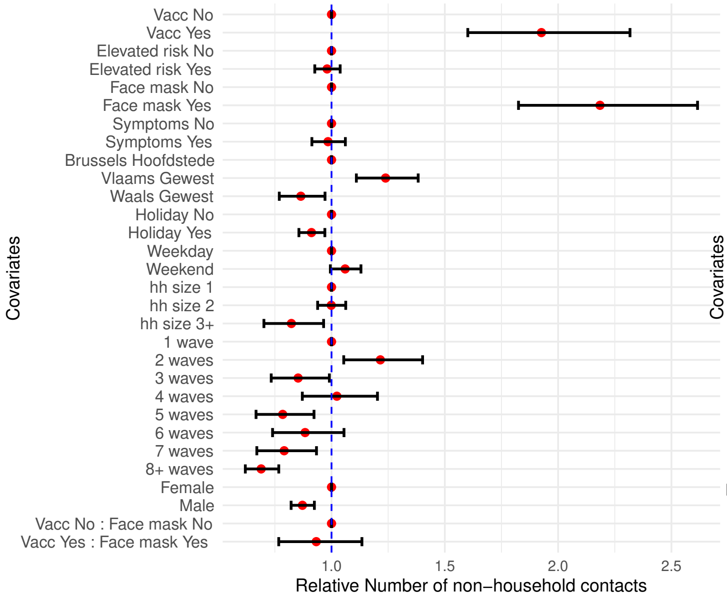
For children, wearing a face mask was positively associated with the average number of non-household contacts (NBI: 1.70 times higher, CI [1.59; 1.82], GPO: 1.90 times higher, CI [1.75; 2.07]). Compared to Brussels Hoofdstedelijk Gewest, participants from Vlaams Gewest reported 273% (CI [213%; 344%] and 369% (CI [286%; 470%]) more non-household contacts based on the NBI and GPO GAMLSS models, respectively. The increase in number of reported non-household contacts in Waals Gewest compared to Brussels Hoofdstedelijk Gewest was much smaller and only significant in the GPO model (1.38 times higher, CI [1.12; 1.68]). However, holiday was a moderating factor which reduced the positive effect of living in Waals Gewest or Vlaams Gewest on the number of non-household contacts. Based on the NBI GAMLSS model, one can conclude that the number of non-household contacts was smaller during holidays (27.5% lower, CI [3.10%; 45.8%]). As opposed to the NBI model, the GPO GAMLSS model indicated that compared to participants who lived together with 1 person, participants living in a household of size 4 or more reported on average 20.7% (CI [9.35%; 30.7%]) fewer non-household contacts.

Based on the results of the NBI and GPO GAMLSS models for elderly, one can see that vaccinated participants reported on average more non-household contacts compared to non-vaccinated participants (NBI: 1.97 times higher, CI [1.82; 2.14], GPO: 1.92 times higher, CI [1.60; 2.32]). Furthermore, wearing a face mask was positively associated with the average number of reported non-household contacts (NBI: 1.46 times higher, CI [1.10; 1.96], GPO: 2.18 times higher, CI [1.83; 2.62]). During holidays, participants reported fewer non-household contacts (NBI: 91.6%, CI [86.7%; 96.8%], GPO: 91.2%, CI [85.6%; 97.1%]). Males also had fewer non-household contacts compared to females (NBI: 88.6%, CI [84.0%; 93.3%], GPO: 87.2%, CI [82.2%; 92.4%]). However, also some differences in effects between both models were observed. First, only the GPO GAMLSS model found a positive effect of living in Vlaams Gewest compared to Brussels Hoofdstedelijk Gewest (1.24, CI[1.11; 1.38]). Secondly, the household size also had a different influence on the average number of reported non-household contacts. Compared to participants who lived alone, the NBI model found that participants living in a household of size 2 reported on average 1.09 times more non-household contacts (CI [1.03; 1.15]) which was non-significant in the GPO model. On

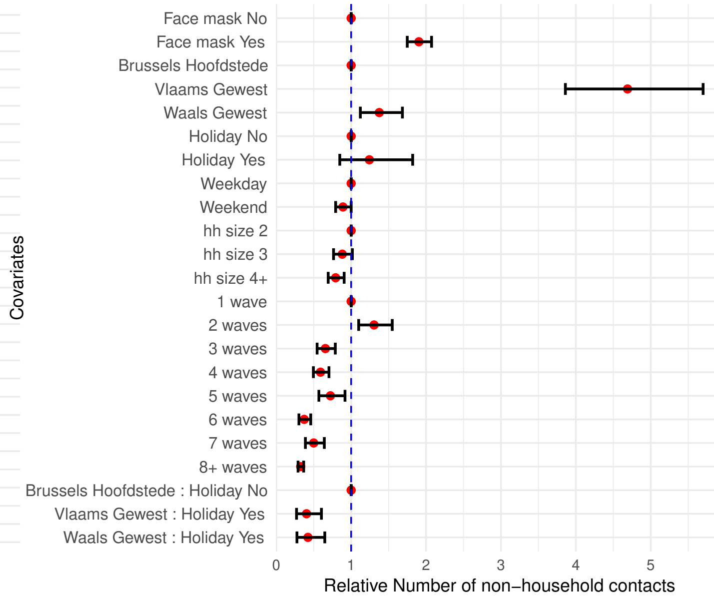


(a) (b) (c)

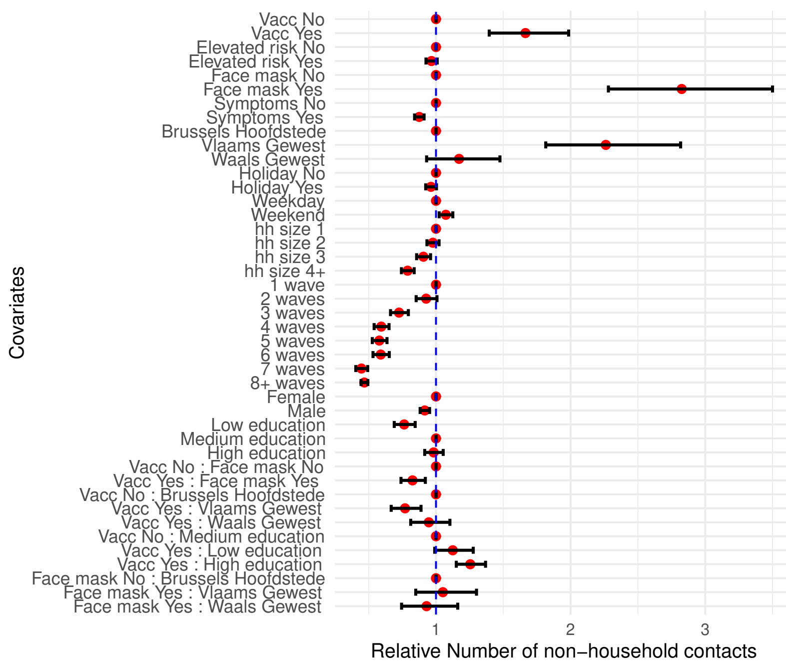
Figure 4: Relative number of non-household contacts and 95% CI based on the NBI GAMLSS model for (a) elderly, (b) children and (c) adults.



(a)



(b)



(c)

Figure 5: Relative number of non-household contacts and 95% CI based on the GPO GAMLSS model for (a) elderly, (b) children and (c) adults.

the other hand, the latter model found a negative effect of living in a household of size 3 on the average number of reported non-household contacts (0.82, CI [0.70; 0.97]).

Similar to the models for the two other age categories, wearing a face mask was positively associated with the average number of reported non-household contacts for adults as participants reported 134% (CI [117%; 152%]) or 183% (CI [128%; 250%]) more contacts outside the household based on the NBI and GPO model, respectively. Vaccinated adult participants reported more non-household contacts compared to non-vaccinated participants (NBI: 1.54 times higher, CI [1.34; 1.77], GPO: 1.66 times higher, CI [1.40; 1.99]). However, vaccinated participants who also wear face masks had fewer contacts as expected. Therefore, being vaccinated acted as a moderating factor which reduced the positive effect of mask-wearing on the number of non-household contacts. Also note that participants experiencing symptoms reported fewer contacts (NBI: 93.8%, CI [91.0%; 96.6%], GPO: 87.5%, CI [84.1%; 91.1%]). Compared to Brussels Hoofdstedelijk gewest, participants from Vlaams Gewest reported 148% more non-household contacts (CI [127%; 171%]) based on the NBI GAMLSS model (126%, CI [81.5%; 182%] for the GPO model). This positive effect of living in Vlaams Gewest was smaller for vaccinated participants. The increase in number of reported non-household contacts was much smaller in Waals Gewest compared to Brussels and only significant in the NBI GAMLSS model (1.16 times higher, CI [1.05; 1.28]). Weekend periods had a positive effect on the number of non-household contacts (NBI: increase of 8.11%, CI [3.87%; 12.5%], GPO: increase of 7.31%, CI [2.37%; 12.5%]), whereas males reported fewer non-household contacts (NBI: 90.3%, CI [87.7%; 92.9%], GPO: 91.7%, CI [88.2%; 95.2%]). The household size also had an influence on the average number of reported non-household contacts of adults. The larger the household size, the smaller the average number of reported non-household contacts.

Finally, note that the number of reported non-household contacts showed a significant downward trend for all three age categories as individuals participated in more waves. This can be related to the under-reporting due to fatigue, which was also observed by Loedy et al. (2023).

3.3 Non-household contacts via zero-inflation and Hurdle models

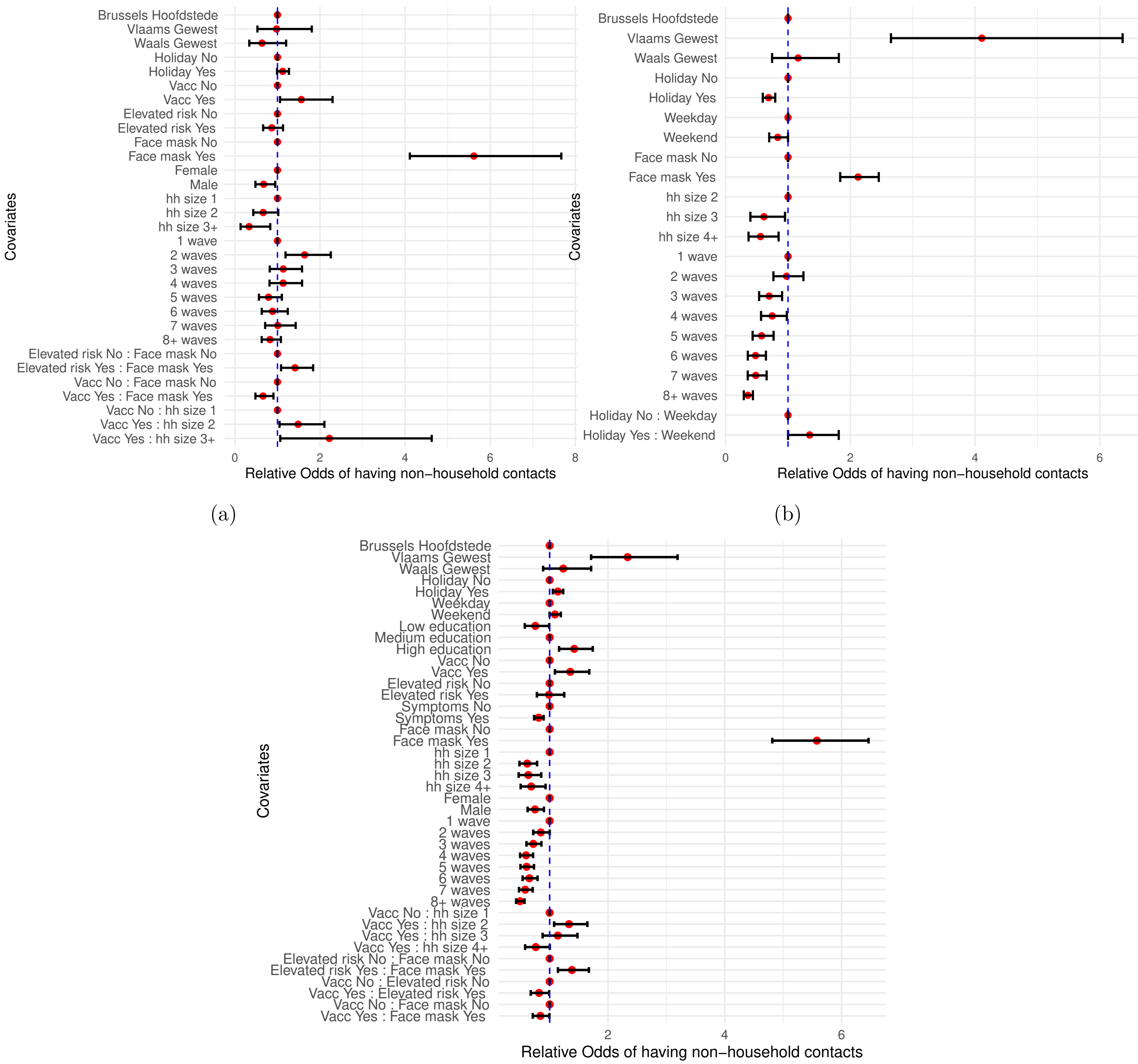
Due to the presence of an excess of zeros, zero-inflation and Hurdle models were considered as well. First note that both zero-inflation Poisson and zero-inflation negative binomial GAMLSS models did not converge, even if only the mean parameter was modeled. Since zero-inflation models require the joint estimation of the count part and the zero-inflation part, the additional layer of complexity by including a random effect for participant resulted into convergence issues. Therefore, only Hurdle models are considered in the sequel.

The summary output of the Hurdle 1 models for elderly, children and adults can be found in Tables 20-22 respectively and are visualised in Figure 6. The DHARMa residual plots for model diagnostics are shown in Figure 16. Although the Kolmogorov-Smirnov and dispersion test in the model for adults were both highly significant, there was not necessarily a problem with the model fit. Given the large sample size, even minor deviations from the model assumptions will be detected by the diagnostic tests, even when the model provides an adequate fit in practice. A formal DHARMa dispersion test revealed that the ratio of the observed standard deviation of the residuals to the expected standard deviation under the fitted model was 0.97 and significantly different from 1 ($p < 0.001$). However, this was a practically negligible deviation from the expected residual variance. For the Hurdle 1 models for elderly and children, the dispersion test was non-significant.

For elderly, the odds of having at least 1 non-household contact were 5.62 (CI [4.11; 7.67]) times higher for participants who were wearing a face mask during their reported contacts and this effect was even higher if the person also had an elevated risk. Vaccinated individuals had 1.56 times higher odds (CI [1.06; 2.29]), although for mask-wearing vaccinated respondents this positive effect of being vaccinated was almost completely diminished. Males had 32% lower odds (OR = 0.68, CI [0.48; 0.95]) of having at least 1 non-household contact compared to females. Also note that living in a larger household decreased the log-odds, but this negative effect was less pronounced for vaccinated participants.

Based on the Hurdle 1 model for children (see Table 21), one can also notice the positive effect of mask wearing (OR = 2.12, CI [1.84; 2.46]) and the negative effect of living in a larger household on the odds of having at least 1 non-household contact. Compared to Brussels Hoofdstedelijk Gewest, the odds of having non-household contacts were 4.11 times higher in Vlaams Gewest (CI [2.65; 6.37]). The odds of having non-household contacts were 31% lower during holidays (OR = 0.69; CI [0.60; 0.80]), although this effect was less negative if the holiday took place during the weekend. Finally, also note that the more waves a respondent already participated, the lower the odds of reporting at least 1 non-household contact.

This fatigue effect was also visible in the Hurdle 1 model for adults (see Table 22). For adults, wearing a face mask was positively associated with the odds of having non-household contacts (OR = 5.58, CI [4.81; 6.46]) and this positive effect was even higher if the participant also had an elevated risk. Having symptoms resulted into 19% lower odds of having non-household contacts (OR = 0.81, CI [0.73; 0.90]) and the odds were also lower for males compared to females (OR = 0.75, CI [0.62; 0.90]). Being vaccinated resulted into 1.35 times higher odds of having at least 1 non-household contact (CI [1.09; 1.68]), although this positive effect was partly counteracted if the respondent was wearing a face mask or was at elevated risk. Living together with other people in the same household had a negative effect



(a) (b) (c)

Figure 6: Relative odds of having non-household contacts and 95% CI based on the Hurdle 1 model for (a) elderly, (b) children and (c) adults.

on the odds of having at least 1 non-household contact. For vaccinated participants, the effect of household size was more negative as the household size was larger. Compared to Brussels Hoofdstedelijk Gewest, the odds of having non-household contacts were 2.34 times higher in Vlaams Gewest (CI [1.71; 3.19]). During holidays, the odds were higher (OR = 1.14, CI [1.06; 1.23]). Finally, the education level of the main earner in the household also influenced the odds of having non-household contacts. Participants for which the highest earner of the household had a low education had 25% lower odds (CI [0.01; 0.43]) compared to the reference category (medium-level education), whereas respondents with a high-level educated main earner in their household had higher odds (OR = 1.42, CI [1.16; 1.74]).

Figure 7 depicts the effects of the factors considered on the relative number of non-household contacts based on the Hurdle 2 model. The specific coefficients can be found in Tables 23-25. The DHARMa residual plots and nonparametric dispersion tests employed for model diagnostics, shown in Figure 17, indicated that there was significant overdispersion present in the Hurdle 2 models for adults and elderly, whereas the children's model showed some underdispersion. Therefore, interpretations of the results should be done with care.

For elderly, being vaccinated was positively associated with the average number of non-household contacts (1.91 times higher, CI [1.60; 2.27]. During holidays and weekends, participants reported 1.13 (CI [1.02; 1.25]) and 1.17 (CI [1.06; 1.29]) times more non-household contacts. Furthermore, respondents from Vlaams Gewest wearing a face mask reported on average 1.95 times more non-household contacts compared to participants from this area of residency who did not wear face masks (CI [1.15; 3.31]). In contrast to elderly, the Hurdle 2 model for children indicated the negative effect of holiday on the average number of non-household contacts. During holidays, 37.9% less non-household contacts were made by children (CI [29.1%; 45.6%]). children from Vlaams Gewest reported 2.11 times more non-household contacts compared to respondents from Brussels Hoofdstedelijk Gewest (CI [1.37; 3.25]). Moreover, there was a positive effect of household size on the average number of non-household contacts and wearing a face mask also increased this average number of contacts by 22.7% (CI [5.08%; 43.2%]). Finally, note that there was a clear fatigue effect present on the average number of reported non-household contacts.

This fatigue effect was also visible in the Hurdle 2 model for adults. Wearing a face mask was positively associated with the average number of non-household contacts (1.54 times higher, CI [1.82; 2.20]), although this effect was partly counteracted if the participant was vaccinated. Moreover, a positive effect of living in a household of size 2 or 3 was also present in this model. Adults who were at elevated risk reported on average 13.2% less non-household contacts (CI [2.37%; 22.8%]). During weekends, 1.11 times more non-household contacts were made (CI [1.03; 1.19]). Compared to adults who were not in labor force, students and employed respondents reported on average approximately 70% more non-

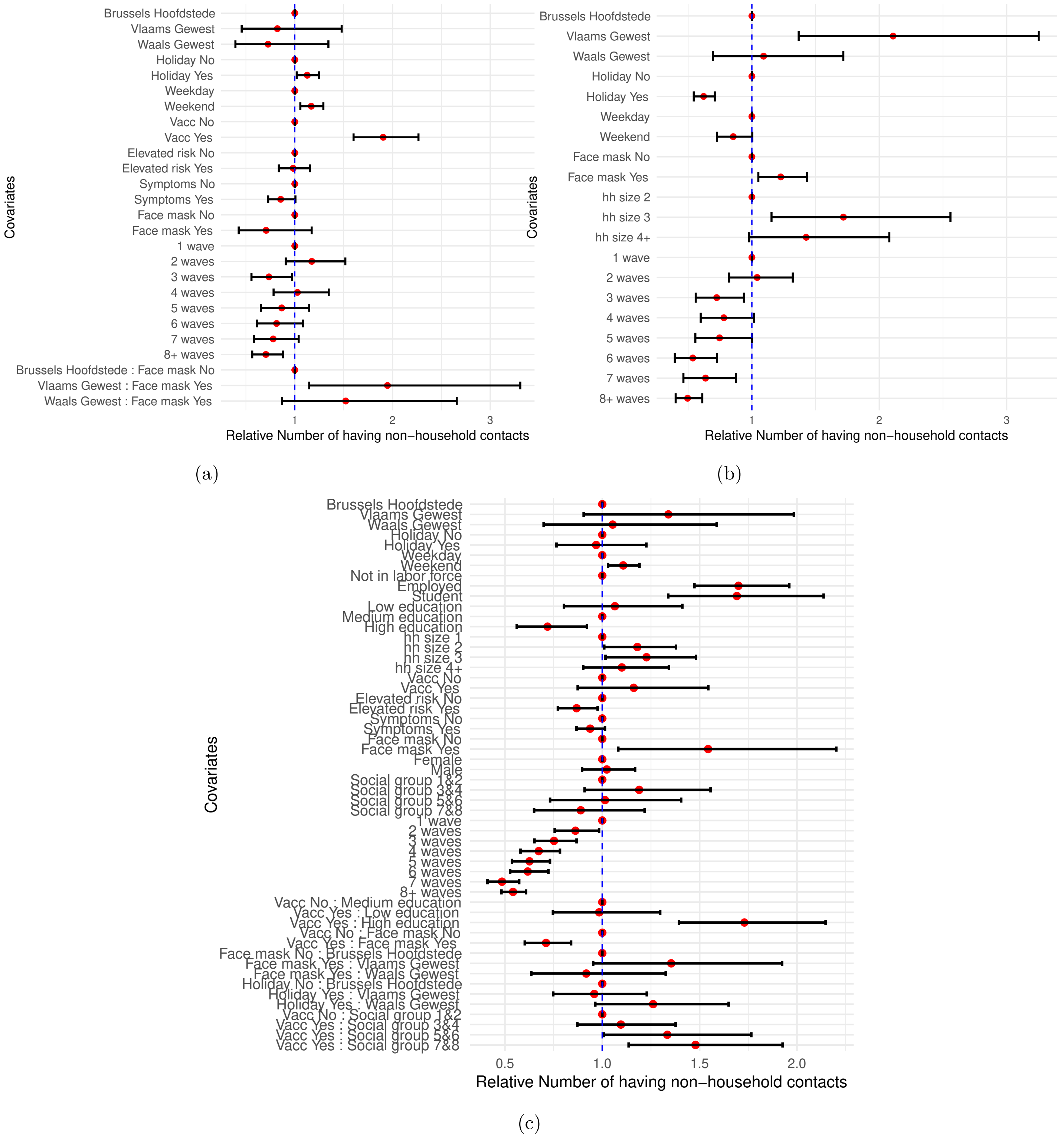


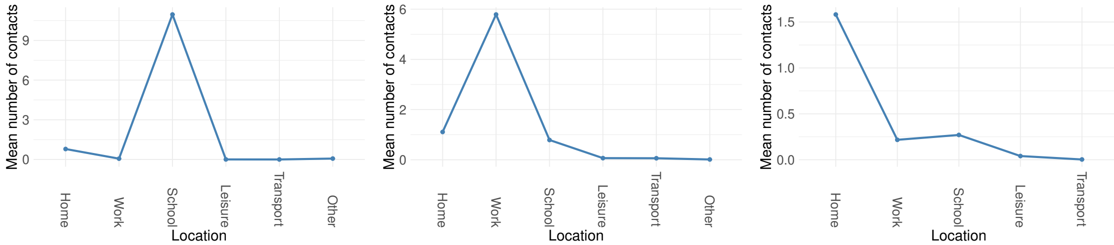
Figure 7: Relative number of having non-household contacts and 95% CI based on the Hurdle 2 model for (a) elderly, (b) children and (c) adults.

household contacts. Moreover, among unvaccinated adults, those with a high-educated main earner in their household had fewer non-household contacts compared to those with a medium-educated main earner (0.72, CI [0.56; 0.92]). However, this pattern reversed for vaccinated adults, where vaccination has led the participants with a high-educated main earner in their household to have on average more non-household contacts. Finally, note that vaccinated respondents from social groups 5&6 and 7&8 reported on average more non-household contacts than unvaccinated adults in these social groups.

3.4 Clustering of contact profiles

As described in the methodology section, the contact patterns of participants were investigated by employing a hierarchical clustering algorithm. First, the contact patterns per age category were considered, as clear differences in (factors that influenced) the number of reported non-household contacts per age category were found in the regression analyses. Since for every age category, the clustering algorithm revealed that there was one large cluster and multiple very small clusters with only a few participants who differed considerable with respect to the locations and number of non-household contacts, the clustering method was executed again on the participants from the large cluster. Nevertheless, this resulted once more in one large cluster and very small clusters. Therefore, the contact pattern from the original large cluster was considered for each age category as typical contact pattern for that category.

The resulting main contact patterns of the three categories, together with the average number of non-household contacts per location can be found in Figure 8. On the one hand, this figure reveals that elderly had, on average, fewer non-household contacts compared to adults and children, where children had even more non-household contacts compared to adults. This is in line with observations made in the exploratory analysis. On the other hand, the locations where most of the non-household contacts took place were different across the three age categories. As expected, children had most of their non-household contacts at school, but also had some non-household contacts at home. These could be contacts with friends or family who were not part of the household. Adults had most of their non-household contacts at work, but also had some contacts at home or at school. The latter could for example be contacts with teachers of their children or with other parents. Finally, it can be seen that most of the non-household contacts of elderly were at home, for example when family members came to visit these participants. Since some participants were still working at the age of 66 or older, some contacts were still made at work. Some elderly respondents also came into contact with other persons at school, for example when they picked up their grandchild. Note that, on average, not many contacts were made at transport, leisure or other locations. The clustering analysis revealed that there were participants with a considerable amount of non-household contacts at these locations, but



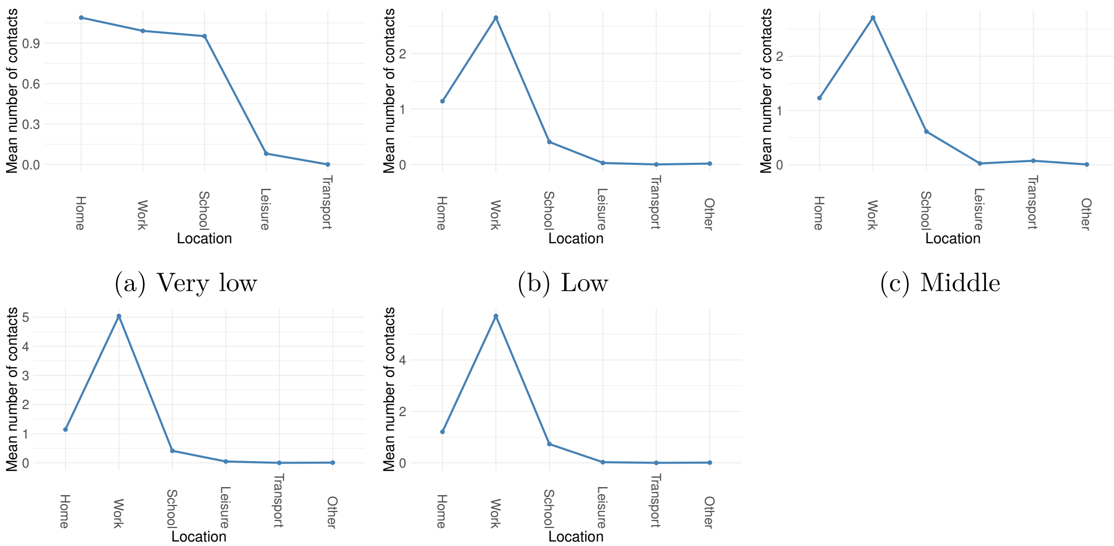
(a) (b) (c)

Figure 8: Main contact patterns with mean number of non-household contacts per location for (a) children, (b) adults and (c) elderly.

these were classified in a separate cluster to indicate that their contact patterns were clearly different from the majority of the participants from that age category.

Figure 9 and Figure 10 depict the results from the clustering analyses performed in the different income and occupation categories respectively. Note that a similar clustering algorithm was performed as with the age categories, where the small clusters were ignored and only the large cluster was considered. Figure 9 reveals that the average number of non-household contacts increased as the income level of the participant was higher. The average number of non-household contacts made at home was approximately 1 across all income categories, although it is noticeable that most contacts were work-related except at the lowest income category. A possible explanation is that respondents in the lowest income category were more likely to be outside the labor force. Therefore, they did not participate in the workplace and will lower the number of non-household contacts at work in the lowest income category. This pattern is also visible in Figure 10, where participants not in the labor force reported, on average, fewer non-household contacts overall, including fewer work-related contacts compared to the other occupation categories. The negative effect in the Hurdle 2 model for adults of not being in labor force can therefore partially be attributed to the smaller number of work-related non-household contacts compared to the other occupation categories. Furthermore, also self-employed participants reported on average a low number of non-household contacts which is in contrast to the high number of non-household contacts observed among service employees. This can be explained by the fact that these latter group of participants worked in the service sector, in roles such as teachers or nurses, where they had many contacts during their work. Finally, note that the contact patterns of symptomatic and non-symptomatic participants were very similar and only differed in the number of reported non-household contacts, where non-symptomatic respondents reported more work-related contacts outside their household.

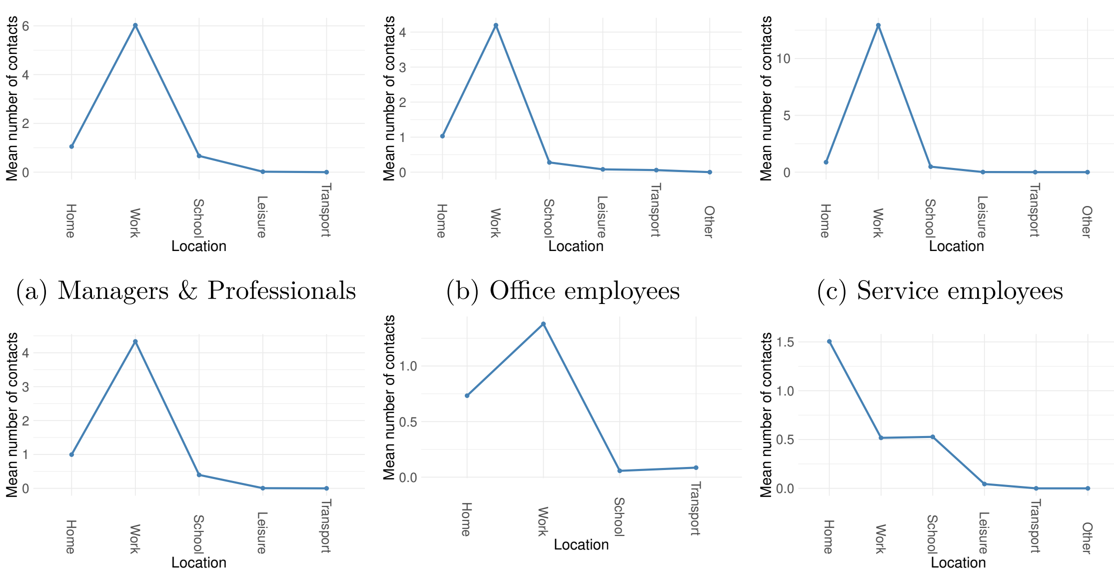
Figure 11 depicts the results of the clustering based on demographic characteristics. The



(a) Very low (b) Low (c) Middle

(d) High (e) Very high

Figure 9: Main contact patterns with mean number of non-household contacts per location for the different income categories when only adults and elderly are considered.



(a) Managers & Professionals (b) Office employees (c) Service employees

(d) Manual workers (e) Self-employed/Small business (f) Not in labor force

Figure 10: Main contact patterns with mean number of non-household contacts per location for the different occupation categories when only adults and elderly are considered.

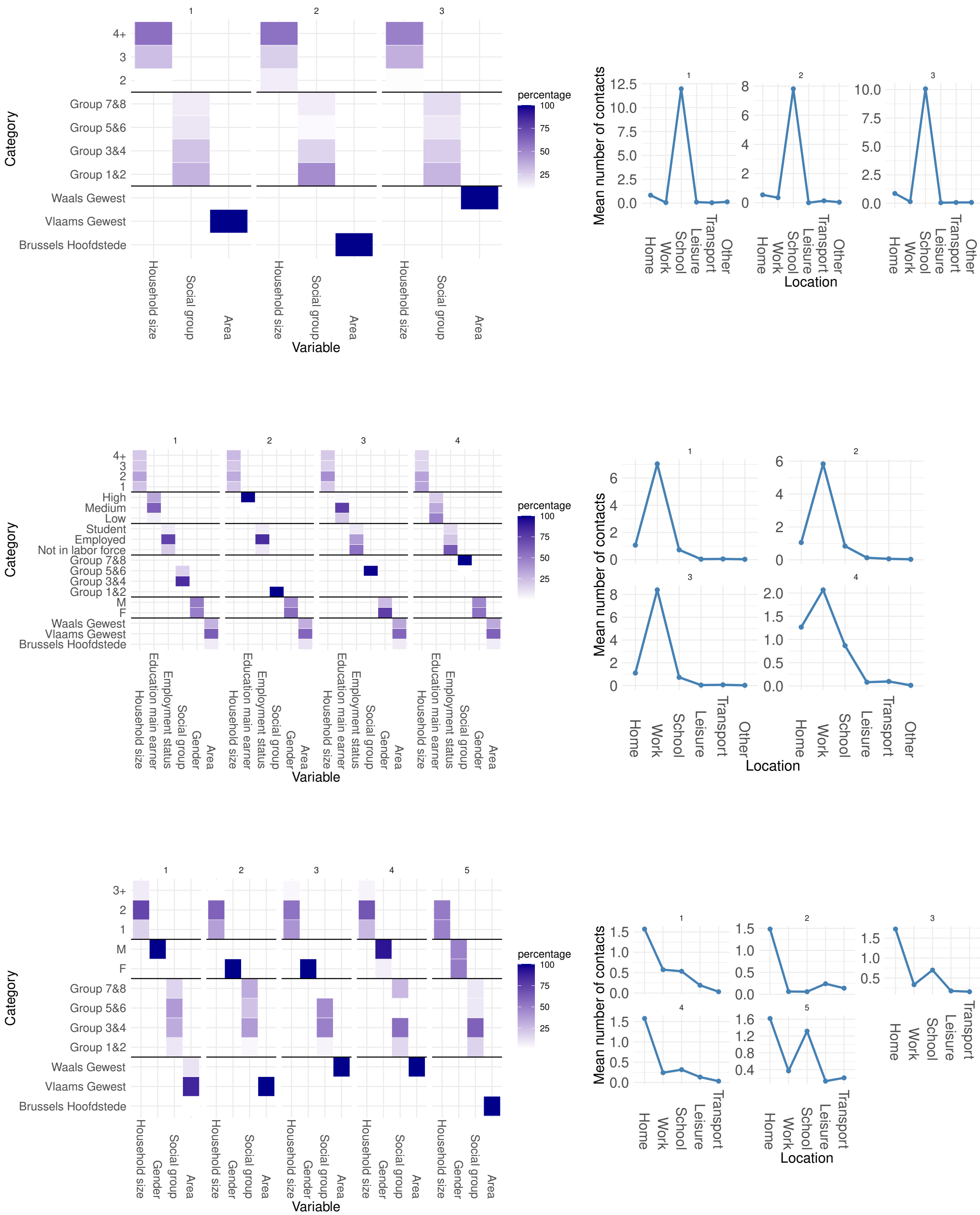


Figure 11: Clustering of participants with at least one non-household contact based on demographic characteristics (left panels) together with corresponding contact profiles per cluster (right panels) for children (upper panels), adults (middle panels) and elderly (bottom panels).

children were separated into three clusters which are clearly primarily characterised by their area of residence. In fact, each cluster only contained respondents from one area of residence. The corresponding contact profiles revealed that the average number of non-household contacts was the highest in the first cluster where only children from Vlaams Gewest are present, which is in line with results from the regression analyses. Based on demographic characteristics, the adult respondents were separated into four clusters. The first cluster predominantly consisted of employed participants from social group 3&4, whereas the second cluster consisted of primarily employed participants from social group 1&2 where the education of the main earner in the household was high. A large amount of participants in the third cluster were from social group 5&6 where the education of the highest earner in the household was medium, whereas the last cluster consisted of participants from social group 7&8 where a considerable amount was not in labor force. On average, a much lower average number of work-related non-household contacts was reported in the latter cluster which can be partly attributed to the employment status of a large proportion of these participants, as already indicated before and in the Hurdle 2 model. Finally, the elderly participants were separated in five clusters, primarily based on their area of residence and gender. Whereas the first two clusters mainly included respondents from Vlaams Gewest and respectively males and females, clusters 3 and 4 consisted of female and male participants from Waals Gewest. The fifth cluster included participants from Brussels Hoofdstedelijk Gewest. As expected, most of the non-household contacts of elderly respondents took place at their home.

4 Discussion

The main objective of this study was to investigate which factors had an influence on the reported number of non-household contacts based on the CoMix study in Belgium between 22 December 2020 and 8 March 2022. Two approaches were considered to estimate these effects. First, both NBI GAMLSS and GPO GAMLSS models were employed for three age categories. Secondly, three two-step Hurdle models were fitted where the drivers of the presence of at least one non-household contact were investigated in the first step, whereafter the factors that influenced the number of non-household contacts were investigated based on only the participations with non-zero number of reported non-household contacts. An agglomerative hierarchical clustering analysis was employed to identify potential explanations for the observed effects in the models.

4.1 Interpretation of the results

The age of the respondent had an influence on the number of non-household contacts, since children reported the highest average number of non-household contacts. This can be explained by the fact that children have a considerable amount of non-household contacts at school, a phenomenon which was also visible in the clustering analysis of contact patterns. This finding is in line with other studies (Bridgen et al., 2022; Gimma et al., 2022; Loedy et al., 2023). Based on the GAMLSS and Hurdle models and similar to Wong et al. (2023) and Loedy et al. (2023), wearing a face mask consistently had a positive effect on the presence and number of non-household contacts across all age categories. Several reasons can be given to explain this phenomenon. First of all, at many times and in various public places, wearing a face mask was mandatory during non-household contacts. Therefore, it will be more likely to have contacts outside the household if a person wears a face mask, which is reflected by the large positive association of face mask wearing in the Hurdle 1 model. People who often had to go outside also had to wear a face mask more often. In this case, a reverse reasoning can be made since having more non-household contacts results into the higher usage of a face mask instead. Moreover, some people could be aware of the risks of having contacts outside the household and therefore chose to wear a face mask for safety reasons. Thirdly, risk compensation could have taken place, where people had a safer feeling when wearing a face mask. This feeling of protection implies having more social interactions and showing higher-risk behavior. Note however that although wearing a face mask results in a lower risk of transmission of COVID-19, this advantage will be counteracted by the trend of making more contacts by face mask wearing participants. Furthermore, there was no information if the face mask was worn well and whether or not the contacted person was wearing a face mask was also not taken into account.

The area of residency of the respondent also had a significant effect on the presence and reported number of non-household contacts of adults and children. More specifically, people from these two age categories living in Vlaams Gewest were more likely to make non-household contacts compared to those in the two other regions, and the average number of non-household contacts was also higher. For elderly, this positive effects were not always present. This effect of area was also found in Coletti et al. (2020). In all GAMLSS models, an association between males and reporting fewer non-household contacts was found. However, this negative effect was only present in the Hurdle 1 model and not in the Hurdle 2 model. Therefore, based on this study, males do have a lower probability of having contacts outside the household than females which implies that males more often reported zero non-household contacts. This latter observation explains the negative effect of male in the GAMLSS models. For adults, this effect could be associated with the fact that women are more often employed in social sector jobs compared to men. Based on the clustering analysis of contact patterns by occupation category, it was also visible that participants who worked in the service sector reported on average a higher number of work-related non-household contacts, as also concluded by Thomas et al. (2021) and Soussand et al. (2025). Moreover, the positive association between income level and number of non-household contacts was also observed by Lucchini et al. (2024).

Based on the Hurdle 1 model, the relative odds of having non-household contacts were consistently lower in higher household sizes. Since these participants already have more contacts within their household, they feel less need to seek for non-household social contacts. This behavioral trend of respondents living in large households will decrease the transmission, although note that when such a participant get infected, more people in its household are prone to being contaminated as well. However, if a person had at least one contact outside its household, children and adults living in larger households had slightly more non-household contacts. Similar to the effect of male, the negative effect of household size in the GAMLSS models of children and adults was mostly attributable to the higher number of zero non-household contacts reported by participants living in larger households. Previous studies found a positive association between household size and the number of contacts (Loedy et al., 2024; Lucchini et al., 2024; Jarvis et al., 2024), although these studies did not restrict their focus to non-household contacts and also considered contacts made within the household, which naturally increase with household size.

In general, vaccinated participants reported having more non-household contacts, which is similar to findings from Reichmuth et al. (2023); Wambua et al. (2023) and Wong et al. (2023). Vaccination can provide a sense of safety, both in terms of one's own health as the health of others, which may encourage in engaging in more non-household contacts. It is well-known that, when infected, vaccinated people typically have a lower viral load, experience milder illness symptoms when infected and remain infectious for a shorter period. Therefore, they are less likely to transmit the virus to other people compared to unvaccinated individuals. However, vaccinated people can still become infected and transmit the virus. The safer feeling of individuals can result into risk compensation where people will make more contacts, also outside their household. This increases the probability of becoming infected and thereafter transmitting the virus. Moreover, adults with symptoms reported on average fewer non-household contacts and isolated themselves to protect others from becoming infected.

Time characteristics of the contacts (holiday and/or weekend) were also present. Based on the Hurdle models of children and adults, holiday periods showed a significant association with a lower odds of reporting non-household contacts and the number of non-household contacts for children, whereas the association between holiday periods and the odds of having non-household contacts was positive for adults. Moreover, the effect of weekend was positive for elderly and adults, but negative for children. These observations for children can be explained by the fact that schools were closed during holidays and weekends and the clustering analysis of contact patterns revealed that most of the non-household contacts of children took place at school. Therefore, as also noted by Backer et al. (2023), holiday periods and weekends lower the probability of transmission between children.

Finally, the results indicated that people reported fewer non-household contacts as they participated in more rounds of the survey. This effect was less pronounced in the Hurdle models for elderly. Loedy et al. (2023) discovered a similar phenomenon in the study of the mechanisms that drove contacts based on the CoMix study in Belgium and attributed this to survey fatigue, while this effect was also observed by Backer et al. (2023) in their study on dynamics of the non-household contacts. O'Reilly-Shah (2017) noted that respondent fatigue is often present in longitudinal survey studies and can be influenced by factors such as survey length and the type of questions. This fatigue effect also influenced the likelihood of reporting non-household contacts, leading to under-reporting and consequently underestimating the true number of non-household contacts.

The exploratory analysis demonstrated the heterogeneity in behavior where two clearly separated groups of contact profiles could be distinguished. On the one hand, a considerable amount of participants (almost) never reported non-household contacts. This highly cautious group systematically avoided non-household contacts, which can be driven by fear of infection, having elevated risk or a strong adherence to lockdown measures. On the other hand, the highly social group almost always reported non-household contacts. This behavior can be due to work-related obligations, a lower perceived risk of infection or less concerns about transmitting the virus to others. Targeted interventions, such as vaccination campaigns, encouraging face mask use or implementing targeted testing strategies, focusing on the highly social group may have a large impact on reducing transmission.

Note that the average number of reported non-household contacts did not increase substantially during the study period, although the stringency index decreased to some extent. This behavioral trend was also observed in previous studies, as already indicated in the introduction. A possible explanation for this phenomenon is the presence of a fatigue effect where respondents tend to report a lower number of non-household contacts as they participated in more survey rounds. Since at later waves, a higher proportion of respondents will be participating in multiple waves, the fatigue effect will have a larger impact and dampen the effective average number of non-household contacts. Moreover, the motivation to contribute to the survey in the CoMix study may initially be high. As time progresses and the number of cases, hospitalisations and lockdown measures were decreasing, participants may have perceived the pandemic as less severe, potentially reducing their motivation to accurately report all their contacts. It is also reasonable that, although restrictions were relaxed and people started working at the office and attending other activities again, people may still try to reduce the number of contacts outside their household and keep physical distance during interactions. This can also partly explain the clustering pattern of the low average number of leisure contacts made during this study.

4.2 Limitations and suggestions for future research

This study has several limitations that should be acknowledged. First of all, the CoMix survey is prone to several sources of bias, which will be further discussed in the section on ethical thinking. Moreover, a fatigue effect was found in the analyses performed in this thesis, which significantly influenced the reported number of non-household contacts. Loedy et al. (2023) estimated the fatigue effects in their study to predict the number of contacts that would have been reported if a respondent participated for the first time to the survey by estimating the effect of participating for the first time. Although not considered in this thesis, this methodology could also be applied to this data in future research to correct for under-reporting due to fatigue. Furthermore, several assumptions were made to account for the missingness present in the data. Future research can dive into a robustness analysis by considering multiple imputation methods to better understand the missing data mechanism whether or not missingness is independent of both observed and unobserved data (missing completely at random (MCAR)), only depends on observed data (missing at random (MAR)) or depends on both (missing not at random (MNAR)), where MNAR cannot be ruled out based on the observed data alone.

No cut-off value for the maximum number of reported contacts was used in this thesis. In contrast, some previous studies considered a cut-off value of 50 (Gimma et al., 2022 and Reichmuth et al., 2024) or 100 (Jarvis et al., 2024) contacts per individual to reduce the influence of reporting high numbers of contacts on the mean. Since over 99.6% of the responses included less than 100 non-household contacts and a large proportion of participants reported no or a small number of non-household contacts, the mean was employed instead of the median. Moreover, Jarvis et al. (2024) demonstrated the limited impact of considering a higher treshold value on their results. As a recommendation for future research, the effect of using different cut-off values on the estimates obtained in the GAMLSS and Hurdle models, together with considering the trimmed mean instead of the arithmetic mean could be investigated.

In the model building procedures, the variables wavecount and household size were treated as nominal instead of ordinal to examine the effect of each category on the average presence and number of non-household contacts compared to a reference category. However, some models showed a linear trend in one of these variables. Although not considered here, an ordinal version of the categorical variable could be considered instead in the corresponding models to improve statistical efficiency. This was not done in this thesis to consistently consider the same type of variable throughout all models. As discussed before, a limitation of the GPO GAMLSS models is that no random effect could be included in the variance model partly due to convergence issues, which had a considerable impact on the overall model fit. This limitation was partly accounted for by also considering the NBI GAMLSS model counterparts and revealed the limited impact on most of the estimates.

Both zero-inflation and Hurdle models were considered for the presence and number of non-household contacts where the overdispersion and excess of zeros were accounted for. Since the ZIP GAMLSS and ZINBI GAMLSS models did not converge, only results from the Hurdle models were reported and interpreted. The Hurdle models modeled on the one hand the odds of having non-household contacts and on the other hand the number of non-household contacts if at least one contact outside the household was reported. As opposed to zero-inflation models, zero counts can only be produced in the first step of Hurdle models as a zero-truncated probability distribution function for the number of contacts outside the household was employed in the second process. Note, however, that the assumption of two independent states (zero vs non-zero counts) makes the Hurdle modeling framework not entirely suitable for this dataset, as some participants almost always reported non-household contacts except in one survey round where they reported none. Nevertheless, the Hurdle models still provide useful insights regarding the contact behavior of the participants.

The DHARMa residual plots for the Hurdle 2 models considered in this thesis revealed that these models did not completely capture the variability in the data. Even though the negative binomial distribution was employed to account for overdispersion, there was still more dispersion than expected. First, there may be unmodeled heterogeneity that was not captured by the variables considered in these models. For example, due to the large number of missing values for socio-economic variables as income, occupation and education, it was only possible to account to some extent for SES of the participant. Instead of considering the education of the participant, the education level of the household's highest earner was employed since the two are assumed to be closely related. Secondly, these Hurdle 2 models could be improved by including a random slope for wavecount. This could account for some of the unexplained variability and reduce overdispersion, as the models considered in this thesis indicated that participants behave differently as the number of participations increased. It is furthermore also possible to explicitely model the dispersion parameter such that the variance could change with covariates. Finally, note that the Hurdle 2 models in this thesis considered a quadratic mean-variance relationship. A linear relation between the mean and variance was considered as well, but these models did not converge. Other relations can be examined in the future.

The clustering analyses of contact profiles made use of the commonly employed Ward's D2 measure of dissimilarity. However, note that also other linkage methods could be considered such as complete or single linkage clustering. Whereas the former method will produce more compact clusters, the latter clustering method will tend to produce long clusters. Group average clustering gives a compromise between the two linkage methods just described. More information about these different linkage methods can be found in

Hastie et al. (2009). Further research can perform the clustering analyses with different linkage methods and compare their performance based on the cophenetic correlation introduced by Sokal and Rohlf (1962). Finally, this thesis focused on modeling the number of non-household contacts, but many other insights could possibly be drawn from the CoMix study. One possible direction is to investigate which factors influence the probability of having physical contacts, which could be examined via a logistic regression model.

4.3 Ethical thinking, societal relevance and stakeholder awareness

This study is based on survey data collected via the CoMix study and involves a random sample of individuals from the Belgian population. For children, one of the parents filled in the survey on behalf of their child. Informed consent was collected and participants could autonomously decide to withdraw from the study or participate again at any time. There was also a possibility to leave specific questions open, including the income level of the respondent. Furthermore, the personal data was pseudo-anonymised to ensure privacy and confidentiality, such that the publicly available data is not personally identifiable.

As discussed before, the CoMix survey is, as a self-completed survey, prone to bias since participants may want to give socially desirable answers instead of their true behaviors (e.g. stating that they did not have any non-household contacts whereas in reality they had). This could make the results less accurate and not fully reflect reality. Anonymisation of the responses was employed to reduce this bias. Moreover, recall bias could be present where participants have difficulties in accurately recalling the number of contacts they had in the past. The disadvantage of this retrospective survey study is limited since only contacts made on the day before filling in the survey had to be remembered. Finally, this study is also prone to nonresponse bias, as individuals who chose not to participate may differ meaningfully from those who did, potentially skewing the results. As the pandemic progresses, changes in public perception about its severity may also result into selection bias, for example if individuals with certain levels of concern are more or less likely to continue participating in the survey over time.

Both significant and non-significant results are stated and discussed in the report, with all necessary details about the analyses given in the methodology section. Both the assumptions and limitations of all models considered in this thesis are discussed as well. Accurate statistical models were fitted to the real data to ensure the relevance of the resulting interpretations. Results from the models can serve as a guidance for decision makers to have a deeper understanding about possible drivers of transmission of COVID-19 due to social contact behavior, which are also relevant for other airborne diseases. Non-pharmaceutical interventions may influence the spread of airborne diseases, but may also have a large impact on the mental well-being of the population (Lwin et al., 2022; Massell et al., 2022; Colella et al., 2023). Therefore, decisions on future strategies about NPIs should not solely be based on models, but its societal impact (e.g. due to social isolation) should be considered as well.

Different stakeholders can be defined. First of all, researchers in social contact behavior can compare their results with the results of this study and, based on observations and limitations in this thesis, decide on what interesting directions could be pursued in further research. Our results can also be employed by researchers in infectious disease modeling to build models for the transmission of airborne diseases. Furthermore, public health officials and the government can consider the information in this thesis to develop data-driven strategies on the use of NPIs for disease control of future outbreaks of airborne diseases. Since the social contact patterns were different between the areas of residence, also regional governments can employ these results to decide on region-specific NPIs. Finally, note that the general population is a relevant stakeholder as well, as decisions from the governments on the NPIs will have an impact on their daily lives during future outbreaks of airborne diseases.

5 Conclusion

Based on data from the CoMix study, several factors were associated with the presence and number of reported non-household contacts. These drivers may have an impact on the spread/transmission of COVID-19 in the population. The average number of reported non-household contacts did not considerably increase after relaxing the lockdown restrictions indicating the longer-term impact of the pandemic on social contact behavior. However, the under-reporting due to participant survey fatigue has to be taken into account as well.

References

L. Angeli, C. P. Caetano, N. Franco, P. Coletti, C. Faes, G. Molenberghs, P. Beutels, S. Abrams, L. Willem, and N. Hens. Assessing the role of children in the covid-19 pandemic in belgium using perturbation analysis. *Nature communications*, 16(1):2230, 2025.

F. J. Anscombe. Sampling theory of the negative binomial and logarithmic series distributions. *Biometrika*, 37(3/4):358–382, 1950.

J. A. Backer, L. Bogaardt, P. Beutels, P. Coletti, W. J. Edmunds, A. Gimma, C. C. van Hagen, N. Hens, C. I. Jarvis, E. R. Vos, et al. Dynamics of non-household contacts during the covid-19 pandemic in 2020 and 2021 in the netherlands. *Scientific Reports*, 13(1): 5166, 2023.

J. A. Backer, E. R. Vos, G. den Hartog, C. C. van Hagen, H. E. de Melker, F. R. van der Klis, and J. Wallinga. Contact behaviour before, during and after the covid-19 pandemic in the netherlands: evidence from contact surveys, 2016 to 2017 and 2020 to 2023. *Eurosurveillance*, 29(43):2400143, 2024.

G. Béraud, S. Kazmercziak, P. Beutels, D. Levy-Bruhl, X. Lenne, N. Mielcarek, Y. Yazdanpanah, P.-Y. Boëlle, N. Hens, and B. Dervaux. The french connection: the first large population-based contact survey in france relevant for the spread of infectious diseases. *PloS one*, 10(7):e0133203, 2015.

J. R. Bridgen, C. Jewell, and J. M. Read. Social mixing patterns in the uk following the relaxation of covid-19 pandemic restrictions, july–august 2020: a cross-sectional online survey. *BMJ open*, 12(12):e059231, 2022.

S. v. Buuren and M. Fredriks. Worm plot: a simple diagnostic device for modelling growth reference curves. *Statistics in medicine*, 20(8):1259–1277, 2001.

S. Colella, F. Dufourt, V. A. Hildebrand, and R. Vivès. Mental health effects of covid-19 lockdowns: a twitter-based analysis. *Economics & Human Biology*, 51:101307, 2023.

P. Coletti, J. Wambua, A. Gimma, L. Willem, S. Vercruysse, B. Vanhoutte, C. I. Jarvis, K. Van Zandvoort, J. Edmunds, P. Beutels, et al. Comix: comparing mixing patterns in the belgian population during and after lockdown. *Scientific reports*, 10(1):21885, 2020.

J. G. Cragg. Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica: journal of the Econometric Society*, pages 829–844, 1971.

L. Di Domenico, M. L. Reichmuth, and C. L. Althaus. Individual-based and neighbourhood-based socio-economic factors relevant for contact behaviour and epidemic control. *medRxiv*, pages 2025–03, 2025.

Z. Dobreva, A. Gimma, H. Rohan, B. Djoudalbaye, A. Tshangela, C. I. Jarvis, K. van Zandvoort, and M. Quaife. Characterising social contacts under covid-19 control measures in africa. *BMC medicine*, 20(1):344, 2022.

D. M. Feehan and A. S. Mahmud. Quantifying population contact patterns in the united states during the covid-19 pandemic. *Nature communications*, 12(1):893, 2021.

C. X. Feng. A comparison of zero-inflated and hurdle models for modeling zero-inflated count data. *Journal of statistical distributions and applications*, 8(1):8, 2021.

A. Gimma, J. D. Munday, K. L. Wong, P. Coletti, K. van Zandvoort, K. Prem, C. C.-. working group, P. Klepac, G. J. Rubin, S. Funk, et al. Comix: Changes in social contacts as measured by the contact survey during the covid-19 pandemic in england between march 2020 and march 2021. *MedRxiv*, pages 2021–05, 2022.

W. H. Greene. Accounting for excess zeros and sample selection in poisson and negative binomial regression models. 1994.

P. L. Gupta, R. C. Gupta, and R. C. Tripathi. Analysis of zero-adjusted count data. *Computational Statistics & Data Analysis*, 23(2):207–218, 1996.

T. Hale, N. Angrist, R. Goldszmidt, B. Kira, A. Petherick, T. Phillips, S. Webster, E. Cameron-Blake, L. Hallas, S. Majumdar, et al. A global panel database of pandemic policies (oxford covid-19 government response tracker). *Nature human behaviour*, 5(4):529–538, 2021.

F. Hartig and M. F. Hartig. Package 'dharma'. *R package*, 531:532, 2017.

T. Hastie and R. Tibshirani. Varying-coefficient models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 55(4):757–779, 1993.

T. Hastie, R. Tibshirani, and J. Friedman. *Unsupervised Learning*, pages 485–585. Springer New York, New York, NY, 2009. ISBN 978-0-387-84858-7. doi: 10.1007/978-0-387-84858-7_14. URL https://doi.org/10.1007/978-0-387-84858-7_14.

C. I. Jarvis, K. Van Zandvoort, A. Gimma, K. Prem, P. Klepac, G. J. Rubin, and W. J. Edmunds. Quantifying the impact of physical distance measures on the transmission of covid-19 in the uk. *BMC medicine*, 18:1–10, 2020.

C. I. Jarvis, P. Coletti, J. A. Backer, J. D. Munday, C. Faes, P. Beutels, C. L. Althaus, N. Low, J. Wallinga, N. Hens, et al. Social contact patterns following the covid-19 pandemic: a snapshot of post-pandemic behaviour from the comix study. *Epidemics*, 48: 100778, 2024.

L. Kaufman and P. J. Rousseeuw. *Finding groups in data: an introduction to cluster analysis*. John Wiley & Sons, 2005.

C. Kremer, L. Willem, J. Boone, W. Arrazola de Oñate, N. Hammami, C. Faes, and N. Hens. Key performance indicators of covid-19 contact tracing in belgium from september 2020 to december 2021. *Plos one*, 18(10):e0292346, 2023.

M. Kretzschmar and R. T. Mikolajczyk. Contact profiles in eight european countries and implications for modelling the spread of airborne infectious diseases. *PloS one*, 4(6): e5931, 2009.

D. Lambert. Zero-inflated poisson regression, with an application to defects in manufacturing. *Technometrics*, 34(1):1–14, 1992.

N. Loedy, P. Coletti, J. Wambua, L. Hermans, L. Willem, C. I. Jarvis, K. L. Wong, W. J. Edmunds, A. Robert, Q. J. Leclerc, et al. Longitudinal social contact data analysis: insights from 2 years of data collection in belgium during the covid-19 pandemic. *BMC Public Health*, 23(1):1298, 2023.

N. Loedy, L. Hermans, M. Bosschaert, A. Torneri, and N. Hens. The role of frailty in shaping social contact patterns in belgium, 2022–2023. *Scientific Reports*, 15(1):12883, 2025.

L. Lucchini, V. Marziano, F. Trentini, C. Chiavenna, E. D'Agnese, V. Offeddu, M. Manica, P. Poletti, D. Balsamo, G. Guzzetta, et al. Post-pandemic social contacts in italy: implications for social distancing measures on in-person school and work attendance. *arXiv preprint arXiv:2412.18549*, 2024.

M. O. Lwin, A. Sheldenkar, J. Lu, P. J. Schulz, W. Shin, C. Panchapakesan, R. K. Gupta, and Y. Yang. The evolution of public sentiments during the covid-19 pandemic: case comparisons of india, singapore, south korea, the united kingdom, and the united states. *JMIR infodemiology*, 2(1):e31473, 2022.

J. Massell, R. Lieb, A. Meyer, and E. Mayor. Fluctuations of psychological states on twitter before and during covid-19. *PLoS One*, 17(12):e0278018, 2022.

J. Mossong, N. Hens, M. Jit, P. Beutels, K. Auranen, R. Mikolajczyk, M. Massari, S. Salmaso, G. S. Tomba, J. Wallinga, et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS medicine*, 5(3):e74, 2008.

F. Murtagh and P. Legendre. Ward's hierarchical agglomerative clustering method: which algorithms implement ward's criterion? *Journal of classification*, 31:274–295, 2014.

V. N. O'Reilly-Shah. Factors influencing healthcare provider respondent fatigue answering a globally administered in-app survey. *PeerJ*, 5:e3785, 2017.

H. T. Phuong, A. Bartz, A. K. Jarynowski, B. Lange, C. I. Jarvis, N. Rübsamen, R. T. Mikolajczyk, S. Scholz, T. Berger, T. Heinsohn, et al. Changes in social contact patterns in germany during the sars-cov-2 pandemic–an analysis based on the covimod study. *BMC Infectious Diseases*, 25(1):588, 2025.

B. J. Quilty, L. A. Chapman, J. D. Munday, K. L. Wong, A. Gimma, S. Pickering, S. J. Neil, R. P. Galão, W. J. Edmunds, C. I. Jarvis, et al. Disentangling the drivers of heterogeneity in sars-cov-2 transmission from data on viral load and daily contact rates. *medRxiv*, pages 2024–08, 2024.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2023. URL `https://www.R-project.org/`.

M. L. Reichmuth, L. Heron, J. Riou, A. Moser, A. Hauser, N. Low, and C. L. Althaus. Socio-demographic characteristics associated with covid-19 vaccination uptake in switzerland: longitudinal analysis of the comix study. *BMC Public Health*, 23(1):1523, 2023.

M. L. Reichmuth, L. Heron, P. Beutels, N. Hens, N. Low, and C. L. Althaus. Social contacts in switzerland during the covid-19 pandemic: Insights from the comix study. *Epidemics*, 47:100771, 2024.

M. Ridout, C. G. Demétrio, and J. Hinde. Models for count data with many zeros. In *Proceedings of the XIXth international biometric conference*, volume 19, pages 179–192. International Biometric Society Invited Papers Cape Town, South Africa, 1998.

R. A. Rigby and D. M. Stasinopoulos. Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 54(3):507–554, 2005.

R. A. Rigby, M. D. Stasinopoulos, G. Z. Heller, and F. De Bastiani. *Distributions for modeling location, scale, and shape: Using GAMLSS in R*. Chapman and Hall/CRC, 2019.

P. Royston and E. M. Wright. Goodness-of-fit statistics for age-specific reference intervals. *Statistics in medicine*, 19(21):2943–2962, 2000.

R. R. Sokal and F. J. Rohlf. The comparison of dendrograms by objective methods. *Taxon*, pages 33–40, 1962.

F. Soussand, A. Y. Abdou, M. Sanchez, B.-T. Huynh, C. Giese, C. Tran-Kiem, G. Béraud, D. Guillemot, S. Cauchemez, L. Opatowski, et al. Evolution of social contacts patterns in france over the sars-cov-2 pandemic: results from the socialcov survey. *BMC Infectious Diseases*, 25(1):224, 2025.

D. Stasinopoulos and R. Rigby. Generalized additive models for location scale and shape (gamlss) in r. *Journal of Statistical Software*, 23, 12 2007. doi: 10.18637/jss.v023.i07.

M. Stasinopoulos, B. Rigby, and C. Akantziliotou. Instructions on how to use the gamlss package in r second edition, 2008.

A. Thomas, L. Danon, H. Christensen, K. Northstone, D. Smith, E. Nixon, A. Trickey, G. Hemani, S. Sauchelli, A. Finn, et al. Limits of lockdown: characterising essential contacts during strict physical distancing. *medRxiv*, pages 2021–03, 2021.

G. Thomas, L. R. Nakamura, R. A. Moral, and C. G. Demétrio. Modeling data with zero inflation and overdispersion using gamlsss. *arXiv preprint arXiv:1810.02618*, 2018.

M. Tizzani and L. Gauvin. Socioeconomic determinants of protective behaviors and contact patterns in the post-covid-19 pandemic era: a cross-sectional study in italy, 2024. URL `https://arxiv.org/abs/2407.20350`.

M. Tizzani, A. De Gaetano, C. I. Jarvis, A. Gimma, K. Wong, W. J. Edmunds, P. Beutels, N. Hens, P. Coletti, and D. Paolotti. Impact of tiered measures on social contact and mixing patterns of in italy during the second wave of covid-19. *BMC Public Health*, 23 (1):906, 2023.

L. Veneti, B. Robberstad, A. Steens, F. Forland, B. A. Winje, D. F. Vestrheim, C. I. Jarvis, A. Gimma, W. J. Edmunds, K. Van Zandvoort, et al. Social contact patterns during the early covid-19 pandemic in norway: insights from a panel study, april to september 2020. *BMC Public Health*, 24(1):1438, 2024.

F. Verelst, L. Hermans, S. Vercruysse, A. Gimma, P. Coletti, J. A. Backer, K. L. Wong, J. Wambua, K. van Zandvoort, L. Willem, et al. Socrates-comix: a platform for timely and open-source contact mixing data during and in between covid-19 surges and interventions in over 20 european countries. *BMC medicine*, 19:1–7, 2021.

J. Walde, M. Chaturvedi, T. Berger, A. Bartz, R. Killewald, D. V. Tomori, N. Rübsamen, B. Lange, S. Scholz, M. Treskova, et al. Effect of risk status for severe covid-19 on individual contact behaviour during the sars-cov-2 pandemic in 2020/2021—an analysis based on the german covimod study. *BMC infectious diseases*, 23(1):205, 2023.

J. Wambua, L. Hermans, P. Coletti, F. Verelst, L. Willem, C. I. Jarvis, A. Gimma, K. L. Wong, A. Lajot, S. Demarest, et al. The influence of risk perceptions on close contact frequency during the sars-cov-2 pandemic. *Scientific reports*, 12(1):5192, 2022.

J. Wambua, N. Loedy, C. I. Jarvis, K. L. Wong, C. Faes, R. Grah, B. Prasse, F. Sandmann, R. Niehus, H. Johnson, et al. The influence of covid-19 risk perception and vaccination status on the number of social contacts across europe: insights from the comix study. *BMC Public Health*, 23(1):1350, 2023.

J. H. Ward Jr. Hierarchical grouping to optimize an objective function. *Journal of the American statistical association*, 58(301):236–244, 1963.

A. H. Welsh, R. B. Cunningham, C. F. Donnelly, and D. B. Lindenmayer. Modelling the abundance of rare species: statistical models for counts with extra zeros. *Ecological Modelling*, 88(1-3):297–308, 1996.

L. Willem, S. Abrams, P. J. Libin, P. Coletti, E. Kuylen, O. Petrof, S. Møgelmose, J. Wambua, S. A. Herzog, C. Faes, et al. The impact of contact tracing and household bubbles on deconfinement strategies for covid-19. *Nature communications*, 12(1): 1524, 2021.

K. L. Wong, A. Gimma, E. S. Paixao, C. Faes, P. Beutels, N. Hens, C. I. Jarvis, and W. J. Edmunds. Pregnancy during covid-19: social contact patterns and vaccine coverage of pregnant women from comix in 19 european countries. *BMC pregnancy and childbirth*, 22(1):757, 2022.

K. L. Wong, A. Gimma, P. Coletti, C. Faes, P. Beutels, N. Hens, V. K. Jaeger, A. Karch, H. Johnson, et al. Social contact patterns during the covid-19 pandemic in 21 european countries–evidence from a two-year study. *BMC infectious diseases*, 23(1):268, 2023.

B. Yadav, L. Jeyaseelan, V. Jeyaseelan, J. Durairaj, S. George, K. Selvaraj, and S. I. Bangdiwala. Can generalized poisson model replace any other count data models? an evaluation. *Clinical Epidemiology and Global Health*, 11:100774, 2021.

Appendices

A.1 Software code

Software code for this thesis can be found via `https://github.com/StijnLapere/CoMix_study_non-household_contacts`.

A.2 Additional descriptive tables and figures

Table 5: Time periods where every wave of data collection took place.

Wave	Period	Wave	Period
12	22 Dec ’20 –04 Jan ’21	28	03 Aug ’21–10 Aug ’21
13	05 Jan ’21 –11 Jan ’21	29	17 Aug ’21–23 Aug ’21
14	19 Jan ’21 –24 Jan ’21	30	31 Aug ’21–07 Sep ’21
15	02 Feb ’21 –07 Feb ’21	31	14 Sep ’21 –20 Sep ’21
16	16 Feb ’21 –23 Feb ’21	32	28 Sep ’21 –04 Oct ’21
17	02 Mar ’21–09 Mar ’21	33	12 Oct ’21 –17 Oct ’21
18	16 Mar ’21–23 Mar ’21	34	27 Oct ’21 –03 Nov ’21
19	30 Mar ’21–06 Apr ’21	35	09 Nov ’21–15 Nov ’21
20	13 Apr ’21 –19 Apr ’21	36	23 Nov ’21–29 Nov ’21
21	27 Apr ’21 –03 May ’21	37	07 Dec ’21 –13 Dec ’21
22	12 May ’21–19 May ’21	38	21 Dec ’21 –28 Dec ’21
23	25 May ’21–01 Jun ’21	39	04 Jan ’22 –11 Jan ’22
24	09 Jun ’21 –16 Jun ’21	40	18 Jan ’22 –23 Jan ’22
25	22 Jun ’21 –27 Jun ’21	41	01 Feb ’22 –08 Feb ’22
26	06 Jul ’21 –14 Jul ’21	42	16 Feb ’22 –22 Feb ’22
27	20 Jul ’21 –26 Jul ’21	43	01 Mar ’22–08 Mar ’22

Table 6: More detailed description of categories of socio-economic variables.

Variable	Category	Groups included
Education main earner	Low	Without a diploma or primary education
		General lower secondary education (first 3 years completed)
		Technical/artistic/professional lower secondary education (first 3 years)
	Medium	General upper secondary education (6 years completed)
		Professional upper secondary (6 years)
		Technical or artistic upper secondary education (6 years)
	High	Higher education: graduate, candidature, bachelor
		University education: bachelor's degree, post-graduate master's degree
		Complementary master
		Doctorate
Employment status	Employed	Employed full-time (34 hours or more)
		Employed part-time (less than 34 hours)
		Self-employed
	Not in labor force	Full-time parent homemaker
		Long-term sick or disabled
		Retired
		Unemployed but looking for a job
		Unemployed and not looking for a job
	Student	Student/pupil
Income	Very low	€0 - €1299
	Low	€1300 - €1899
	Middle	€1900 - €3200
	High	€3200 - €4499
	Very high	$\geq$ €4500
Occupation	Managers/Professionals	Liberal profession or profession with qualification required
		Member of general management senior executive
		Middle management not part of general management
	Office employees	Other employee who mainly perform office work
	Service employees	Other employee who does not do office work (eg teacher, nurses ...)
	Manual workers	Non-skilled worker
		Skilled worker
	Self-employed/	Craftsman trader with 5 employees or less
	Small business	Industrial wholesaler with 6 employees or more
		Farmer
	Not in labor force	Houseman or housewife
		Unable to work
		Never worked
		Unemployed
		(pre-)retired
		Student

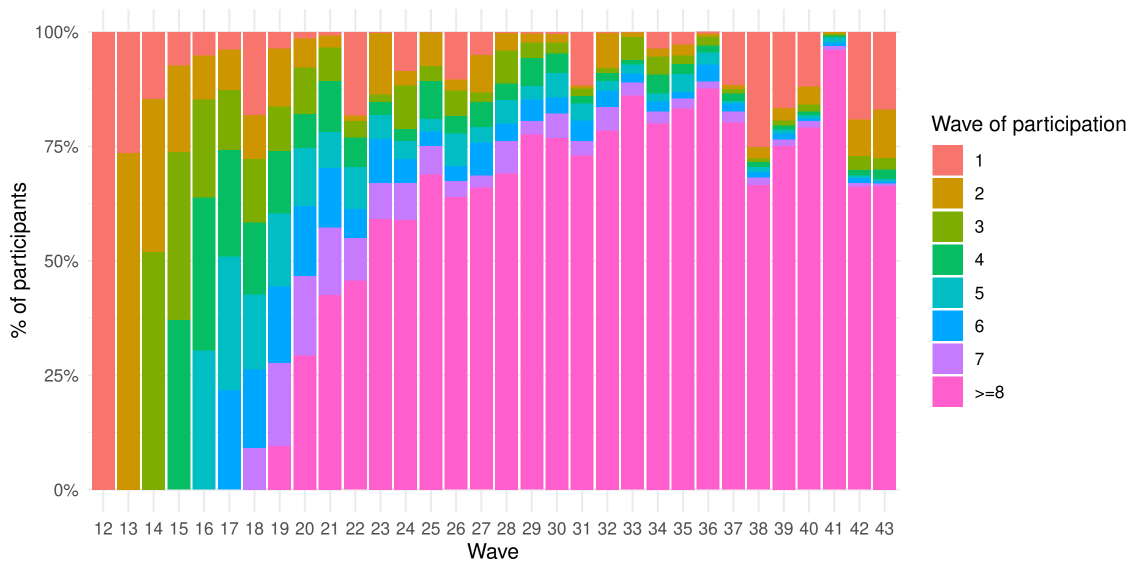


Figure 12: Proportion of the number of times a participant already participated to the CoMix study per wave.

A.3 Coefficients of NBI and GPO GAMLSS models

Table 7: Coefficients for mu for NBI GAMLSS elderly.

Parameter	Estimate	SE	p-value
Intercept	-0.591262	0.152045	< 0.001
Vacc Yes	0.680269	0.040828	< 0.001
Face mask Yes	0.381465	0.147819	0.010
Symptoms Yes	-0.055198	0.034870	0.113
Vlaams Gewest	0.022659	0.144789	0.876
Waals Gewest	-0.178782	0.150306	0.234
Holiday Yes	-0.087728	0.028273	0.002
Weekend	0.039205	0.028273	0.166
hh size 2	0.086573	0.028278	0.002
hh size 3+	-0102832	0.070358	0.144
2 waves	0.113599	0.068108	0.095
3 waves	-0.236361	0.072007	0.001
4 waves	-0.008987	0.075504	0.905
5 waves	-0.297800	0.077364	< 0.001
6 waves	-0.350155	0.079367	< 0.001
7 waves	-0.313700	0.078051	< 0.001
8+ waves	-0.454829	0.052329	< 0.001
Male	-0.121366	0.026794	< 0.001
Vlaams Gewest : Face mask Yes	0.247030	0.153813	0.108
Waals Gewest : Face mask Yes	0.074947	0.161006	0.642

Table 8: Coefficients for mu for NBI GAMLSS children.

Parameter	Estimate	SE	p-value
Intercept	0.17979	0.09960	0.071
Face mask Yes	0.53104	0.03518	< 0.001
Vlaams Gewest	1.31550	0.08915	< 0.001
Waals Gewest	0.12408	0.09316	0.183
Holiday Yes	−0.32183	0.14813	0.030
Weekend	−0.04058	0.05013	0.418
2 waves	0.05629	0.06455	0.383
3 waves	−0.35745	0.07412	< 0.001
4 waves	−0.33201	0.07304	< 0.001
5 waves	−0.33650	0.08984	< 0.001
6 waves	−0.60372	0.09164	< 0.001
7 waves	−0.45255	0.10364	< 0.001
8+ waves	−0.70211	0.04597	< 0.001
Vlaams Gewest : Holiday Yes	−0.17951	0.15644	0.251
Waals Gewest : Holiday Yes	−0.16132	0.16985	0.342

Table 9: Coefficients for mu for NBI GAMLSS adults.

Parameter	Estimate	SE	p-value
Intercept	-1.05014	0.06439	< 0.001
Social Group 3&4	0.07557	0.02192	< 0.001
Social Group 5&6	0.04732	0.02760	0.086
Social Group 7&8	-0.17937	0.02783	< 0.001
Vacc Yes	0.43206	0.07144	< 0.001
Elevated risk Yes	-0.18980	0.04702	< 0.001
Face mask Yes	0.85116	0.03801	< 0.001
Symptoms Yes	-0.06448	0.01513	< 0.001
Vlaams Gewest	0.90859	0.04522	< 0.001
Waals Gewest	0.14886	0.04967	0.003
Holiday Yes	-0.02578	0.01573	0.101
Weekend	0.07799	0.02042	< 0.001
hh size 2	-0.05192	0.01857	0.005
hh size 3	-0.15338	0.02215	< 0.001
hh size 4+	-0.24405	0.02275	< 0.001
2 waves	0.09714	0.03379	0.004
3 waves	-0.07670	0.03727	0.040
4 waves	-0.21297	0.03952	< 0.001
5 waves	-0.27232	0.04083	< 0.001
6 waves	-0.29780	0.04067	< 0.001
7 waves	-0.49281	0.04308	< 0.001
8+ waves	-0.46260	0.01798	< 0.001
Male	-0.10252	0.01463	< 0.001
Low education	-0.06530	0.04251	0.124
High education	0.04133	0.03024	0.172
Vacc Yes : Face mask Yes	-0.16838	0.04458	< 0.001
Elevated risk Yes : Face mask Yes	0.18100	0.05021	< 0.001
Vacc Yes : Low education	-0.03514	0.05005	0.483
Vacc Yes : High education	0.12371	0.03229	< 0.001
Vacc Yes : Vlaams Gewest	-0.14788	0.05783	0.011
Vacc Yes : Waals Gewest	0.05245	0.06289	0.404

Table 10: Coefficients for sigma for NBI GAMLSS elderly.

Parameter	Estimate	SE	p-value
Intercept	-0.50952	0.38087	0.181
Vacc Yes	0.39576	0.15593	0.011
Face mask Yes	-1.22537	0.34344	< 0.001
Symptoms Yes	0.19692	0.10656	0.065
Vlaams Gewest	0.31339	0.32326	0.332
Waals Gewest	1.03256	0.33127	0.002
Holiday Yes	0.06745	0.08712	0.439
Weekend	0.42585	0.08106	< 0.001
2 waves	-0.03755	0.22691	0.036
3 waves	-0.50911	0.27048	0.060
4 waves	0.48012	0.22870	0.036
5 waves	0.24199	0.24786	0.329
6 waves	0.29340	0.23225	0.207
7 waves	0.06759	0.24453	0.782
8+ waves	0.04028	0.18116	0.824
Male	0.46056	0.08445	< 0.001
Vlaams Gewest : Face mask Yes	-0.08379	0.36047	0.816
Waals Gewest : Face mask Yes	-0.89672	0.37605	0.017

Table 11: Coefficients for sigma for NBI GAMLSS children.

Parameter	Estimate	SE	p-value
Intercept	0.69004	0.15889	< 0.001
Face mask Yes	-0.59891	0.05812	< 0.001
Vlaams Gewest	-0.88301	0.13691	< 0.001
Waals Gewest	-0.50381	0.14520	< 0.001
Holiday Yes	-0.49432	0.25639	0.054
Weekend	0.13098	0.07383	0.076
2 waves	1.00702	0.11630	< 0.001
3 waves	1.18894	0.12686	< 0.001
4 waves	1.03239	0.13536	< 0.001
5 waves	1.54811	0.13955	< 0.001
6 waves	1.49510	0.15163	< 0.001
7 waves	1.81874	0.15130	< 0.001
8+ waves	1.86716	0.10750	< 0.001
Vlaams Gewest : Holiday Yes	0.88196	0.26687	< 0.001
Waals Gewest : Holiday Yes	0.77680	0.28783	< 0.001

Table 12: Coefficients for sigma for NBI GAMLSS adults.

Parameter	Estimate	SE	p-value
Intercept	-4.184320	0.165465	< 0.001
Vacc Yes	-0.382833	0.051819	< 0.001
Elevated risk Yes	0.008556	0.104322	0.935
Face mask Yes	-1.302687	0.056924	< 0.001
Symptoms Yes	0.055563	0.045670	0.224
Vlaams Gewest	-0.077878	0.086241	0.367
Waals Gewest	-0.160117	0.092857	0.085
Holiday Yes	0.157468	0.047161	< 0.001
Weekend	0.038639	0.050661	0.446
2 waves	6.053423	0.143036	< 0.001
3 waves	6.160135	0.147742	< 0.001
4 waves	6.196243	0.152166	< 0.001
5 waves	6.117824	0.155786	< 0.001
6 waves	5.946735	0.160240	< 0.001
7 waves	5.793944	0.165900	< 0.001
8+ waves	6.066425	0.134969	< 0.001
Male	0.183651	0.043661	< 0.001
Low education	0.145399	0.073560	0.048
High education	-0.312642	0.046334	< 0.001
Elevated risk Yes : Face mask Yes	-0.466475	0.118947	< 0.001

Table 13: Coefficients for mu for GPO GAMLSS elderly.

Parameter	Estimate	SE	p-value
Intercept	0.486520	0.115675	< 0.001
Vacc Yes	0.655594	0.094251	< 0.001
Elevated risk Yes	-0.019628	0.029134	0.501
Face mask Yes	0.781521	0.091757	< 0.001
Symptoms Yes	-0.015642	0.038261	0.683
Vlaams Gewest	0.213975	0.056081	< 0.001
Waals Gewest	-0.145527	0.059404	0.014
Holiday Yes	-0.092588	0.032005	0.004
Weekend	0.058483	0.032478	0.072
hhsize 2	-0.001194	0.031644	0.970
hhsize 3+	-0.194896	0.081320	0.017
2 waves	0.195183	0.072784	0.007
3 waves	-0.159409	0.076471	0.037
4 waves	0.023364	0.082265	0.776
5 waves	-0.242724	0.082816	0.003
6 waves	-0.123869	0.090532	0.171
7 waves	-0.234290	0.084516	0.006
8+ waves	-0.371801	0.054508	< 0.001
Male	-0.137434	0.030027	< 0.001
Vacc Yes : Face mask Yes	-0.069611	0.099750	0.485

Table 14: Coefficients for mu for GPO GAMLSS children.

Parameter	Estimate	SE	p-value
Intercept	3.06955	0.12707	< 0.001
Face mask Yes	0.64410	0.04344	< 0.001
Vlaams Gewest	1.54548	0.09938	< 0.001
Waals Gewest	0.31855	0.10350	0.002
Holiday Yes	0.21760	0.19487	0.264
Weekend	-0.11558	0.05913	0.051
hhsize 3	-0.12601	0.07236	0.082
hhsize 4+	-0.23211	0.06834	< 0.001
2 waves	0.26659	0.08713	0.002
3 waves	-0.42193	0.09385	< 0.001
4 waves	-0.52649	0.08946	< 0.001
5 waves	-0.32428	0.12180	0.008
6 waves	-0.98464	0.10673	< 0.001
7 waves	-0.69363	0.12748	< 0.001
8+ waves	-1.11953	0.05828	< 0.001
Vlaams Gewest : Holiday Yes	-0.90540	0.20415	< 0.001
Waals Gewest : Holiday Yes	-0.85949	0.21733	< 0.001

Table 15: Coefficients for mu for GPO GAMLSS adults.

Parameter	Estimate	SE	p-value
Intercept	1.51018	0.11769	< 0.001
Vacc Yes	0.50959	0.08994	< 0.001
Elevated risk Yes	−0.03406	0.02197	0.121
Face mask Yes	1.03867	0.10926	< 0.001
Symptoms Yes	−0.13302	0.02046	< 0.001
Vlaams Gewest	0.81612	0.11211	< 0.001
Waals Gewest	0.15820	0.11760	0.179
Holiday Yes	−0.03810	0.02121	0.072
Weekend	0.07055	0.02406	0.003
hhsize 2	−0.02354	0.02347	0.316
hhsize 3	−0.09835	0.02911	< 0.001
hhsize 4+	−0.23687	0.03058	< 0.001
2 waves	−0.07609	0.04255	0.074
3 waves	−0.32101	0.04656	< 0.001
4 waves	−0.52213	0.04756	< 0.001
5 waves	−0.54687	0.04804	< 0.001
6 waves	−0.52876	0.05220	< 0.001
7 waves	−0.80727	0.05049	< 0.001
8+ waves	−0.75989	0.02746	< 0.001
Male	−0.08706	0.01950	< 0.001
Low education	−0.26977	0.05188	< 0.001
High education	−0.01767	0.03562	0.620
Vacc Yes : Face mask Yes	−0.19186	0.05581	< 0.001
Vacc Yes : Vlaams Gewest	−0.26141	0.07321	< 0.001
Vacc Yes : Waals Gewest	−0.05482	0.07815	0.483
Vacc Yes : Low education	0.11738	0.06476	0.070
Vacc Yes : High education	0.22694	0.04413	< 0.001
Face mask Yes : Vlaams Gewest	0.04995	0.10869	0.646
Face mask Yes : Waals Gewest	−0.07278	0.11353	0.521

Table 16: Coefficients for sigma for GPO GAMLSS elderly.

Parameter	Estimate	SE	p-value
Intercept	-1.353458	0.263408	< 0.001
Vacc Yes	-0.024746	0.196598	0.900
Elevated risk Yes	-0.371119	0.062329	< 0.001
Face mask Yes	-1.356939	0.196333	< 0.001
Symptoms Yes	0.223195	0.085458	0.009
Vlaams Gewest	0.012477	0.112029	0.911
Waals Gewest	-0.141419	0.123960	0.254
Holiday Yes	-0.009736	0.067587	0.885
Weekend	0.304400	0.067587	< 0.001
hhsize 2	0.110857	0.068581	0.106
hhsize 3+	0.513540	0.123124	< 0.001
2 waves	0.263773	0.190974	0.167
3 waves	0.038162	0.229924	0.868
4 waves	0.663957	0.194954	< 0.001
5 waves	0.598480	0.202950	0.003
6 waves	0.942402	0.198214	< 0.001
7 waves	0.554680	0.198214	0.005
8+ waves	0.405440	0.155484	0.009
Male	0.296946	0.066210	0.036
Vacc Yes : Face mask Yes	0.675805	0.210241	0.001

Table 17: Coefficients for sigma for GPO GAMLSS children.

Parameter	Estimate	SE	p-value
Intercept	-0.47674	0.13494	< 0.001
Face mask Yes	-0.29354	0.04724	< 0.001
Vlaams Gewest	-0.67994	0.09896	< 0.001
Waals Gewest	-0.23230	0.10535	0.027
Holiday Yes	0.26843	0.05225	< 0.001
Weekend	0.08800	0.06149	0.152
hhsize 3	0.09095	0.08504	0.285
hhsize 4+	0.26285	0.08021	0.001
2 waves	0.68206	0.08885	< 0.001
3 waves	0.78705	0.09881	< 0.001
4 waves	0.60192	0.10471	< 0.001
5 waves	1.11812	0.10840	< 0.001
6 waves	0.90287	0.12342	< 0.001
7 waves	1.20692	0.12128	< 0.001
8+ waves	1.21878	0.08163	< 0.001

Table 18: Coefficients for sigma for GPO GAMLSS adults.

Parameter	Estimate	SE	p-value
Intercept	-1.117336	0.186636	< 0.001
Vacc Yes	0.624263	0.142172	< 0.001
Elevated risk Yes	-0.010594	0.038911	0.785
Face mask Yes	-1.181165	0.158363	< 0.001
Symptoms Yes	0.117262	0.035248	< 0.001
Vlaams Gewest	0.014376	0.176758	0.935
Waals Gewest	0.306036	0.190210	0.108
Holiday Yes	0.021503	0.037432	0.566
Weekend	-0.009564	0.040828	0.815
hhsize 2	0.298901	0.042417	< 0.001
hhsize 3	0.348997	0.048875	< 0.001
hhsize 4+	0.264342	0.056861	< 0.001
2 waves	0.836547	0.072317	< 0.001
3 waves	0.874209	0.077984	< 0.001
4 waves	0.856696	0.085793	< 0.001
5 waves	0.811996	0.089810	< 0.001
6 waves	0.992357	0.088934	< 0.001
7 waves	0.723333	0.101949	< 0.001
8+ waves	1.058194	0.060206	< 0.001
Male	0.031334	0.034517	0.364
Low education	0.003061	0.059844	0.959
High education	-0.142263	0.035474	< 0.001
Vacc Yes : Vlaams Gewest	-0.419240	0.146168	0.004
Vacc Yes : Waals Gewest	-0.631203	0.159811	< 0.001
Face mask Yes : Vlaams Gewest	0.250498	0.165210	0.129
Face mask Yes : Waals Gewest	-0.150463	0.178032	0.398

A.4 Model diagnostic plots for GAMLSS models

Table 19: The mean, variance, coefficient of skewness and coefficient of kurtosis of the quantile residuals of the different GAMLSS models.

Model	Mean	Variance	Skewness	Kurtosis
NBI children	−0.0193	0.9363	−0.1468	2.9583
NBI adults	−0.0193	0.8888	−0.0842	3.1905
NBI elderly	−0.0088	0.8791	−0.0357	3.3395
GPO children	−0.0010	1.0467	−0.0885	2.8152
GPO adults	−0.0140	0.9816	0.0196	3.4537
GPO elderly	−0.0093	0.9438	0.0892	3.8500

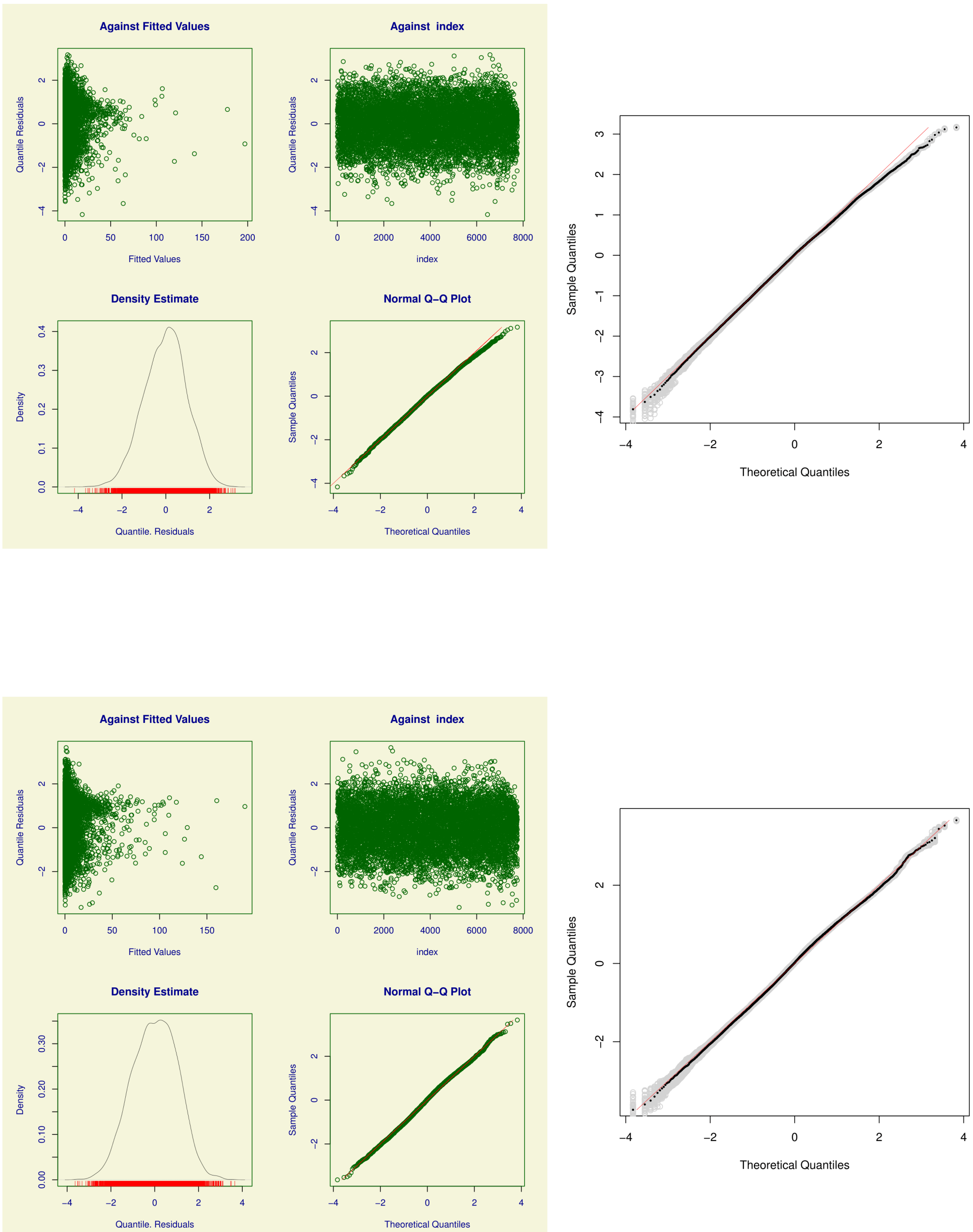


Figure 13: Model diagnostics based on the randomised quantile residuals (left panels) and QQ-plot of the median of 40 realisations of the randomised quantile residuals (right panels) for the NBI GAMLSS model (upper panels) and GPO GAMLSS model (lower panels) for children.

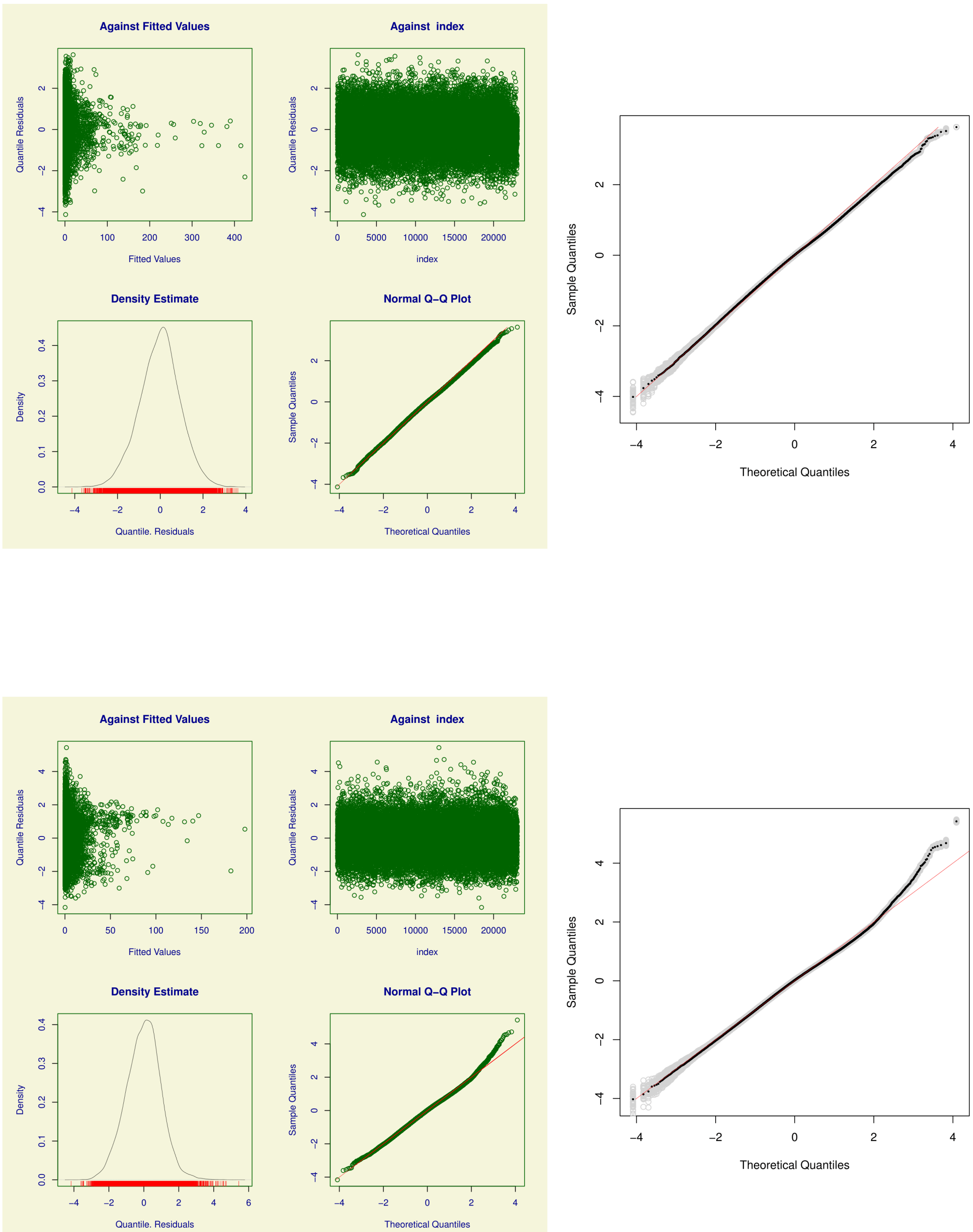


Figure 14: Model diagnostics based on the randomised quantile residuals (left panels) and QQ-plot of the median of 40 realisations of the randomised quantile residuals (right panels) for the NBI GAMLSS model (upper panels) and GPO GAMLSS model (lower panels) for adults.

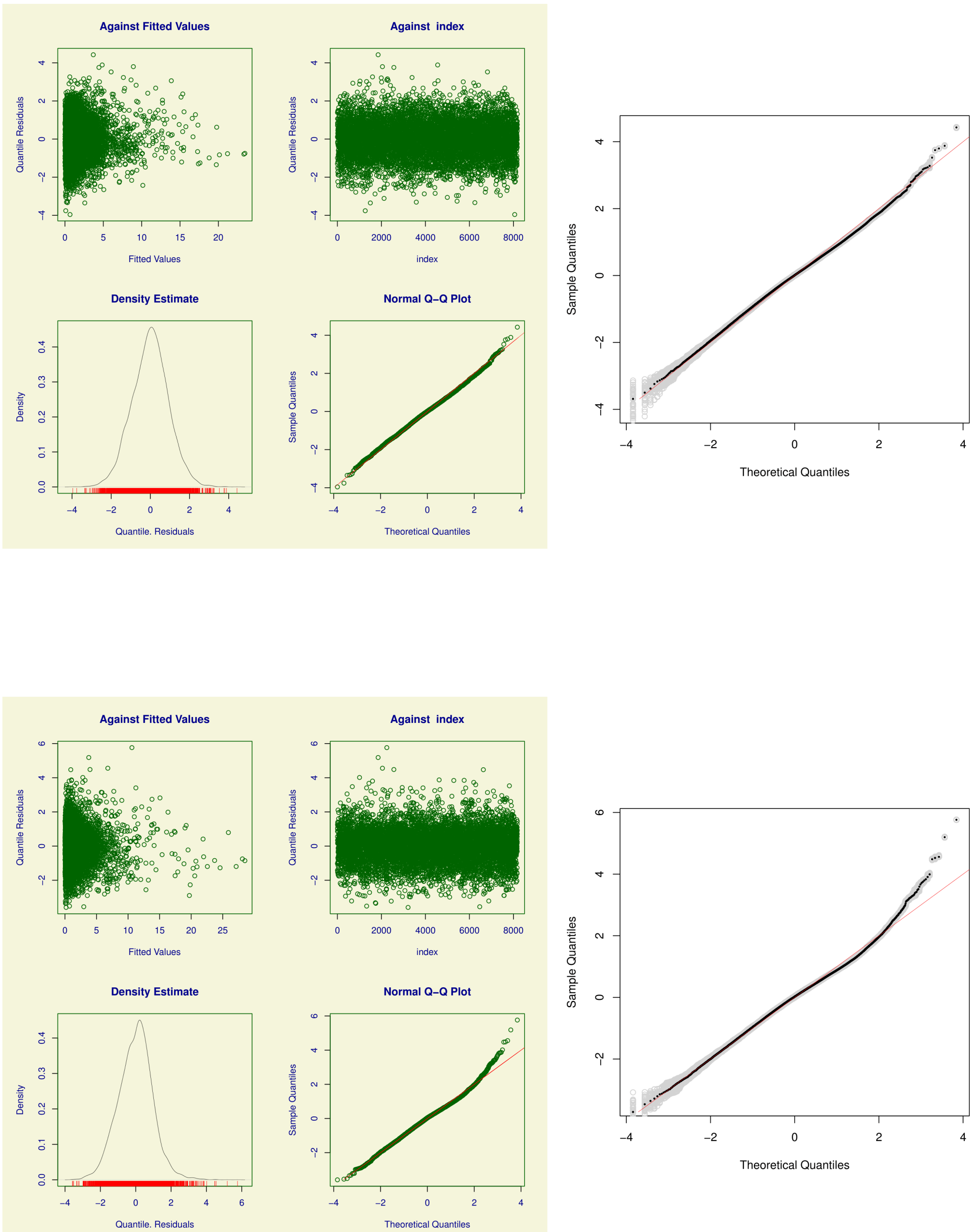


Figure 15: Model diagnostics based on the randomised quantile residuals (left panels) and QQ-plot of the median of 40 realisations of the randomised quantile residuals (right panels) for the NBI GAMLSS model (upper panels) and GPO GAMLSS model (lower panels) for elderly.

A.5 Coefficients of Hurdle models

Table 20: Coefficients for Hurdle 1 model for elderly.

Parameter	Estimate	SE	p-value
Intercept	−0.711330	0.370124	0.055
Vlaams Gewest	−0.022986	0.313367	0.942
Waals Gewest	−0.452020	0.324909	0.164
Holiday Yes	0.114661	0.063356	0.070
Vacc Yes	0.444576	0.196907	0.024
Elevated risk Yes	−0.144448	0.136176	0.289
Face mask Yes	1.725492	0.159138	< 0.001
Male	−0.392107	0.172636	0.023
hh size 2	−0.411383	0.219339	0.061
hh size 3+	−1.101318	0.467023	0.018
2 waves	0.492847	0.163078	0.003
3 waves	0.127269	0.167397	0.447
4 waves	0.124492	0.169102	0.462
5 waves	−0.237321	0.170783	0.165
6 waves	−0.123039	0.173370	0.478
7 waves	0.006912	0.178130	0.969
8+ waves	−0.193477	0.137896	0.161
Elevated risk Yes : Face mask Yes	0.345069	0.134543	0.010
Vacc Yes : Face mask Yes	−0.415289	0.161142	0.010
Vacc Yes : hh size 2	0.394946	0.177694	0.026
Vacc Yes : hh size 3+	0.797104	0.374711	0.033

Table 21: Coefficients for Hurdle 1 model for children.

Parameter	Estimate	SE	p-value
Intercept	−0.46670	0.27705	0.092
Vlaams Gewest	1.41299	0.22347	< 0.001
Waals Gewest	0.15099	0.22712	0.506
Holiday Yes	−0.37373	0.07377	< 0.001
Weekend	−0.17913	0.09164	0.051
hh size 3	−0.48692	0.22348	0.029
hh size 4+	−0.58241	0.21387	0.006
Face mask Yes	0.75331	0.07403	< 0.001
2 waves	−0.02253	0.12434	0.856
3 waves	−0.36051	0.13314	0.007
4 waves	−0.29258	0.13949	0.036
5 waves	−0.55091	0.14679	< 0.001
6 waves	−0.73604	0.15240	< 0.001
7 waves	−0.72873	0.15755	< 0.001
8+ waves	−1.03180	0.10426	< 0.001
Holiday Yes : Weekend	0.29808	0.15172	0.049

Table 22: Coefficients for Hurdle 1 model for adults.

Parameter	Estimate	SE	p-value
Intercept	-1.32878	0.20512	< 0.001
Vlaams Gewest	0.84838	0.15921	< 0.001
Waals Gewest	0.20921	0.16720	0.211
Holiday Yes	0.13139	0.03938	< 0.001
Weekend	0.08638	0.04493	0.055
Low education	-0.28140	0.14003	0.044
High education	0.35113	0.10279	< 0.001
hh size 2	-0.48526	0.12423	< 0.001
hh size 3	-0.45668	0.15349	0.003
hh size 4+	-0.37944	0.15683	0.016
Vacc Yes	0.30219	0.11017	0.006
Elevated risk Yes	-0.01198	0.11992	0.920
Face mask Yes	1.71873	0.07515	< 0.001
Symptoms Yes	-0.20915	0.05117	< 0.001
Male	-0.28911	0.09443	0.002
2 waves	-0.16487	0.08440	0.051
3 waves	-0.32955	0.09055	< 0.001
4 waves	-0.51955	0.09451	< 0.001
5 waves	-0.50286	0.09712	< 0.001
6 waves	-0.42750	0.09993	< 0.001
7 waves	-0.54417	0.10211	< 0.001
8+ waves	-0.70912	0.07438	0.001
Vacc Yes : hh size 2	0.28607	0.10838	0.008
Vacc Yes : hh size 3	0.13037	0.13215	0.324
Vacc Yes : hh size 4+	-0.27181	0.13978	0.052
Elevated risk Yes : Face mask Yes	0.32328	0.09742	< 0.001
Vacc Yes : Elevated risk Yes	-0.20077	0.09793	0.040
Vacc Yes : Face mask Yes	-0.17285	0.08603	0.045

Table 23: Coefficients for Hurdle 2 model for elderly.

Parameter	Estimate	SE	p-value
Intercept	0.04136	0.31422	0.895
Vlaams Gewest	−0.19596	0.29991	0.514
Waals Gewest	−0.31952	0.31388	0.309
Holiday	0.12024	0.05137	0.019
Weekend	0.15602	0.05131	0.002
Vacc Yes	0.64438	0.08850	< 0.001
Elevated risk Yes	−0.01671	0.08213	0.839
Face mask Yes	−0.34694	0.25826	0.179
Symptoms Yes	−0.15582	0.08292	0.060
2 waves	0.16023	0.13101	0.221
3 waves	−0.30736	0.14189	0.030
4 waves	0.02683	0.13849	0.846
5 waves	−0.14388	0.14363	0.316
6 waves	−0.20608	0.14566	0.157
7 waves	−0.24968	0.14726	0.090
8+ waves	−0.35092	0.11279	0.002
Vlaams Gewest : Face mask Yes	0.66708	0.26995	0.013
Waals Gewest : Face mask Yes	0.41927	0.28472	0.141

Table 24: Coefficients for Hurdle 2 model for children.

Parameter	Estimate	SE	p-value
Intercept	1.12475	0.27007	< 0.001
Vlaams Gewest	0.74603	0.22098	< 0.001
Waals Gewest	0.08782	0.23125	0.704
Holiday	−0.47659	0.06766	< 0.001
Weekend	−0.15833	0.08241	0.055
hh size 3	0.54145	0.20310	0.008
hh size 4+	0.35535	0.19210	0.064
Face mask Yes	0.20427	0.07894	0.010
2 waves	0.04055	0.12141	0.738
3 waves	−0.32295	0.13182	0.014
4 waves	−0.24859	0.13532	0.066
5 waves	−0.29275	0.14991	0.051
6 waves	−0.62410	0.15497	< 0.001
7 waves	−0.45236	0.16261	0.005
8+ waves	−0.70276	0.10692	< 0.001

Table 25: Coefficients for Hurdle 2 model for adults.

Parameter	Estimate	SE	p-value
Intercept	−0.06260	0.26672	0.814
Vlaams Gewest	0.29229	0.20044	0.145
Waals Gewest	0.05189	0.20938	0.804
Holiday	−0.03216	0.12044	0.789
Weekend	0.10211	0.03723	0.006
Employed	0.53032	0.07284	< 0.001
Student	0.52555	0.11928	< 0.001
Low education	0.06264	0.14373	0.663
High education	−0.33020	0.12661	0.009
hh size 2	0.16541	0.07931	0.037
hh size 3	0.20464	0.09610	0.033
hh size 4+	0.09555	0.10138	0.346
Vacc Yes	0.14983	0.14532	0.303
Elevated risk Yes	−0.14148	0.05996	0.018
Face mask Yes	0.43411	0.18120	0.017
Symptoms Yes	−0.06427	0.03972	0.106
Male	0.02292	0.06773	0.735
Social group 3&4	0.17342	0.13710	0.206
Social group 5&6	0.01389	0.16641	0.933
Social group 7&8	−0.11730	0.16021	0.464
2 waves	−0.14855	0.06725	0.027
3 waves	−0.28487	0.07304	< 0.001
4 waves	−0.39504	0.07653	< 0.001
5 waves	−0.46811	0.07918	< 0.001
6 waves	−0.48263	0.08077	< 0.001
7 waves	−0.72389	0.08528	< 0.001
8+ waves	−0.61347	0.05928	< 0.001
Vacc Yes : Low education	−0.01603	0.14097	0.909
Vacc Yes : High education	0.54810	0.11020	< 0.001
Vacc Yes : Face mask Yes	−0.34117	0.08498	< 0.001
Face mask Yes : Vlaams Gewest	0.30301	0.17912	0.091
Face mask Yes : Waals Gewest	−0.08585	0.18772	0.647
Holiday Yes : Vlaams Gewest	−0.04223	0.12654	0.739
Holiday Yes : Waals Gewest	0.23180	0.13694	0.091
Vacc Yes : Social group 3&4	0.09113	0.11660	0.434
Vacc Yes : Social group 5&6	0.28803	0.14301	0.044
Vacc Yes : Social group 7&8	0.39120	0.13484	0.004

A.6 Model diagnostic plots for Hurdle models

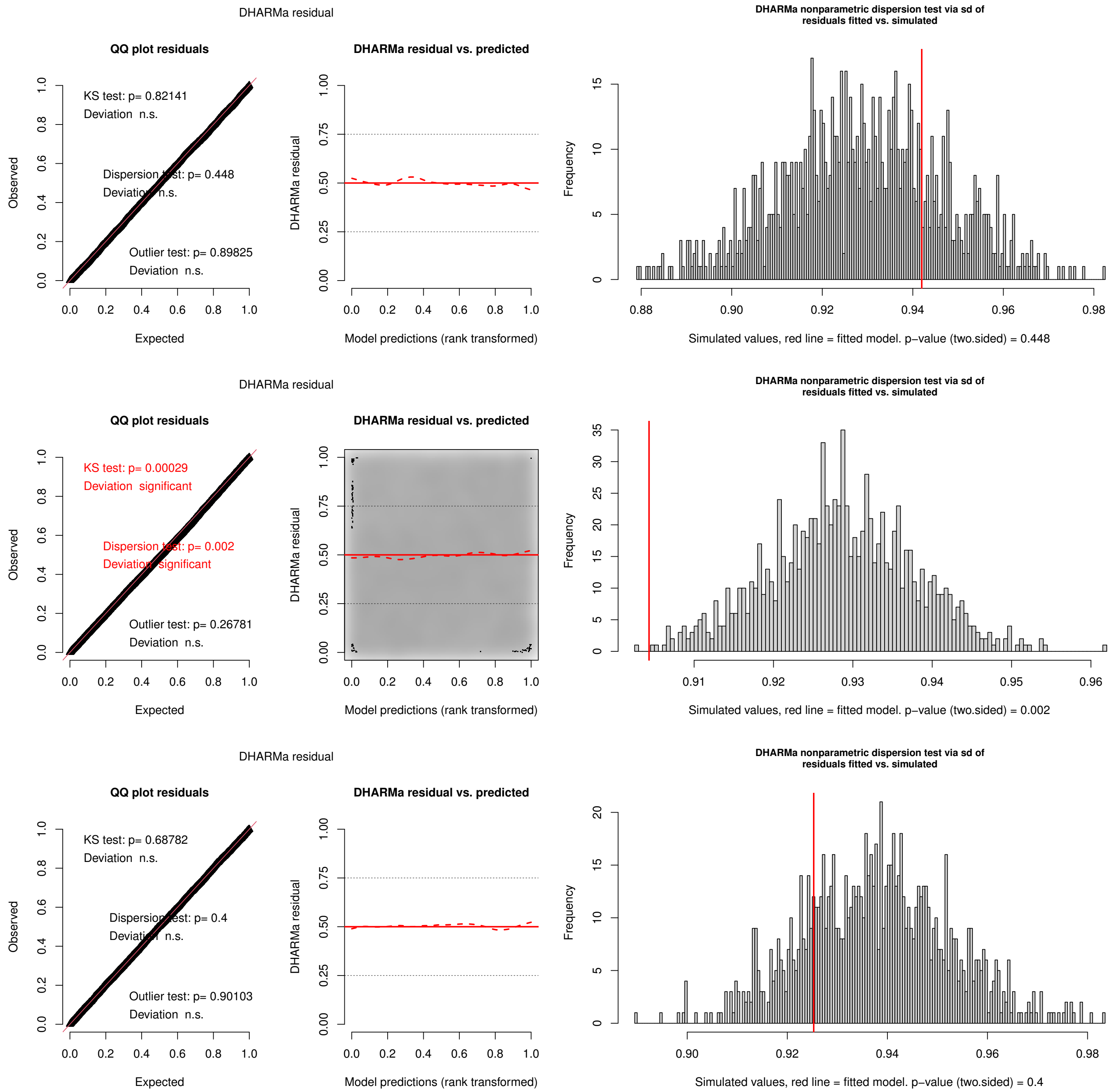


Figure 16: DHARMa residual plots (left) and nonparametric dispersion test (right) for model diagnostics of Hurdle 1 models for children (top panels), adults (middle panels) and elderly (bottom panels).

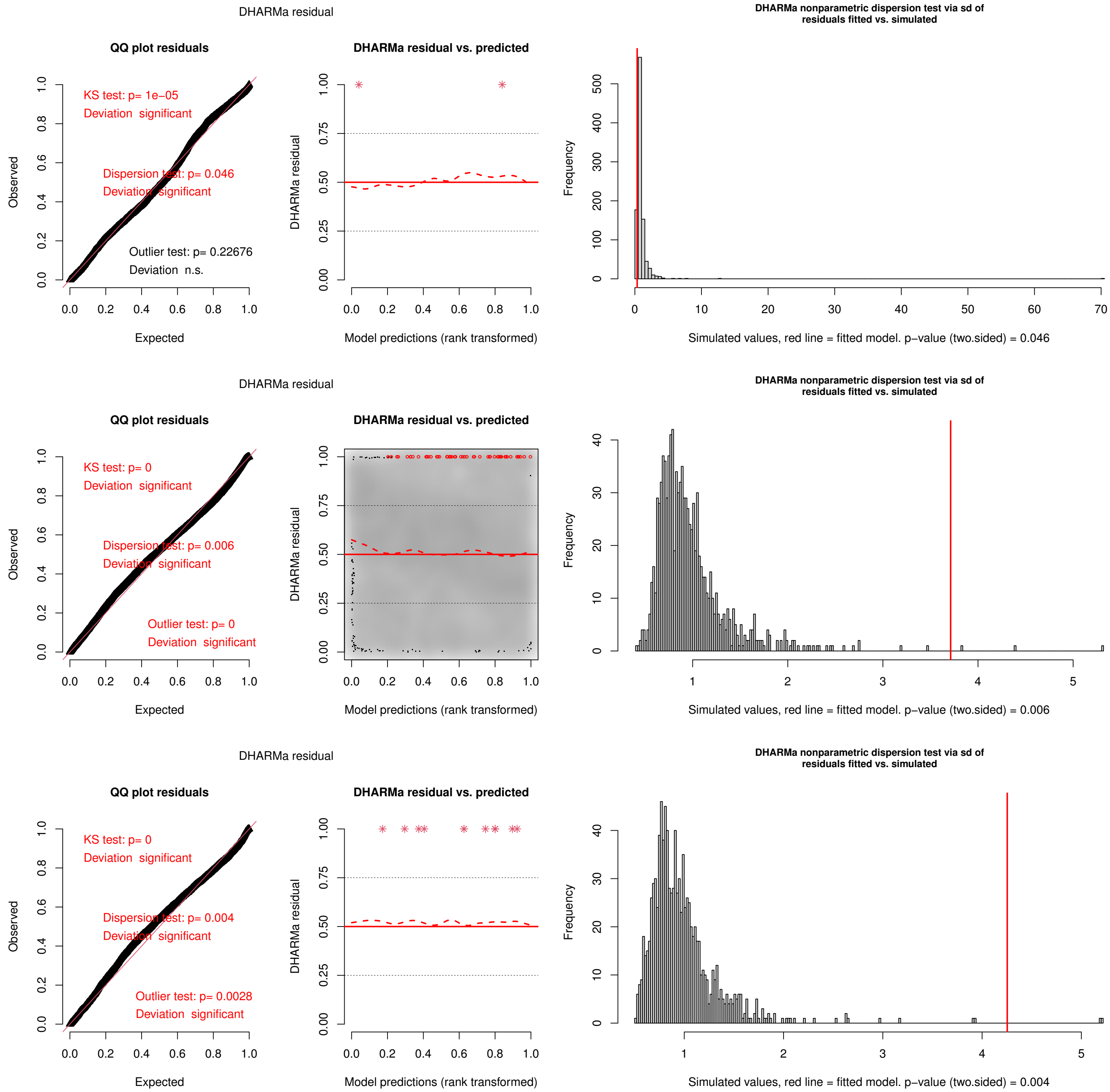


Figure 17: DHARMa residual plots (left) and nonparametric dispersion test (right) for model diagnostics of Hurdle 2 models for children (top panels), adults (middle panels) and elderly (bottom panels).