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

UHASSELT
KNOWLEDGE IN ACTION

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

UHASSELT
KNOWLEDGE IN ACTION

Maastricht University

Faculty of Sciences

School for Information Technology

Master of Statistics and Data Science

Master's thesis

The differential effect of socio-economic status on social contacts during and after COVID-19 lockdown in Belgium.

Bryan Sumalinab

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

SUPERVISOR :
dr. Pietro COLETTI

2020
2021

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

Master's thesis

The differential effect of socio-economic status on social contacts during and after COVID-19 lockdown in Belgium.

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

SUPERVISOR :
dr. Pietro COLETTI

Acknowledgements

The author would like to express his wholehearted gratitude and appreciation to:

The VLIR-UOS for the scholarship grant throughout his masters' program at University of Hasselt.

Dr. Pietro Coletti, his thesis supervisor, for the advice, recommendations, guidance, insightful suggestions and comments aided in the realization and improvement of this paper. He has inspired and motivated the author in his pursuit as an aspiring researcher.

The faculty and staff of the University of Hasselt's Center for Statistics (CenStat) for their support and imparting their knowledge.

The MSU-IIT, where he is currently employed, for their help and support.

His friends Paulo, Roxanne, Neilshan, Paul, Huyen, Hafsah, Esmee, Joshline, Mel, MJ, Luigi, Donna, Malai, Marsha, Admire, Rein, JV and special thanks to Ma'am Chella.

Dubhe Joy, for the love and happiness they have shared and for accompanying him in times of difficulties and worries.

His parents, Vicente and Lydia Sumalinab, and his siblings Kathybeth, Kristine Mae, Vincent Lloyd, Krismarie, Vince Lawrence and Krizza Mae who are the main source of his inspiration.

His CFC - Youth for Christ family for their spiritual guidance.

Above all, GOD ALMIGHTY, the most merciful and most compassionate, for continuously showering blessings, wisdom, courage, guidance, protection, good health and providing the opportunities to realize this work.

Abstract

Background: Social contacts, or specifically closed contact with infected individuals, play a significant role in transmitting respiratory viruses such as the coronavirus. During the COVID-19 pandemic, different countries have implemented several non-pharmaceutical interventions to minimize the virus transmission. These measures have been found to reduce social contacts and decrease disease transmission. However, there are growing concerns regarding the inequalities brought by the pandemic due to socioeconomic disparities.

Objectives: The main objective of this study is to investigate the differential effect of the socioeconomic status on the number of contacts during and after lockdown.

Methodology: The dataset was obtained from a representative survey of the adult population (18+ years) in Belgium, containing information about the participants' characteristics and contact behaviour. The Hurdle Negative Binomial mixed models were used to achieve the study's objective with the number of contacts as the outcome variable and the socioeconomic and other factors as the covariates.

Result: Results showed that the probability of making contacts is higher for those with at least two household members than for those who live alone. People in Flanders are more likely to make contacts than those living in Brussels and Wallonia. In addition, people having higher incomes have a higher chance of making contacts than those with lower incomes after the lockdown. The number of contacts generally increases with the household size, with the effect more pronounced during the lockdown period. Moreover, those in Flanders have more social contacts than in Brussels and Wallonia, and these differences are more significant after the lockdown. Furthermore, the students (and unemployed) made more (less) contact during the re-opening period than employed individuals. After the lockdown, full-time parents/homemakers and retired individuals have fewer contacts compared to those employed.

Conclusion: We've shown how social contacts differed among various groups in the population during and after the lockdown. This can help enhance the implementation and designing of non-pharmaceutical interventions that are equitable to all social groups and improve communication and targeting efforts.

Keywords: Social contacts, Lockdown, Socioeconomic status, Hurdle model, Generalized Linear Mixed Model

Contents

1 Introduction

1.1 Background of the Study

COVID-19 remains a major global health challenge since its initial outbreak on December 31, 2019, where a cluster of pneumonia cases was reported in Wuhan, Hubei Province, China. The novel coronavirus was eventually identified as the causative agent of the outbreak and was named coronavirus disease 2019 (COVID-19) by the World Health Organization (WHO) [1]. As of April 22, 2021, there is a total of approximately 142 million reported cases already, including more than 3 million fatalities worldwide [2]. These numbers are evident of the devastating impact of this pandemic, including the different burdens that it brings to the economy [3], education [4], travel and tourism [5], mental health [6], and more. Nicola et al. [7] discussed a detailed review of the economic implications of COVID-19, which focused on the primary sector (industries involved in the extraction of raw materials), secondary sector (manufacturing industries), and tertiary sector (e.g., education, healthcare, tourism, etc.). For instance, the adverse effect in the agricultural sector is that the price of commodities decreased by 20% because of the drop in bulk demands due to closures of hotels and restaurants. In the manufacturing industry, the negative impact of the pandemic is mainly on the disruption of supply chains and workforce losses due to isolation measures. Moreover, in the tertiary sector, some of the effects are the lack of access to online classes for low-income families, high risk for healthcare workers, and economic losses implied by the hardest-hit tourism sector due to travel restrictions. With such devastating effects worldwide, designing appropriate measures to limit the virus from spreading and eventually bring the pandemic to an end is, therefore, a serious concern.

The virus is transmitted through respiratory droplets primarily by close contact with infected individuals or indirect contact through surfaces or objects used by a person infected with the virus [8]. Hence, the transmission potential of respiratory-spread agents like this can be estimated using some proxy measures (e.g., face-to-face conversation) obtained through surveys [9]. The disease transmission rates between groups (e.g., people of different ages) in the population can be described by the so-called “Who Acquires Infection From Whom” (WAIFW) matrix [10]. To utilize information from the social contact surveys, the age-specific transmission rate is assumed to be proportional to the social contact rates, which is known as the “social contact hypothesis”. The contact rates are derived from the social contact matrix containing information about the average number of contacts made by individual from different age classes [9]. Although we have available information on social contacts from pre-pandemic surveys, this might differ from those collected during the pandemic. That is, people may drastically reduce their contacts during a pandemic, to the point that they make no contact at all. Thus, understanding the social contact patterns of individuals can help identify subgroups in the population at greater risk of the disease and inform epidemiological models by using empirical data on mixing patterns [11, 12]. Consequently, it is essential for curbing this type of pandemic and will support policy and decision-makers to decide for appropriate measures that will help control the pandemic. Moreover, it can also help to evaluate the impact of the different interventions implemented on social contacts and assess people’s compliance to such measures.

A systematic review of the pre-pandemic social contact surveys was done by Hoang et al. [11] in which the top three determinants of social contacts that were identified are age, weekday/weekend, and household size. During the COVID-19 pandemic, several studies [12–19] about social contacts were conducted. All of these studies have concurred that the number of social contacts was reduced following the implementation of non-pharmaceutical interventions such as physical distancing and lockdown measures. Moreover, a rapid review of social contact studies during the COVID-19 pandemic reported a 65% - 87% average reduction in social contacts from pre-pandemic setting, lowering the reproduction number below one in several countries [20]. Since the early stages of the pandemic and up to the present, different countries have implemented several non-pharmaceutical intervention strategies [21, 22] to control the spread of the disease, while vaccines are partially utilized. These measures include different quarantine guidelines, lockdown, social distancing, travel ban, closure of schools and businesses, mask-wearing, limiting social bubbles, etc. Several studies have shown that these control measures are effective in reducing disease transmission and risk of infection [17–19, 23–26].

Despite the effectiveness of the lockdown measures in preventing the virus from spreading, there are concerns about the inequalities that it brings as a result of socioeconomic disparities. In particular, socioeconomically deprived groups tend to have high risk of exposure to COVID-19 [27]. According to studies in the United States [28] and Europe [29], the pandemic has a greater effect on women and people with lower levels of education. That is, they are more likely to experience economic instability as a result of the pandemic, such as job losses, financial, food, and housing insecurity. Also, studies in the United Kingdom showed that individuals of lower income classes or socioeconomic positions have faced more financial hardship and adverse events induced by the COVID-19 crisis [30, 31]. Meanwhile, areas in England [32] and France [33] with higher socioeconomic status have greater reductions in mobility during the lockdown period than areas with much lower socioeconomic levels. This may be due to the fact that low-wage employees have works that need to be performed in person even during the lockdown, while higher-paying jobs can be done from home [33].

In Belgium, interventions started last March 13, 2020, that included prohibiting public gatherings and closure of schools, shops offering non-essential services, and hospitality industries [16]. To assess the impact of and compliance to these lockdown strategies, a survey (called CoMix survey) was conducted every two weeks from April 24, 2020 to present, with a stop during August - November 2020. It's a representative survey of the adult population (18+ years) in Belgium, which contains information about their contact behaviour. The first eight waves of the survey were analysed by Coletti et al. [16] by comparing social contact changes from the pre-pandemic to the COVID-19 pandemic scenario and assessing compliance to some measures such as face-mask wearing and social distancing. This work will investigate the differential effect of socioeconomic status on the number of contacts throughout the different intervention strategies as this was not done in the paper of Coletti et al [16]. This will allow us to look at how lockdown measures affect the different strata of the socioeconomic population. This is also relevant to the previously discussed unbalanced effect of the pandemic. In doing so, we expect to gain a better understanding of the true impact at the societal level, which will help us to design better intervention measures and improve communication and targeting efforts.

1.2 Objectives of the Study

The study's main objective is to investigate the differential effect of the socioeconomic status on the number of contacts. Specifically, modelling will be done with the number of contacts as the outcome variable and the socioeconomic and other factors as the covariates using multilevel Hurdle count models. The analysis will be stratified according to the group of survey waves, such that each group is assumed to be homogeneous in terms of non-pharmaceutical interventions. Moreover, we want to see how lockdown strategies affect the different socioeconomic groups in order to improve the implementation and designing of non-pharmaceutical interventions.

2 Methodology

This section describes the different models used in this paper. Previous social contact studies [34, 35] have used the negative binomial regression to model the number of contacts. Similar methodology was also used by some studies [12, 18] during the COVID-19 pandemic. Here, we start with the usual approach to model count data which is the *Poisson regression*. Then, the *negative binomial regression* is introduced to account for the overdispersion issue in Poisson regression. Moreover, since this survey was done during the COVID-19 pandemic, more people tend to report zero contacts than usual, especially during the lockdown period. Hence, a *Hurdle model* is also proposed to address these excess zeros. Furthermore, since the CoMix survey is a panel survey where we expect correlated responses from the same subject, a *generalized linear mixed model* (GLMM) is used to account for this correlation. More details about (Hurdle) count models and GLMM can be found in [37] and [38], respectively.

2.1 Poisson Regression

Poisson regression is the most commonly used method when modelling data with count responses. For a given vector of covariates $\mathbf{X}_i$, the count response y_i is assumed to follow a Poisson distribution given below,

$$f(y_i|\mathbf{X}_i) = \frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!}, \qquad y_i = 0, 1, 2, \ldots \tag{1}$$

where the linear predictor $\mathbf{X}_i^{'}\boldsymbol{\beta}$ is related to the parameter μ_i through a *link* function. The *log-link* function is often used such that $\log(\mu_i) = \mathbf{X}_i^{'}\boldsymbol{\beta}$. For random samples y_i, $i = 1, .., n$, the log-likelihood function can then be written as

$$\mathcal{L} = \sum_{i=1}^{n} \left(y_i \ln \mu_i - \mu_i - \ln y_i! \right).$$

Moreover, for Poisson regression, the conditional mean is equal to the conditional variance, such that

$$E(y_i|X_i) = Var(y_i|X_i) = \mu_i = \exp\left(\mathbf{X}_i^{'}\boldsymbol{\beta}\right).$$

This property of equality for the conditional mean and variance is called *equidispersion*. This assumption, however, is mostly not satisfied in real-life data sets. Hence, another phenomenon that could happen is when we have greater variability than what we would expect from our model, or equivalently, the variance is greater than the mean. This is termed as *overdispersion*. The most common reasons for overdispersion are correlated responses and unobserved predictors or omitting important explanatory variables. Failing to account for overdispersion would obviously violate our distributional assumption for Poisson regression. It will also cause underestimating the standard error of the estimates where we might detect the statistical significance of explanatory variable(s) by chance [36, 37]. One way to handle overdispersion as an alternative to Poisson regression is by using Negative Binomial Regression, which will be discussed next.

2.2 Negative Binomial Regression

As previously mentioned, overdispersion is a common issue in datasets. *Negative Binomial regression* is one of the candidates to address this issue. There are several parameterizations of a negative binomial, and the one presented here is just one of those. Let Y be a negative binomial distributed random variable. Then the probbaility distribution of Y is given by

$$f(y|p,r) = \binom{y+r-1}{r-1} p^r (1-p)^y, \qquad y = 0, 1, 2, \ldots \tag{2}$$

For independent Bernoulli trials with success probability p, the negative binomial can be used to model the number of failures (y) before obtaining r successes. From (2), it can be shown that the mean and variance of Y is given by

$$E(Y) = \mu = \frac{r(1-p)}{p} \qquad \text{and} \qquad Var(Y) = \frac{r(1-p)}{p^2}.$$

The variance of Y can further be re-written as

$$Var(Y) = \mu + \alpha\mu^2$$

where $\alpha = \frac{1}{r} > 0$. Hence, it can be seen that the variance of a negative binomial distributed random variable is greater than its mean which makes it to be a potential candidate to address the problem of overdispersion in Poisson regression.

2.2.1 Poisson-Gamma Mixture

Another formulation of negative binomial can be derived as a mixture of Poisson and Gamma distribution. This is done by formulating the distribution of random variable Y, for each observation i given by

$$f(y_i|u_i) = \frac{(-\lambda_i u_i)^y e^{-\lambda_i u_i}}{y_i!}, \qquad y = 0, 1, 2, \ldots$$

or equivalently, by modifying the Poisson regression model as

$$E(y_i|\mathbf{X}_i, u_i) = \lambda_i u_i = e^{\mathbf{X}_i' \boldsymbol{\beta} + \varepsilon_i}$$

with additional heterogeneity term $u_i = e^{\varepsilon_i}$, independent of the covariates, that accounts for overdispersed Poisson model. Given u_i and vector of covariates $\mathbf{X}_i$, y_i is Poisson distributed with conditional mean and variance equals to $\lambda_i u_i$.

Suppose that $g(u_i)$ is the probability density function of u_i. Then the unconditional distribution $f(y_i|\mathbf{X}_i)$ can be derived by integrating $f(y_i|\mathbf{X}_i, u_i)g(u_i)$ over u_i, that is,

$$f(y_i|\mathbf{X}_i) = \int_0^\infty f(y_i|\mathbf{X}_i, u_i)g(u_i)d(u_i). \tag{3}$$

In this case, the heterogeneity parameter u_i is assumed to follow a gamma (θ, θ) distribution with mean and variance, 1 and $\frac{1}{\theta}$, respectively. The density function of u_i is given by

$$g(u_i) = \frac{\theta^\theta}{\Gamma(\theta)} u_i^{\theta-1} e^{-\theta u_i}$$

where $\theta > 0$ and $\Gamma(\cdot)$ is a gamma function. From this, it can be shown that the analytical solution for the integral in equation (3) is given by

$$f(y_i|\mathbf{X}_i) = \binom{y_i + \frac{1}{\alpha} - 1}{\frac{1}{\alpha} - 1} \left(\frac{1}{1+\alpha\mu_i}\right)^{\frac{1}{\alpha}} \left(\frac{\alpha\mu_i}{1+\alpha\mu_i}\right)^{y_i} \tag{4}$$

where $\alpha = \frac{1}{\theta} > 0$ is called the *overdispersion parameter.* Equation (4) above is a reparameterization of the negative binomial distribution which is equivalent to (2) with $p = \frac{1}{1+\alpha\mu_i}$ and $r = \frac{1}{\alpha}$. Moreover we have,

$$E(y_i|\mathbf{X}_i) = \mu_i = e^{\mathbf{X}_i\boldsymbol{\beta}} \qquad \text{and} \qquad Var(y_i|\mathbf{X}_i) = \mu_i + \alpha\mu_i^2.$$

Maximum likelihood estimation method can be used to estimate the parameters of the negative binomial regression with the log-likelihood function given by

$$\mathcal{L} = \sum_{i=1}^{n} y_i \ln\left(\frac{\alpha\mu_i}{1+\alpha\mu_i}\right) - \frac{1}{\alpha}\ln(1+\alpha\mu_i) + \ln\Gamma\left(y_i + \frac{1}{\alpha}\right) - \ln\Gamma(y_i+1) - \ln\Gamma\left(\frac{1}{\alpha}\right).$$

2.3 Hurdle Model

Hurdle models can be used when we have an excess count of zeros than what we expect from our underlying distributional assumption (Poisson, Negative Binomial). These extra zeros can also be a source of overdispersion. It is a mixture of two processes where it assume that the zeros and positive counts are generated by two separate processes: the *binary* and the *count* process, respectively. The hurdle model gets its name from the fact that non-zero counts are only generated when a zero barrier/hurdle is crossed [37].

A logit model is often used for the binary process, while the count process is estimated using a zero-truncated count model. The poisson and negative binomial are altered in the zero-truncated model to exclude the probability of zero counts such that the probability mass function still sum to one. For example, the density of zero-truncated poisson model is given by

$$P(Y = y|y > 0) = \frac{f(y)}{1 - f(0)}, \qquad y = 1, 2, 3, \ldots \tag{5}$$

where $f(\cdot)$ is given in (1).

For the ***Hurdle Poisson*** model, the probability distribution of a count response Y, can be written as

$$f(y) = \begin{cases} \pi, & y = 0 \\ (1-\pi)\left(\frac{\lambda^y e^{-\lambda}}{y!}\right)\left(\frac{1}{1-e^{-\lambda}}\right), & y = 1, 2, 3\ldots \end{cases}$$

where π is the probability of observing zero and can be estimated using a binary model. For example, using a logit model, we have $\pi = \frac{1}{1+e^{z\gamma}}$ where z is the binary covariate with the coeffecient γ. Moreover, the parameter λ is estimated using the Poisson model such that $\lambda = \exp{(X\beta)}$ with $X\beta$ as the linear predictor. The expression $\frac{1}{1-e^{-\lambda}}$ for the count component is a result of truncating the Poisson distribution in (5) where $f(0) = e^{-\lambda}$.

Similar formulation can be obtained for the ***Hurdle Negative Binomial***, where the probability distribution can be written as,

$$f(y) = \begin{cases} \pi, & y = 0 \\ (1-\pi)\binom{y + \frac{1}{\alpha} - 1}{\frac{1}{\alpha} - 1}\left(\frac{1}{1+\alpha\lambda}\right)^{\frac{1}{\alpha}}\left(\frac{\alpha\lambda}{1+\alpha\lambda}\right)^{y}\left(\frac{1}{1-(1+\alpha\lambda)^{-\frac{1}{\alpha}}}\right), & y = 1, 2, 3\ldots \end{cases}$$

where α is a negative binomial dispersion parameter. Similar to hurdle poisson, π and λ are to be estimated from the binary and count component, respectively. Here, the probability of obtaining zero counts for the negative binomial distribution is $f(0) = (1+\alpha\lambda)^{-\frac{1}{\alpha}}$

2.4 Generalized Linear Mixed Model

A *generalized linear mixed model* (GLMM), also called as the *multilevel* or *random-effects model*, can be used to model dataset when observations are correlated due to clustering or repeated measurements from the same subject [38]. The CoMix survey is a type of longitudinal survey where participants were asked to participate in each wave. Hence, GLMM can be considered in modelling this data. Suppose we have measurements Y_{ij} of *ith* subject, $i = 1, \ldots, n$ and $j = 1, \ldots, n_i$, where n_i is the number of measurements for each subject i. The general formulation of GLMM is given by,

$$\eta(\mu_{ij}) = \eta(E(Y_{ij}|\mathbf{b}_i) = \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{b}_i \tag{6}$$

where $\eta(\cdot)$ is a link function of a generalized linear model, $\mathbf{x}'_{ij}$ and $\mathbf{z}'_{ij}$ are vector of known covariates, $\boldsymbol{\beta}$ a vector of unknown fixed regression coefficients, and $\mathbf{b}_i$ are called the random-effects assumed to follow a normal distribution with mean zero and variance covariance matrix $\mathbf{D}$, that is, $\mathbf{b}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{D})$.

Maximum likelihood estimation can then be used to estimate the parameters by maximizing the marginal likelihood (integrating out the random effects $\mathbf{b}_i$),

$$\mathcal{L}(\boldsymbol{\beta}, \mathbf{D}) = \prod_{i=1}^{n} \int \prod_{j=1}^{n_i} f(y_{ij}|\boldsymbol{\beta}, \mathbf{b}_i) f(\mathbf{b}_i|\mathbf{D})\, d\mathbf{b}_i \tag{7}$$

where $f(y_{ij})$ and $f(\mathbf{b}_i)$ are the probability distributions of the responses y_{ij} and random-effects $\mathbf{b}_i$, respectively. Most often, especially for non-continuous outcomes, there's no analytical solution for the integral in equation (7). Hence, numerical integration techniques (e.g., Gaussian quadrature) are needed to evaluate/approximate the integral.

For ***Poisson*** and ***Negative Binomial GLMM*** used in this paper, we consider a *random-intercept model* as a special case of (6) given by,

$$\eta(E(Y_{ij}|u_i)) = \log(\mu_{ij}) = \mathbf{x}'_{ij}\boldsymbol{\beta} + u_i \tag{8}$$

where $u_i \sim \mathcal{N}(0, \sigma^2)$ is a subject-specific random-intercept that accounts for between-subject variability or equivalently accounting for correlated responses from the same subject.

As introduced in [39] the ***Hurdle Poisson and Negative Binomial GLMM*** can be written as,

$$\log(\lambda_{ij}) = \mathbf{x}'_{ij}\boldsymbol{\beta} + v_i \tag{9}$$

$$\text{logit}(\pi_{ij}) = \mathbf{z}'_{ij}\boldsymbol{\gamma} + w_i \tag{10}$$

where equation (9) and (10) corresponds to the count and binary process of hurdle model with random-intercepts v_i and w_i, respectively. Moreover, the random-intercepts v_i and w_i follows a bivariate normal distribution with mean zero. Here, we assume that v_i and w_i are independent with variances σ_v^2 and σ_w^2, respectively.

2.5 Software

All the analysis were performed using R version 3.6.1 [40], and all of the models were fitted with the R package glmmTMB [41]. The R code is provided in the appendix.

3 Results

3.1 Data

The data was collected from a representative survey (CoMix study) of the adult population (18+ years) in Belgium, which contains information about the participants' number of contacts, demographic and socioeconomic characteristics together with other relevant information related to their attitudes and contact behaviour during the COVID-19 pandemic. The participants were asked to report all the contacts they made between 5 am the day preceding the survey and 5 am the day of the survey. A contact was defined as exchanging at least a few words with anyone the participant met in person, or that involve skin-to-skin contact [16]. It's an on-going survey every two weeks that started last April 24, 2020. Here, we utilized only the 8 waves of survey (until July 30, 2020) and the interest is on the socioeconomic factors as the covariates and the number of social contacts as the outcome variable.

The description of all the variables of interest and the distribution of sample sizes are shown in in the appendix in Table 9 and 10, respectively. The income variable was re-categorized for better interpretability, according to the quartile distribution of income in Belgium as of 2019 [42]. Also, some participants do not report their income for some of the waves in which they participated. Hence, these "missing values" were filled in using the observed value (mode income) from other waves of the same individual. Moreover, the subjects who prefer not to answer (PNTA) about their income remain unchanged and are included as part of the income level in the analysis.

Moreover, we formed three different group of waves, namely:

- Group 1 : Waves 1 and 2 (Dates: April 24 and May 8, 2020)
- Group 2 : Waves 3, 4 and 5 (Dates: May 21, June 4 and June 18, 2020)
- Group 3 : Waves 6, 7 and 8 (Dates: July 2, July 16, and July 30, 2020)

These groups were created on the assumption that the waves belonging to the same group are somewhat similar in terms of non-pharmaceutical intervention. This can be seen in Figure 1 depicting the timeline of interventions imposed in Belgium since March and the day of the survey for each wave. For *group 1*, which consists of waves 1 and 2, we can see in Figure 1 that during these times full lockdown was still in place. In *group 2*, comprising waves 3, 4, and 5, the schools have been re-opened partially. Also, the food and beverage industries and the companies and shops offering non-essential services were already allowed to re-open. Hence, we can think of the second group as partial lockdown. Finally, *group 3*, including waves 6, 7, and 8, is quite similar to group 2 in terms of interventions but with the difference that the third group is during summer holidays where schools are closed and that the border has already been opened.

Figure 1: Timeline of interventions in Belgium and waves of survey

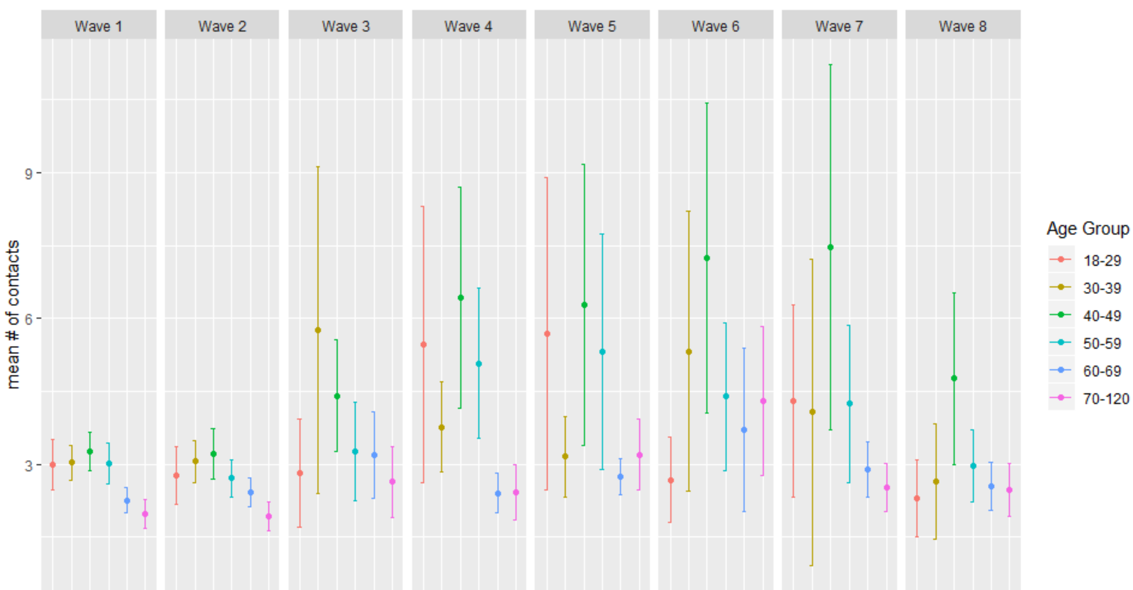

Figure 2: Plot for the average number of contacts by age group per wave with 95% CI

Table 1 provides some summary statistics such as mean and standard deviation (sd) number of contacts for each level of socioeconomic variables. It's worth noting that for group 1, the mean number of contacts appears to be mostly similar across levels of all variables, except for household size, which seems to have considerable differences. This observation is also consistent with the standard deviations for group 1, indicating that the number of contacts appears to be homogeneous during lockdown (waves 1 and 2). For groups 2 and 3, however, differences in the mean number of contacts between levels of variables can be observed. Hence, all of these factors chosen for inclusion as covariates in the model may be able to explain variations in the number of social contacts. Furthermore, groups 2 and 3 show larger standard deviations than group 1, implying more heterogeneity in the number of contacts. Using the age variable as an example, Figure 1 shows that waves 1 and 2 have a lower average number of contacts and smaller confidence intervals. On the other hand, the mean number of contacts is higher for the subsequent waves with wider confidence intervals.

Table 1: Summary statistics for the number of contacts

	Group 1	Group 2	Group 3
	mean(sd) number of contacts		
Age			
18-29	2.9 (2.77)	4.59 (9.05)	3.03 (3.94)
30-39	3.04 (2.36)	4.36 (11.87)	4.09 (11.08)
40-49	3.24 (3.32)	5.59 (12.71)	6.5 (16.29)
50-59	2.88 (2.88)	4.45 (10.29)	3.85 (7.81)
60-69	2.34 (2.13)	2.79 (4.49)	3.08 (7.39)
70-120	1.95 (1.64)	2.75 (3.41)	3.12 (4.9)
Area			
Center (Bruxelles)	2.51 (2.29)	2.84 (6.94)	2.36 (3.65)
North (Flandre)	2.86 (2.82)	4.81 (10.92)	4.98 (12.07)
South (Wallonie)	2.57 (2.38)	2.78 (5.3)	2.46 (4.18)
Day of Survey			
Weekday	2.69 (2.63)	4.15 (9.9)	4.11 (10.26)
Weekend	3.12 (2.79)	3.2 (4.06)	3.74 (8.81)
Educational Level			
primary	2.98 (2.59)	4.78 (12.01)	2.22 (2.03)
l_secondary	2.26 (2.3)	2.72 (3.77)	3.13 (6.07)
h_secondary	2.75 (2.81)	4.19 (9.23)	4.22 (9.21)
higher_ed	2.79 (2.64)	3.58 (4.62)	3.69 (9.78)
bs/ms_degree	2.89 (2.55)	4.82 (14.64)	4.84 (13.47)
phd	3.39 (2.33)	6.58 (10.95)	4.05 (8.06)
Employment Status			
employed	3.18 (2.82)	5.4 (12.49)	5.2 (13.06)
others	2.17 (2.01)	2.6 (3.85)	2.81 (3.74)
self-employed	3.42 (3.79)	5.01 (6.44)	6.72 (19.51)
student	3.69 (5.58)	5 (9.01)	4.92 (8.99)
unemployed	2.04 (1.9)	2.81 (8.89)	2.54 (3.38)
Gender			
female	2.72 (2.69)	3.61 (6.68)	4.37 (11.61)
male	2.76 (2.62)	4.27 (10.6)	3.76 (8.5)
Household Size			
1	1.71 (2.09)	3.6 (13.76)	2.7 (8.51)
2	2.22 (2.39)	3.3 (7.19)	4.65 (12.69)
3	2.88 (2.02)	4.28 (6.14)	3.78 (4.91)
4	3.68 (2.56)	5.2 (6.34)	4.36 (6.29)
5	5.24 (3.5)	5.85 (4.11)	7.04 (9.15)
6	6.88 (3.5)	10.52 (14.32)	6.11 (4.23)
7	8.4 (3.81)	16.83 (19.52)	9 (12.17)
Household Type			
depend_children	3.78 (2.67)	5.16 (6.97)	5.1 (9.71)
no_depend_children	2.33 (2.53)	3.62 (9.87)	3.65 (9.88)
Income			
Q1	1.84 (1.93)	3.09 (8.71)	2.59 (7.34)
Q2	2.38 (2.9)	3.84 (10.89)	3.68 (9.37)
Q3	2.83 (2.7)	4.23 (13.32)	4.09 (9.23)
Q4	3.27 (2.72)	4.69 (7.95)	4.7 (10.47)
PNTA	2.71 (2.41)	3.26 (4.23)	4.16 (11.99)

3.2 Negative Binomial GLMM

The Poisson GLMM (results presented in section 7.2 in the appendix) and Negative Binomial (NB) GLMM were first fitted to account for correlated responses from the same subject. Using Akaike Information Criterion (AIC) as the model fit criteria, it seems that the NB GLMM provides better fit than Poisson GLMM since it has lower AIC for all three groups as shown in Table 7. To formally test for the presence of overdispersion, a boundary likelihood ratio test (LRT) [37] was used since the overdispersion parameter α lies on the parameter space boundary under the null hypothesis ($\alpha = 0$). In particular, we noticed that by formulation, when $\alpha = 0$, the negative binomial reduces to a Poisson distribution. However, α is strictly greater than zero, hence, the assumption of standard chi-square distribution for the LRT statistic under the null hypothesis cannot be assumed. To implement boundary LRT, the p-value of the standard LRT is simply divided by 2. This yields a p-value < 0.0001 for all three groups, indicating the presence of overdispersion, or equivalently, that the negative binomial model is more appropriate for our data.

In order to determine the socioeconomic factors associated with the number of contacts for each group of waves, the LRT was also used and the result is shown in Table 2. It can be seen in the table that the covariates household size and area are highly significant for all group of waves. This means that the number of contacts varies significantly according to the size of the household and the area where the participant lives. The only other significant covariates are employment status and age for group 2 and 3, respectively. When comparing the results to the Poisson GLMM in Table 11 in the appendix, the results for group 1 were similar with the NB GLMM. However, the Poisson GLMM shows that the number of contacts also differs significantly by day of survey and household type for groups 2 and 3. Hence, we have shown that some variables appears to be statistically significant when only using the Poisson model. This demonstrates that without accounting for overdispersion (as addressed by the negative binomial), the standard errors may be deflated, causing statistical significance to be detected by chance [36, 37].

Looking at the parameter estimates in Table 3, an increasing trend in the coefficient of household size variable can be noticed. Hence, given that the reference category is hh_size=1, we can say that the number of contacts increases as the number of people living in the household also increases. Moreover, people living in Flanders (reference category) have higher average number of contacts than those in Brussels and Wallonia with the effect stronger for group 2 and group 3. In terms of age, where it is significant for group 3, it can be seen that only people of ages 30-39 years old have significant difference in the number of contacts with people who are 60-69 years old. Similarly, for group 2, the number of contacts for participants that are employed (reference level) only differs with those that are unemployed and with "other" type of employment.

Table 2: Likelihood Ratio Test for NB GLMM

		Group 1		Group 2		Group 3	
	df	LRT.stat	p-value	LRT.stat	p-value	LRT.stat	p-value
incomeQuartile	4	6.666	0.155	3.943	0.414	2.732	0.604
age	5	5.629	0.344	5.319	0.378	17.195	**0.004**
gender	1	1.729	0.188	0.468	0.494	3.282	0.070
hh_size	6	191.198	**<0.001**	45.327	**<0.001**	44.774	**<0.001**
hh_type	1	0.020	0.888	2.446	0.118	2.742	0.098
educ	5	6.108	0.296	3.277	0.657	5.758	0.331
employ	4	3.026	0.553	17.842	**0.001**	3.686	0.450
day	1	2.416	0.120	1.163	0.281	0.636	0.425
area	2	12.254	**0.002**	53.938	**<0.001**	56.069	**<0.001**

Table 3: Parameter Estimates for NB GLMM

	Group 1	Group 2	Group 3
(Intercept)	0.7280***	1.1105***	0.5521
incomeQuartile2	0.1327	0.0910	0.1719
incomeQuartile3	0.2050*	0.0876	0.1983
incomeQuartile4	0.1796*	0.1978	0.0757
incomeQuartilePNTA	0.1337	0.0406	0.1166
age18-29	-0.1536	-0.2678	-0.3560
age30-39	-0.1050	-0.2650	-0.4624**
age40-49	-0.1462	-0.1831	0.0666
age50-59	-0.0622	-0.1498	-0.1246
age70-120	-0.1197	0.0958	0.0159
gendermale	-0.0549	0.0472	-0.1421
hh_size2	0.2156***	0.1835*	0.5621***
hh_size3	0.5122***	0.3812***	0.5980***
hh_size4	0.7493***	0.5853***	0.5989***
hh_size5	1.0545***	0.7994***	1.1184***
hh_size6	1.3309***	1.2226***	1.0489***
hh_size7	1.5942***	1.7331***	1.4909**
hh_typedepend_children	0.0076	0.1413	0.1825
educl_secondary	-0.2607*	-0.2167	0.1200
educh_secondary	-0.2221	-0.0401	0.2872
educhigher_ed	-0.1558	-0.0495	0.2388
educbs/ms_degree	-0.1642	-0.0029	0.4151
educphd	-0.1396	-0.2081	0.1320
employothers	-0.1106	-0.4570***	-0.2190
employself-employed	0.0391	0.0014	0.1257
employstudent	-0.0163	-0.0138	-0.0801
employunemployed	0.0018	-0.3920*	-0.0340
dayWeekend	0.0847	-0.0691	0.0453
areaCenter (Bruxelles)	-0.1714*	-0.5807***	-0.6249***
areaSouth (Wallonie)	-0.1400**	-0.4962***	-0.5858***
$\hat{\sigma}^2$	0.1569	0.5552	0.5793
$\hat{\alpha}$	0.0675	0.3585	0.4131

*p-value<0.05, **p-value<0.01 , ***p-value<0.001
$\hat{\sigma}^2$ - random intercept variance estimate
$\hat{\alpha}$ - overdispersion parameter estimate

3.3 Hurdle Negative Binomial GLMM

This survey was done during the COVID-19 pandemic, in a period of time in which it is possible that people tend to avoid making contacts to prevent getting infected. In fact, more people were having zero contacts than usual which leads to the issue of excess zeros. The percentage of zeros in our data are 12%, 16% and 17% for group 1, group 2 and group 3, respectively, which seem to suggest that we have extra zeros in the data. To address this, the Hurdle Poisson GLMM (results provided in section 7.3 in the appendix) and Hurdle Negative Binomial (NB) GLMM were fitted. Looking at the AIC in Table 7, the Hurdle models have lower AIC compared to their Poisson and Negative Binomial counterpart, indicating the presence of excess zeros. Moreover, the Hurdle NB GLMM appears to fit better than the Hurdle Poisson GLMM having lower AIC in Table 7, which demonstrates again the evidence of overdispersion in our data.

Table 4 presents the results for the likelihood ratio test for the binary component of the hurdle model. This enables us to determine which subgroups of people are more likely to make zero contacts. The table shows that the household size and area have an effect on whether a person makes contact or not, for all three groups. There is also a difference in the likelihood of making contact between those who have and do not have dependent children. However, this effect is no longer significant in group 3. The additional effects can only be seen in group 2 for age and income. Looking at the parameter estimates in Table 5, we noticed a negative estimates for household size variable. This means that those who live in larger households have a higher probability of making contacts than those who live alone. This is similar to the effect of having dependent children, in which they are more likely to make contacts than those without dependent children, given that the estimates are also negative. Meanwhile, in terms of income, people that receive higher salary have higher chance of making contacts during group 2. Another notable result is that the coefficients of those who prefer not to answer (PNTA) about their income is largely negative similar to the 3rd and 4rth quartile of the income level.

In terms of the count component, which compares the number of non-zero contacts, the result for the likelihood ratio test in Table 6 showed that the effect of household size and area remains consistent across all three groups. The effect of household size however is closed to borderline significance in group 3. A similar scenario can be seen for area, which is also borderline significant in group 1, showing that area has a lesser impact under full lockdown. The employment status is the only other significant variable for group 2 and 3. The estimates for the household size in Table 8 are consistent with the other fitted models, indicating that the number of contacts increases with larger household members, and this effect is stronger for group 1 than in the other groups. Moreover, after lockdown, the number of contacts is substantially higher in Flanders than in Brussels and Wallonia. Furthermore, during the partial re-opening of schools and businesses in group 2, the students/unemployed have more/less contacts, respectively, compared to those employed. For other types of employment, such as full-time parent/homemaker and retired individuals, a stronger effect can be seen in groups 2 and 3, having fewer contacts than the employed.

Table 4: Likelihood Ratio Test for binary component of Hurdle NB GLMM

		Group 1		Group 2		Group 3	
	df	LR.stat	*p*-value	LR.stat	*p*-value	LR.stat	*p*-value
incomeQuartile	4	7.667	0.105	11.519	**0.021**	2.453	0.653
age	5	3.334	0.649	11.833	**0.037**	10.809	0.055
gender	1	2.284	0.131	0.006	0.937	1.717	0.190
hh_size	6	36.228	**<0.001**	49.864	**<0.001**	64.133	**<0.001**
hh_type	1	7.152	**0.007**	4.975	**0.026**	2.155	0.142
educ	5	9.958	0.076	7.277	0.201	7.160	0.209
employ	4	4.765	0.312	6.997	0.136	4.824	0.306
day	1	0.688	0.407	0.158	0.691	0.084	0.772
area	2	10.455	**0.005**	17.282	**<0.001**	10.657	**0.005**

Table 5: Parameter Estimates for binary component Hurdle NB GLMM

	Group 1	Group 2	Group 3
(Intercept)	-2.6437***	-1.1803	-0.8542
incomeQuartile2	0.1925	-0.8679*	-0.5676
incomeQuartile3	-0.3802	-1.3675*	-0.7477
incomeQuartile4	-0.5357	-1.3864**	-0.8124
incomeQuartilePNTA	-0.5061	-1.3348**	-0.8000
age18-29	0.5973	0.8553	0.9111
age30-39	0.5447	1.6462**	1.6898*
age40-49	0.6759	1.3152*	0.1411
age50-59	0.4166	0.8926	0.3205
age70-120	0.3752	-0.3196	-0.1911
gendermale	0.3265	0.0213	0.4218
hh_size_new2	-1.6989***	-1.8343***	-3.1066***
hh_size_new3	-2.0261***	-2.3637***	-3.0197***
hh_size_new4	-1.8697***	-2.0862***	-2.1934***
hh_size_new5	-2.8476***	-3.5255***	-4.057***
hh_typedepend_children	-0.9531**	-0.8600*	-0.6561
educl_secondary	0.4783	-0.0643	0.4752
educh_secondary	0.8823	-0.6296	-0.1324
educhigher_ed	0.1749	-0.7787	-0.6921
educbs/ms_degree	0.1757	-0.9097	-0.7311
educphd	-0.1008	1.8364	1.1864
employothers	0.3170	0.2151	-0.8485
employself-employed	0.1093	-0.5845	-0.9316
employstudent	1.6987*	2.6680*	1.6106
employunemployed	0.3192	0.1888	-0.8151
dayWeekend	-0.2682	0.1080	-0.0686
areaCenter (Bruxelles)	0.6087	1.6771***	1.4746**
areaSouth (Wallonie)	0.7144**	0.8451**	0.8347*
$\hat{\sigma}_w^2$	2.3235	6.0022	7.4340

*p-value<0.05, **p-value<0.01 , ***p-value<0.001
$\hat{\sigma}_w^2$ - random intercept variance estimate for binary component

Table 6: Likelihood Ratio Test for count component of Hurdle NB GLMM

		Group 1		Group 2		Group 3	
	df	LR.stat	p-value	LR.stat	p-value	LR.stat	p-value
incomeQuartile	4	6.717	0.152	8.355	0.079	2.896	0.575
age	5	4.543	0.474	2.621	0.758	6.233	0.284
gender	1	0.518	0.472	0.773	0.379	1.024	0.312
hh_size	6	171.638	**<0.001**	40.320	**<0.001**	12.848	**0.046**
hh_type	1	0.540	0.462	0.081	0.776	1.446	0.229
educ	5	4.877	0.431	2.763	0.736	4.572	0.470
employ	4	3.672	0.452	27.757	**<0.001**	11.334	**0.023**
day	1	2.336	0.126	0.756	0.384	0.520	0.471
area	2	6.073	**0.048**	43.852	**<0.001**	55.897	**<0.001**

Table 7: AIC of the models

	Group 1	Group 2	Group 3
Poisson GLMM	7751	11560	10381
NB GLMM	7721	9999	8301
Hurdle Poisson GLMM	7669	11132	10089
Hurdle NB GLMM	7635	9694	8032

Table 8: Parameter Estimates for count component of Hurdle NB GLMM

	Group 1	Group 2	Group 3
(Intercept)	0.8194***	1.7153***	0.9015*
incomeQuartile2	0.2222*	-0.0957	0.1491
incomeQuartile3	0.2320*	-0.1665	0.1773
incomeQuartile4	0.2014*	0.0666	-0.0082
incomeQuartilePNTA	0.1485	-0.2187	-0.0085
age18-29	-0.1407	-0.2683	-0.3998
age30-39	-0.0773	-0.0547	-0.2673
age40-49	-0.1426	-0.0044	0.0403
age50-59	-0.0478	-0.0453	-0.1499
age70-120	-0.1346	0.0411	6e-04
gendermale	-0.0356	0.0703	-0.0968
hh_size2	-0.0584	-0.2715*	0.1427
hh_size3	0.3341***	-0.0185	0.2190
hh_size4	0.639***	0.2472	0.3678
hh_size5	0.9554***	0.4153*	0.8605***
hh_size6	1.2187***	0.942**	0.6557
hh_size7	1.4645***	1.4735**	1.0021
hh_typedepend_children	-0.0457	0.0305	0.1655
educl_secondary	-0.3266*	-0.2980	0.3199
educh_secondary	-0.2131	-0.2087	0.3959
educhigher_ed	-0.2040	-0.2382	0.2327
educbs/ms_degree	-0.2078	-0.1625	0.4431
educphd	-0.1727	0.2206	0.4838
employothers	-0.1105	-0.5937***	-0.5113**
employself-employed	0.0699	-0.0378	0.0273
employstudent	0.1801	1.088*	0.7154
employunemployed	0.0745	-0.4686*	-0.2271
dayWeekend	0.0968	-0.0709	0.0558
areaCenter (Bruxelles)	-0.1740	-0.4749**	-0.5670**
areaSouth (Wallonie)	-0.1026	-0.5528***	-0.7586***
$\hat{\sigma}_v^2$	0.1947	0.5374	0.6066
$\hat{\alpha}$	0.0916	0.5541	0.7230

*p-value<0.05, **p-value<0.01 , ***p-value<0.001
$\hat{\sigma}_v^2$ - random intercept variance estimate for count component
$\hat{\alpha}$ - overdispersion parameter estimate

4 Discussion

The main aim of the study is to examine the differential effect of socioeconomic status on social contacts during and after the lockdown. To achieve this goal, the Negative Binomial model, which accounts for overdispersion, was fitted. This modelling strategy was employed in previous cross-sectional social contact studies [34, 35]. Moreover, our data was collected at several time points (a panel survey) where participants can participate in at least one wave of the survey. Hence, we expect correlated responses from the same subject. To account for this correlation, a generalized linear mixed model (GLMM) was used. In addition, during the pandemic, people tend to avoid making contacts to prevent getting infected. As a result, more people were having zero contacts than usual, leading to excess zeros. The Hurdle model was also used to address this issue of excess zeros.

It was shown that our data are overdispersed, making the Negative Binomial model more suitable than the Poisson model. Also, the Hurdle model improves the model fit of the Negative Binomial GLMM, indicating that the data contains more zeros than what we expect from our distributional assumption. In addition to providing a better fit, the benefit of hurdle model also allow us to understand the two underlying mechanism of how different social groups can make contacts. The first component of the hurdle model is the binary part that enables us to see who are those people that are most likely to have zero contact. This is especially important during a pandemic because the lockdown's primary goal is to contain individuals and prevent them from spreading the disease. Similarly, identifying those who are more likely to make contacts will also help us better understand peoples' contact behaviour. The second part of the Hurdle model is the count component, where we want to measure how many contacts each social group has generated especially for those who have a high chance of making contacts.

Among all the fitted models, the Hurdle Negative Binomial GLMM was found to be the best fit for our data as overdispersion, excess zeros, and correlated responses are all accounted for in this model. The effect of household size and area remains consistent for all the fitted models across all different group of waves. Results showed that the number of contacts increases with the size of the household. This effect, however, is less pronounced during the summer holidays (group 3) and is stronger during the lockdown (group 1). Since everyone is expected to stay at home during the lockdown, the number of people in the household largely determines the number of potential contacts. On the other hand, when people are given the opportunity to go out, like for group 2 and 3, smaller family members don't seem to matter as much because we can potentially make more interactions outside the house. Furthermore, people in Flanders are more likely to make contacts than those in Brussels and Wallonia. In addition, in terms of number of contacts, those who are living in Flanders tend to have more social contacts and that this difference is stronger after the lockdown.

In general, the socioeconomic position has a minor effect on the number of contacts (count component of hurdle model), but it is more relevant in determining who is more likely to make zero or non-zero contacts (binary component). The only socioeconomic variable that appears to have an impact in the count component is the employment status, which is only significant after the lockdown. That is, during the re-opening phase, students made more contact than employed individuals. On the other hand, the unemployed made less contact than the employed. In addition, after the lockdown, full-time parent/homemaker and retired individuals also made less contacts relative to those employed. With regards to whether a person makes contact or not, we've seen more differences during the re-opening period. Results showed that people who have a higher salary are less likely to make zero contacts after the full lockdown. Moreover, those who are 30-49 years old have higher chance of making no contacts during the re-opening period. Finally, during the lockdown, we saw lesser heterogeneity (and fewer contacts) in our data for all fitted models, suggesting that people were compliant and that the lockdown was effective in reducing number of contacts. As a result, this leads to the reduction of the virus transmission potential (reproduction number), and subsequently decreases hospital admissions and confirmed cases [43, 44].

There are several limitations to this study. First, the Hurdle model's limitation is that it assumes that the underlying mechanisms generating the two components (binary and count components) are independent of one another. Another way to address this is by using the zero-inflated models which assumes that both binary and count models are used to generate the binary component. However, we encounter some fitting problems with this model. One possible reason for this is that the zero-inflated models requires excessive amount of zeros [37, 39]. Perhaps, the zeros in our data are enough to indicate the presence of extra zeros but possibly insufficient for what the zero-inflated model requires. Second, since we group the survey waves into three groups based on the homogeneity of non-pharmaceutical interventions, we assume that the contact behaviour of participants in different waves of the same group are similar. Third, the participant's missing income in some waves were filled in using the (observed) mode income from other waves of the same participant. In addition, those who prefer not to answer (PNTA) regarding their income are included as part of the income level in the analysis. More suitable missing data analysis techniques may be able to address these issues. Lastly, there might be a recall bias because the participants were asked to report their contacts from the previous day.

5 Conclusion

In conclusion, we examined social contact differences among the demographic and socioeconomic groups during and after the lockdown. This enable us to determine which groups are more affected by the lockdown and which of them are (not) compliant with the lockdown. Consequently, this information can aid in designing better intervention measures that are equitable for all social groups and improving communication and targeting efforts.

6 References

[1] WHO Timeline - COVID-19. https://www.who.int/news/item/27-04-2020-who-timeline—covid-19.

[2] COVID-19 situation update worldwide, as of week 15, updated 22 April 2021. https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases.

[3] Chudik, A., Mohaddes, K., Pesaran, M. H., Raissi, M., & Rebucci, A. (2020). Economic consequences of Covid-19: A counterfactual multi-country analysis. Center for Economic and Policy Research: Washington DC Retrieved from https://voxeu.org/article/economic-consequences-covid-19-multicountry-analysis.

[4] Di Pietro, G., Biagi, F., Costa, P., Karpiński, Z., & Mazza, J. (2020). The likely impact of COVID-19 on education: Reflections based on the existing literature and recent international datasets (Vol. 30275). Publications Office of the European Union.

[5] Škare, M., Soriano, D. R., & Porada-Rochoń, M. (2021). Impact of COVID-19 on the travel and tourism industry. Technological Forecasting and Social Change, 163, 120469.

[6] Benke, C., Autenrieth, L. K., Asselmann, E., & Pané-Farré, C. A. (2020). Lockdown, quarantine measures, and social distancing: Associations with depression, anxiety and distress at the beginning of the COVID-19 pandemic among adults from Germany. Psychiatry research, 293, 113462.

[7] Bai, H. M., Zaid, A., Catrin, S., Ahmed, K., & Ahmed, A. J. (2020). The socio-economic implications of the coronavirus pandemic (COVID-19): A review. Int. J. Surg., 8(4), 8-17.

[8] Modes of transmission of virus causing COVID-19: implications for IPC precaution recommendations. https://www.who.int/news-room/commentaries/detail/modes-of-transmission-of-virus-causing-covid-19-implications-for-ipc-precaution-recommendations.

[9] Wallinga, J., Teunis, P., & Kretzschmar, M. (2006). Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents. American journal of epidemiology, 164(10), 936-944.

[10] Anderson, R. M., & May, R. M. (1992). Infectious diseases of humans: dynamics and control. Oxford university press.

[11] Hoang, T., Coletti, P., Melegaro, A., Wallinga, J., Grijalva, C. G., Edmunds, J. W., ... & Hens, N. (2019). A systematic review of social contact surveys to inform transmission models of close-contact infections. Epidemiology (Cambridge, Mass.), 30(5), 723.

[12] Feehan, D. M., & Mahmud, A. S. (2021). Quantifying population contact patterns in the United States during the COVID-19 pandemic. Nature communications, 12(1), 1-9.

[13] Latsuzbaia, A., Herold, M., Bertemes, J. P., & Mossong, J. (2020). Evolving social contact patterns during the COVID-19 crisis in Luxembourg. PloS one, 15(8), e0237128.

[14] McCarthy, Z., Xiao, Y., Scarabel, F., Tang, B., Bragazzi, N. L., Nah, K., ... & Wu, J. (2020). Quantifying the shift in social contact patterns in response to non-pharmaceutical interventions. Journal of Mathematics in Industry, 10(1), 1-25.

[15] Quaas, M. F., Meya, J. N., Schenk, H., Bos, B., Drupp, M. A., & Requate, T. (2021). The social cost of contacts: Theory and evidence for the first wave of the COVID-19 pandemic in Germany. PloS one, 16(3), e0248288.

[16] Coletti, P., Wambua, J., Gimma, A., Willem, L., Vercruysse, S., Vanhoutte, B., ... & Hens, N. (2020). CoMix: comparing mixing patterns in the Belgian population during and after lockdown. Scientific reports, 10(1), 1-10.

[17] Quaife, M., Van Zandvoort, K., Gimma, A., Shah, K., McCreesh, N., Prem, K., ... & Austrian, K. (2020). The impact of COVID-19 control measures on social contacts and transmission in Kenyan informal settlements. BMC medicine, 18(1), 1-11.

[18] Del Fava, E., Cimentada, J., Perrotta, D., Grow, A., Rampazzo, F., Gil-Clavel, S., & Zagheni, E. (2020). The differential impact of physical distancing strategies on social contacts relevant for the spread of COVID-19. medRxiv.

[19] Jarvis, C. I., Van Zandvoort, K., Gimma, A., Prem, K., Klepac, P., Rubin, G. J., & Edmunds, W. J. (2020). Quantifying the impact of physical distance measures on the transmission of COVID-19 in the UK. BMC medicine, 18, 1-10.

[20] Liu, C. Y., Berlin, J., Kiti, M. C., Del Fava, E., Grow, A., Zagheni, E., ... & Nelson, K. (2021). Rapid review of social contact patterns during the COVID-19 pandemic. medRxiv.

[21] Oxford COVID-19 government response tracker. https://www.bsg.ox.ac.uk/research/research-projects/covid-19-government-response-tracker

[22] Data on country response measures to COVID-19. https://www.ecdc.europa.eu/en/publications-data/download-data-response-measures-covid-19.

[23] Aquino, E. M., Silveira, I. H., Pescarini, J. M., Aquino, R., & Souza-Filho, J. A. D. (2020). Social distancing measures to control the COVID-19 pandemic: potential impacts and challenges in Brazil. Ciencia & saude coletiva, 25, 2423-2446.

[24] Chinazzi, M., Davis, J. T., Ajelli, M., Gioannini, C., Litvinova, M., Merler, S., ... & Vespignani, A. (2020). The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. Science, 368(6489), 395-400.Chinazzi, M., Davis, J. T., Ajelli, M., Gioannini, C., Litvinova, M., Merler, S., ... & Vespignani, A. (2020). The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. Science, 368(6489), 395-400.

[25] Lau, H., Khosrawipour, V., Kocbach, P., Mikolajczyk, A., Schubert, J., Bania, J., & Khosrawipour, T. (2020). The positive impact of lockdown in Wuhan on containing the COVID-19 outbreak in China. Journal of travel medicine, 27(3), taaa037.

[26] Chu, D. K., Akl, E. A., Duda, S., Solo, K., Yaacoub, S., Schünemann, H. J., ... & Reinap, M. (2020). Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. The Lancet, 395(10242), 1973-1987.

[27] Patel, J. A., Nielsen, F. B. H., Badiani, A. A., Assi, S., Unadkat, V. A., Patel, B., ... & Wardle, H. (2020). Poverty, inequality and COVID-19: the forgotten vulnerable. Public health, 183, 110.

[28] Perry, B. L., Aronson, B., & Pescosolido, B. A. (2021). Pandemic precarity: COVID-19 is exposing and exacerbating inequalities in the American heartland. Proceedings of the National Academy of Sciences, 118(8).

[29] Adams-Prassl, A., Boneva, T., Golin, M., & Rauh, C. (2020). Inequality in the impact of the coronavirus shock: Evidence from real time surveys. Journal of Public Economics, 189, 104245.

[30] Witteveen, D. (2020). Sociodemographic inequality in exposure to COVID-19-induced economic hardship in the United Kingdom. Research in Social Stratification and Mobility, 69, 100551.

[31] Wright, L., Steptoe, A., & Fancourt, D. (2020). Are we all in this together? Longitudinal assessment of cumulative adversities by socioeconomic position in the first 3 weeks of lockdown in the UK. J Epidemiol Community Health, 74(9), 683-688.

[32] Do Lee, W., Qian, M., & Schwanen, T. (2021). The association between socioeconomic status and mobility reductions in the early stage of England's COVID-19 epidemic. Health & Place, 69, 102563.

[33] Pullano, G., Valdano, E., Scarpa, N., Rubrichi, S., & Colizza, V. (2020). Evaluating the effect of demographic factors, socioeconomic factors, and risk aversion on mobility during the COVID-19 epidemic in France under lockdown: a population-based study. The Lancet Digital Health, 2(12), e638-e649.

[34] Mossong, J., Hens, N., Jit, M., Beutels, P., Auranen, K., Mikolajczyk, R., ... & Edmunds, W. J. (2008). Social contacts and mixing patterns relevant to the spread of infectious diseases. PLoS Med, 5(3), e74.

[35] Van Hoang, T., Coletti, P., Kifle, Y. W., Van Kerckhove, K., Vercruysse, S., Willem, L., ... & Hens, N. (2021). Close contact infection dynamics over time: insights from a second large-scale social contact survey in Flanders, Belgium, in 2010-2011. BMC infectious diseases, 21(1), 1-15.

[36] Agresti, A. (2013). Categorical data analysis 3rd ed. John Willey and Sons, USA.

[37] Hilbe, J. M. (2011). Negative binomial regression. Cambridge University Press.

[38] Molenberghs, G., & Verbeke, G. (2005). Models for discrete longitudinal data. Statistics 2005: Springer.

[39] Min, Y., & Agresti, A. (2005). Random effect models for repeated measures of zero-inflated count data. Statistical modelling, 5(1), 1-19.

[40] R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

[41] Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., ... & Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. R J. 9, 378–400. doi: 10.32614. RJ-2017-066.

[42] Statistics Belgium; Eurostat (2020), European Union Statistics on Income and Living Conditions (EU-SILC), ilc_di01, https://ec.europa.eu/eurostat

[43] Coletti, P., Libin, P., Petrof, O., Willem, L., Abrams, S., Herzog, S. A., ... & Hens, N. (2021). A data-driven metapopulation model for the Belgian COVID-19 epidemic: assessing the impact of lockdown and exit strategies. BMC Infectious Diseases, 21(1), 1-12.

[44] Abrams, S., Wambua, J., Santermans, E., Willem, L., Kuylen, E., Coletti, P., ... & Hens, N. (2021). Modelling the early phase of the Belgian COVID-19 epidemic using a stochastic compartmental model and studying its implied future trajectories. Epidemics, 35, 100449.

7 Appendix

7.1 List of variables and sample distribution

Table 9: List of Variables

Variables	Levels	Description	
Income	Q1 (*ref*) Q2 Q3 Q4 PNTA	-	1st quartile 2nd quartile 3rd quartile 4th quartile Prefer not to answer
Age	18-29 (*ref*), 30-39, 40-49, 50-59, 60-69, 70+	-	-
Gender	male (*ref*), female	-	-
HH_size	1 (*ref*), 2, 3, 4, 5, 6, 7	Household size	-
HH_type	no_depend_children (*ref*) depend_children	Household type	Without dependent children With dependent children
Educ	primary (*ref*) lower_secondary higher_secondary higher_ed bs/ms degree phd	Educational attainment	Primary education Lower secondary education Upper secondary education Higher education Bachelor's/Master's degree Doctorate
Employ	employed (*ref*), unemployed, self-employed, student, others	Employment status	-
Day	weekday(*ref*), weekend	Day of survey	-
Area	North (Flandre) (*ref*), Center (Bruxelles, South (Wallonie)	-	-

**ref* means reference category

Table 10: Distribution of sample sizes

	Group 1	Group 2	Group 3
	N=2806	N=3003	N=2484
	n(%)		
Age			
18-29	193 (9.82%)	152 (6.96%)	113 (6.21%)
30-39	278 (14.14%)	289 (13.23%)	216 (11.87%)
40-49	409 (20.8%)	427 (19.54%)	334 (18.35%)
50-59	393 (19.99%)	449 (20.55%)	396 (21.76%)
60-69	459 (23.35%)	582 (26.64%)	494 (27.14%)
70-120	234 (11.9%)	286 (13.09%)	267 (14.67%)
Area			
Center (Bruxelles)	163 (8.29%)	170 (7.78%)	139 (7.64%)
North (Flandre)	1188 (60.43%)	1317 (60.27%)	1115 (61.26%)
South (Wallonie)	615 (31.28%)	698 (31.95%)	566 (31.1%)
Day of Survey			
Weekday	1740 (88.5%)	1849 (84.62%)	1289 (70.82%)
Weekend	226 (11.5%)	336 (15.38%)	531 (29.18%)
Educational Level			
primary	59 (3%)	67 (3.07%)	54 (2.97%)
l_secondary	238 (12.11%)	263 (12.04%)	219 (12.03%)
h_secondary	766 (38.96%)	848 (38.81%)	713 (39.18%)
higher_ed	530 (26.96%)	587 (26.86%)	480 (26.37%)
bs/ms_degree	355 (18.06%)	401 (18.35%)	334 (18.35%)
phd	18 (0.92%)	19 (0.87%)	20 (1.1%)
Employment Status			
employed	968 (49.24%)	1000 (45.77%)	804 (44.18%)
others	797 (40.54%)	987 (45.17%)	872 (47.91%)
self-employed	95 (4.83%)	84 (3.84%)	61 (3.35%)
student	26 (1.32%)	20 (0.92%)	12 (0.66%)
unemployed	80 (4.07%)	94 (4.3%)	71 (3.9%)
Gender			
female	828 (42.12%)	868 (39.73%)	723 (39.73%)
male	1138 (57.88%)	1317 (60.27%)	1097 (60.27%)
Household Size			
1	448 (22.79%)	564 (25.81%)	531 (29.18%)
2	673 (34.23%)	829 (37.94%)	726 (39.89%)
3	422 (21.46%)	429 (19.63%)	302 (16.59%)
4	262 (13.33%)	244 (11.17%)	185 (10.16%)
5	112 (5.7%)	88 (4.03%)	55 (3.02%)
6	34 (1.73%)	25 (1.14%)	18 (0.99%)
7	15 (0.76%)	6 (0.27%)	3 (0.16%)
Household Type			
depend_children	559 (28.43%)	549 (25.13%)	436 (23.96%)
no_depend_children	1407 (71.57%)	1636 (74.87%)	1384 (76.04%)
Income			
Q1	309 (15.72%)	351 (16.06%)	286 (15.71%)
Q2	389 (19.79%)	466 (21.33%)	411 (22.58%)
Q3	218 (11.09%)	274 (12.54%)	229 (12.58%)
Q4	771 (39.22%)	803 (36.75%)	683 (37.53%)
PNTA	279 (14.19%)	291 (13.32%)	211 (11.59%)

7.2 Results for Poisson GLMM

Table 11: Likelihood Ratio Test for Poisson GLMM

		Group 1		Group 2		Group 3	
	df	LRT.stat	*p*-value	LRT.stat	*p*-value	LRT.stat	*p*-value
incomeQuartile	4	6.529	0.163	8.422	0.077	4.017	0.404
age	5	5.806	0.326	3.556	0.615	15.967	**0.007**
gender	1	1.752	0.186	1.355	0.244	1.939	0.164
hh_size	6	202.734	**<0.001**	69.492	**<0.001**	74.666	**<0.001**
hh_type	1	0.162	0.687	12.944	**<0.001**	3.929	**0.047**
educ	5	6.213	0.286	2.098	0.835	4.821	0.438
employ	4	3.048	0.550	13.742	**0.008**	3.074	0.546
day	1	2.644	0.104	24.769	**<0.001**	19.880	**<0.001**
area	2	12.331	**0.002**	47.822	**<0.001**	56.112	**<0.001**

Table 12: Parameter Estimates for Poisson GLMM

	Group 1	Group 2	Group 3
(Intercept)	0.7103***	0.9662***	0.3963
incomeQuartile2	0.1286	0.1405	0.2146
incomeQuartile3	0.2001*	0.1770	0.2711
incomeQuartile4	0.1750*	0.3151**	0.1338
incomeQuartilePNTA	0.1226	0.1958	0.1822
age18-29	-0.1577	-0.1207	-0.2020
age30-39	-0.1030	-0.1255	-0.4329*
age40-49	-0.1408	0.0000	0.0823
age50-59	-0.0539	-0.0437	-0.1520
age70-120	-0.1181	0.1375	0.0475
gendermale	-0.0550	0.0804	-0.1069
hh_size2	0.2143***	0.0428	0.4456***
hh_size3	0.5276***	-0.0200	0.5189***
hh_size4	0.7785***	0.3025**	0.7509***
hh_size5	1.0441***	0.3858**	1.2713***
hh_size6	1.3218***	1.0773***	0.7994***
hh_size7	1.6006***	1.4968***	1.6493***
hh_typedepend_children	-0.0208	0.2601***	0.1779*
educl_secondary	-0.2580*	-0.1941	0.0871
educh_secondary	-0.2179	-0.0499	0.2478
educhigher_ed	-0.1512	-0.0709	0.2182
educbs/ms_degree	-0.1545	-0.0357	0.3691
educphd	-0.1280	-0.2305	0.0875
employothers	-0.1134	-0.3568***	-0.1686
employself-employed	0.0285	0.0228	0.0696
employstudent	-0.0153	0.0241	-0.2068
employunemployed	-0.0020	-0.3861*	0.0709
dayWeekend	0.0833	-0.2016***	0.1497***
areaCenter (Bruxelles)	-0.1693*	-0.5771***	-0.6198***
areaSouth (Wallonie)	-0.1408**	-0.4655***	-0.5870***
$\hat{\sigma}^2$	0.1940	0.7437	0.7692

*p-value<0.05, **p-value<0.01 , ***p-value<0.001
$\hat{\sigma}^2$ - random intercept variance estimate

7.3 Results for Hurdle Poisson GLMM

Table 13: Likelihood Ratio Test for binary component of Hurdle Poisson GLMM

		Group 1		Group 2		Group 3	
	df	LRT.stat	p-value	LRT.stat	p-value	LRT.stat	p-value
incomeQuartile	4	7.667	0.105	11.519	**0.021**	2.453	0.653
age	5	3.334	0.649	11.833	**0.037**	10.809	0.055
gender	1	2.284	0.131	0.006	0.937	1.717	0.190
hh_size	6	36.228	**<0.001**	49.864	**<0.001**	64.133	**<0.001**
hh_type	1	7.152	**0.007**	4.975	**0.026**	2.155	0.142
educ	5	9.958	0.076	7.277	0.201	7.160	0.209
employ	4	4.765	0.312	6.997	0.136	4.824	0.306
day	1	0.688	0.407	0.158	0.691	0.084	0.772
area	2	10.455	**0.005**	17.282	**<0.001**	10.657	**0.005**

Table 14: Likelihood Ratio Test for count component of Hurdle Poisson GLMM

		Group 1		Group 2		Group 3	
	df	LRT.stat	p-value	LRT.stat	p-value	LRT.stat	p-value
incomeQuartile	4	7.028	0.134	8.649	0.070	4.374	0.358
age	5	4.940	0.423	5.298	0.381	6.631	0.250
gender	1	0.540	0.462	2.153	0.142	0.610	0.435
hh_size	6	184.022	**<0.001**	90.975	**<0.001**	62.070	**<0.001**
hh_type	1	2.023	0.155	7.581	**0.006**	0.172	0.678
educ	5	4.787	0.442	3.009	0.699	3.366	0.644
employ	4	3.854	0.426	20.894	**<0.001**	8.984	0.061
day	1	2.599	0.107	25.742	**<0.001**	23.499	**<0.001**
area	2	5.983	0.050	35.065	**<0.001**	56.521	**<0.001**

Table 15: Parameter Estimates for binary component Hurdle Poisson GLMM

	Group 1	Group 2	Group 3
(Intercept)	-2.6437***	-1.1803	-0.8542
incomeQuartile2	0.1925	-0.8678*	-0.5676
incomeQuartile3	-0.3802	-1.3674*	-0.7477
incomeQuartile4	-0.5357	-1.3864**	-0.8125
incomeQuartilePNTA	-0.5061	-1.3348**	-0.8000
age18-29	0.5973	0.8553	0.9111
age30-39	0.5447	1.6462**	1.6897*
age40-49	0.6759.	1.3152*	0.1411
age50-59	0.4166	0.8926	0.3205
age70-120	0.3752	-0.3196	-0.1911
gendermale	0.3265	0.0213	0.4217
hh_size_new2	-1.6989***	-1.8343***	-3.1066***
hh_size_new3	-2.0261***	-2.3637***	-3.0196***
hh_size_new4	-1.8697***	-2.0862***	-2.1934***
hh_size_new5	-2.8476***	-3.5255***	-4.0570***
hh_typedepend_children	-0.9531**	-0.8601*	-0.6561
educl_secondary	0.4783	-0.0643	0.4753
educh_secondary	0.8822	-0.6297	-0.1324
educhigher_ed	0.1748	-0.7788	-0.6921
educbs/ms_degree	0.1756	-0.9098	-0.7310
educphd	-0.1008	1.8362	1.1865
employothers	0.3169	0.2151	-0.8485
employself-employed	0.1093	-0.5845	-0.9316
employstudent	1.6987*	2.6680*	1.6105
employunemployed	0.3192	0.1889	-0.8151
dayWeekend	-0.2682	0.1080	-0.0685
areaCenter (Bruxelles)	0.6087	1.6771***	1.4746**
areaSouth (Wallonie)	0.7144**	0.8450**	0.8347*
$\hat{\sigma}_w^2$	2.3235	6.0022	7.4340

*p-value<0.05, **p-value<0.01 , ***p-value<0.001

$\hat{\sigma}_w^2$ - random intercept variance estimate for binary component

Table 16: Parameter Estimates for count component Hurdle Poisson GLMM

	Group 1	Group 2	Group 3
(Intercept)	0.8131***	1.5825***	0.8573**
incomeQuartile2	0.2208*	-0.0144	0.1751
incomeQuartile3	0.2269*	-0.0292	0.1910
incomeQuartile4	0.1968*	0.2025	-0.0041
incomeQuartilePNTA	0.1432	0.0296	0.0072
age18-29	-0.1457	-0.1014	-0.2069
age30-39	-0.0745	0.1014	-0.2212
age40-49	-0.1340	0.2016	0.0855
age50-59	-0.0343	0.0902	-0.1464
age70-120	-0.1296	0.0646	0.0338
gendermale	-0.0352	0.1064	-0.0643
hh_size2	-0.0542	-0.3979***	0.1185
hh_size3	0.3499***	-0.4608***	0.2463*
hh_size4	0.6667***	-0.1019	0.6854***
hh_size5	0.9192***	-0.1265	1.1066***
hh_size6	1.1874***	0.6124***	0.5392*
hh_size7	1.4411***	1.0222***	1.3319***
hh_typedepend_children	-0.0817	0.2135**	0.0420
educl_secondary	-0.3116*	-0.2544	0.2157
educh_secondary	-0.2007	-0.2154	0.2947
educhigher_ed	-0.1904	-0.2659	0.1887
educbs/ms_degree	-0.1894	-0.1853	0.3498
educphd	-0.1528	0.2236	0.4576
employothers	-0.1148	-0.3997***	-0.3803**
employself-employed	0.0515	0.0148	-0.0549
employstudent	0.1822	0.9194*	0.4059
employunemployed	0.0751	-0.3868*	-0.0011
dayWeekend	0.0922	-0.2225***	0.1767***
areaCenter (Bruxelles)	-0.1668	-0.4333**	-0.4898**
areaSouth (Wallonie)	-0.0992	-0.4434***	-0.6727***
$\hat{\sigma}_v^2$	0.2367	0.7002	0.7512

*p-value<0.05, **p-value<0.01 , ***p-value<0.001

$\hat{\sigma}_v^2$ - random intercept variance estimate for count component

7.4 R Code

```
library(readr)
library(glmmTMB)

group1 <- read_csv("group1.csv")
group2 <- read_csv("group2.csv")
group3 <- read_csv("group3.csv")

var<-c("incomeQuartile","age","gender",
       "hh_size","hh_type","educ","employ","day","area")

############# (1) GLMM Poisson ##########
glmm_pois_group1 = glmmTMB(as.formula(paste("n_cnt_all ~",
                                            paste(var,collapse  = "+"),
                                            paste("(1|panel_id)"),sep = "+")),
                           family=poisson, data=group1)

## Likelihood Ratio test for Poisson GLMM
LRT1<-matrix(ncol=3,nrow=length(var))
colnames(LRT1)<-c("df","LR stat.","Pr(Chi)")
rownames(LRT1)<-var
lrt<-list()
for (i in 1:length(var)){
  m<-update(glmm_pois_group1,as.formula(paste(".~.-",var[i],sep="")))
  lrt=anova(m,glmm_pois_group1)
  LRT1[i,1]<-lrt$`Chi Df`[2]
  LRT1[i,2]<-lrt$Chisq[2]
  LRT1[i,3]<-lrt$`Pr(>Chisq)`[2]
}
LRT1

############# (2) GLMM Negative Binomial ##########
glmm_nb_group1 = glmmTMB(as.formula(paste("n_cnt_all ~",
                                          paste(var,collapse  = "+"),
                                          paste("(1|panel_id)"),sep = "+")),
                         family=nbinom2, data=group1)

## Likelihood Ratio test for Negative Binomial GLMM
LRT1<-matrix(ncol=3,nrow=length(var))
colnames(LRT1)<-c("df","LR stat.","Pr(Chi)")
rownames(LRT1)<-var
lrt<-list()
for (i in 1:length(var)){
  m<-update(glmm_nb_group1,as.formula(paste(".~.-",var[i],sep="")))
  lrt=anova(m,glmm_nb_group1)
  LRT1[i,1]<-lrt$`Chi Df`[2]
  LRT1[i,2]<-lrt$Chisq[2]
  LRT1[i,3]<-lrt$`Pr(>Chisq)`[2]
}
LRT1


############ HURDLE MODEL ##############

#covariates for count component of hurdle model
var<-c("incomeQuartile","age","gender",
       "hh_size","hh_type","educ","employ","day","area")

#covariates for binary component of hurdle model
var1<-c("incomeQuartile","age","gender",
```

```
        "hh_size_new","hh_type","educ","employ","day","area")

############# (3) Poisson Hurdle GLMM ##########
hurdle_glmm_pois_group1 = glmmTMB(as.formula(paste("n_cnt_all ~", paste(var,collapse  = "+"),
                                       paste("(1|panel_id)"),sep = "+")),
                          ziformula =as.formula(paste(" ~", paste(var1,collapse  = "+"),
                                                  paste("(1|panel_id)"),sep = "+")),
                          family=truncated_poisson, data=group1)

## Likelihood Ratio test for binary component of Poisson Hurdle GLMM
LRT1_binary<-matrix(ncol=3,nrow=length(var1))
colnames(LRT1_binary)<-c("df","LR stat.","Pr(Chi)")
rownames(LRT1_binary)<-var1

for (i in 1:length(var1)){
  m<-glmmTMB(as.formula(paste("n_cnt_all ~", paste(var,collapse  = "+"),
                              paste("(1|panel_id)"),sep = "+")),
             ziformula =as.formula(paste(" ~", paste(var1[-i],collapse  = "+"),
                                         paste("(1|panel_id)"),sep = "+")),
             family=truncated_poisson, data=group1)
  sum<-summary(m)
  LRT1_binary[i,1]<-(length(coef(s1)$cond[,1])+length(coef(s1)$zi[,1])) -
                    (length(coef(sum)$cond[,1])+length(coef(sum)$zi[,1]))
  LRT1_binary[i,2]<-c(2*(logLik(hurdle_glmm_pois_group1)-logLik(m)))
  LRT1_binary[i,3]<-c(1-pchisq(c(2*(logLik(hurdle_glmm_pois_group1)-logLik(m))),
                        df=LRT1_binary[i,1]))
}
LRT1

## Likelihood Ratio test for count component of Poisson Hurdle GLMM
LRT1<-matrix(ncol=3,nrow=length(var))
colnames(LRT1)<-c("df","LR stat.","Pr(Chi)")
rownames(LRT1)<-var

for (i in 1:length(var)){
  m<-glmmTMB(as.formula(paste("n_cnt_all ~", paste(var[-i],collapse  = "+"),
                              paste("(1|panel_id)"),sep = "+")),
             ziformula =as.formula(paste(" ~", paste(var1,collapse  = "+"),
                                         paste("(1|panel_id)"),sep = "+")),
             family=truncated_poisson, data=group1)
  sum<-summary(m)
  LRT1[i,1]<-(length(coef(s1)$cond[,1])+length(coef(s1)$zi[,1]))-
                (length(coef(sum)$cond[,1])+length(coef(sum)$zi[,1]))
  LRT1[i,2]<-c(2*(logLik(hurdle_glmm_pois_group1)-logLik(m)))
  LRT1[i,3]<-c(1-pchisq(c(2*(logLik(hurdle_glmm_pois_group1)-logLik(m))),
                        df=LRT1[i,1]))
}
LRT1


############# (4) Negative Binomial Hurdle GLMM ##########
hurdle_glmm_nb_group1 = glmmTMB(as.formula(paste("n_cnt_all ~", paste(var,collapse  = "+"),
                                                 paste("(1|panel_id)"),sep = "+")),
                                  ziformula =as.formula(paste(" ~", paste(var1,collapse  = "+"),
                                                              paste("(1|panel_id)"),sep = "+")),
                                  family=truncated_nbinom2, data=group1)

## Likelihood Ratio test for binary component of Negative Binomial Hurdle GLMM
LRT1_binary<-matrix(ncol=3,nrow=length(var1))
colnames(LRT1_binary)<-c("df","LR stat.","Pr(Chi)")
```

```
rownames(LRT1_binary)<-var1

for (i in 1:length(var1)){
  m<-glmmTMB(as.formula(paste("n_cnt_all ~", paste(var,collapse  = "+"),
                              paste("(1|panel_id)"),sep = "+")),
             ziformula =as.formula(paste(" ~", paste(var1[-i],collapse  = "+"),
                                         paste("(1|panel_id)"),sep = "+")),
             family=truncated_nbinom2, data=group1)
  sum<-summary(m)
  LRT1_binary[i,1]<-(length(coef(s1)$cond[,1])+length(coef(s1)$zi[,1]))-
                    (length(coef(sum)$cond[,1])+length(coef(sum)$zi[,1]))
  LRT1_binary[i,2]<-c(2*(logLik(hurdle_glmm_nb_group1)-logLik(m)))
  LRT1_binary[i,3]<-c(1-pchisq(c(2*(logLik(hurdle_glmm_nb_group1)-logLik(m))),
                               df=LRT1_binary[i,1]))
}

## Likelihood Ratio test for count component of Negative Binomial Hurdle GLMM
LRT1<-matrix(ncol=3,nrow=length(var))
colnames(LRT1)<-c("df","LR stat.","Pr(Chi)")
rownames(LRT1)<-var

for (i in 1:length(var)){
  m<-glmmTMB(as.formula(paste("n_cnt_all ~", paste(var[-i],collapse  = "+"),
                              paste("(1|panel_id)"),sep = "+")),
             ziformula =as.formula(paste(" ~", paste(var1,collapse  = "+"),
                                         paste("(1|panel_id)"),sep = "+")),
             family=truncated_nbinom2, data=group1)
  sum<-summary(m)
  LRT1[i,1]<-(length(coef(s1)$cond[,1])+length(coef(s1)$zi[,1]))-
                (length(coef(sum)$cond[,1])+length(coef(sum)$zi[,1]))
  LRT1[i,2]<-c(2*(logLik(hurdle_glmm_nb_group1)-logLik(m)))
  LRT1[i,3]<-c(1-pchisq(c(2*(logLik(hurdle_glmm_nb_group1)-logLik(m))),
                        df=LRT1[i,1]))
}
```