UHASSELT
KNOWLEDGE IN ACTION

Maastricht University

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

Master's thesis

Social contact patterns relevant to the spread of SARS-CoV-2 and other infectious diseases in a rural sub-Saharan setting

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

SUPERVISOR :
dr. Pietro COLETTI

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

2020
2021

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

Master's thesis

Social contact patterns relevant to the spread of SARS-CoV-2 and other infectious diseases in a rural sub-Saharan setting

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

SUPERVISOR :
dr. Pietro COLETTI

Abstract

Introduction: The recently emerged disease, COVID-19 has been a heavy burden on health systems around the world including those in sub-saharan areas where healthcare resources are limited. Quantifying the social mixing behaviour relevant to the spread of SARS-COV-2 and other respiratory diseases is crucial in disease modeling and intervention optimization, yet this information is scarce in rural sub-saharan areas.

Methods: We conducted a Sero-survey with a two-months follow-up of participants, living in Kimpese Health Demographic and Surveillance System, along with social mixing surveys, where participants were asked to record details about contacts they had the previous day of the interview. Reported contacts and their characteristics were analysed using Negative binomial model and SOCRATES tool.

Results: Most contacts occurred within households, which accounted for 75% and 81% of all contacts during the first and the second wave respectively, followed by the field and river. In the majority of contacts, participants interacted with household members within their households, as well as outside household(field and river). The mixing pattern was found to be an inter-generational mixing, highly observed between young(0-17years) and adult participants(17-50years). We observed a small but significant increase in the average number of reported contacts during the second wave, when measures were re-installed. We also observed low sero-prevalence(2,1%).

Conclusion: Our study gives in-depth information on the mixing behaviour of rural sub-saharan areas. It therefore fills an important knowledge gap that will help to more accurately predict the disease transmission dynamics, and the impact of control strategies in such areas.

Contents

Acknowledgements

I would like to express my sincere gratitude to my supervisors Dr. Pietro Coletti and Dr. Esther Van Kleef for their invaluable contribution, assistance, motivation and support over the course of my thesis. I appreciate the time invested on me providing explanations, correcting and advising me, which have enabled me to finish my thesis on the predicted time. Without your persistent assistance, the goal of this thesis would not have been fulfilled. My gratitude to all lecturers for their contribution in expanding my knowledge and skills.

I would like to thank the CRSK(Centre de Recherche en Santé Kimpese) and HDSS(Health Demographic Surveillance System) team, for their enthusiastic participation and contribution in data collection and data curation, as well as their amazing teamwork. I would want to specifically thank Dr. Inocencio da Luz Raquel, who facilitated this collaboration and provided significant assistance over the course of my thesis.

I also thank my classmates and friends for their supports and encouragements. Lastly, I would want to express my appreciation to my family for their support and prayers.

1 Introduction

In December 2019, pneumonia cases of unknown cause were identified in China, in Wuhan city [21]. This was found to be caused by a novel coronavirus, coined Severe Acute Respiratory Syndrome Coronavirus 2(SARS-COV-2) and the disease was coined COVID-19 [2]. This virus can be transmitted via respiratory droplets among people who are physically close each other, touching contaminated objects and then touch their eyes and nose without washing hands, via aerosol transmission in some cases such as indoor or clouded areas with no ventilator [7]. Since the virus is highly contagious, it has spread quickly through 222 nations, resulting in a large number of cases and fatalities. As a result, on March 11, 2020, WHO declared COVID-19 a pandemic. Up until March, more than 176 millions COVID-19 cases have been reported, and more than 3 million patients have died worldwide [1]. Though the outbreak situation is different in various countries, this pandemic remains an important issue and threat to public health, education, economy and development of all countries.

While Sub-Saharan African countries have less covid-19 cases and fatalities than Asian, American, and European countries, the disease's burden is high due to inadequate health-care facilities, scarce financial resources, protective equipment, poor testing and treatment capacities, and lack of research funding [19]. The first COVID-19 case registered in the Republic Democratic of Congo (DRC) was on 10^{th} March 2020, in the city of Kinshasa, and later spread to neighboring regions. At the moment, the virus has spread through 23 provinces, with 34 695 cases registered and 831 fatalities, with Kongo Central province ranking the fourth in terms of severity, with 1 806 cases [1]. We should not forget that these are official figures, which are unreliable due to a lack of regular surveillance, likely resulting in under-reporting.

Physical distancing policies have been implemented by the government in order to reduce and restrict social contacts in the population, thus decreasing disease transmission(https://covidtracker.bsg.ox.ac.uk/). Some of them, such as introduction of school closure, workplace closure, public events ban, requirements to stay at home, and internal movement limits, have been shown in studies to play a significant role in reducing the spread of the virus in the population [8] [14]. However, the implementation of these measures cannot be sustained for an extended period of time because it has a negative impact on the economy, education, and development of countries [17]. As a result, governments chose to implement some measures in phases, such as school closure, workplace closure by encouraging remote working where possible while also enforcing others such as mandatory mask wearing, hand disinfection, and maintaining a 1.5 meter distance, among other things. To accurately predict the disease dynamics and the impact of such interventions, mathematical modeling of disease transmission have proven to be a useful tool [13], but the model outcomes are critically dependent on individual contact rates [3]. Such models often assume homogeneous mixing patterns but such assumptions may result in incorrect model estimates if such contact patterns are not representative. Therefore, understanding the mixing behavior of a specific population is critical for accurately estimating model parameters and making predictions.

For the time being, considerable knowledge is available on mixing behavior in the European population, where these countries are likely to share the mixing behavior of age assortativity, which means that people tend to have contacts with others of the same age, and it is highly observed in young ages than in the elderly [16]. This knowledge aided researchers in their analysis of the effect of specific intervention measures in reducing COVID-19 transmission and monitoring the change in mixing patterns in different waves of this pandemic [20] [10] [4]. However, these results cannot be generalized to Sub-Saharan countries because the mixing behavior, household structure, and inter-generational mixing of Europeans population are vastly different from those of Sub-Saharan Africans, who account for 14% of the world's population.

Unfortunately, little is known about mixing behavior in Sub-Saharan African countries. To date only five studies have been conducted in four sub-saharan countries including South Africa [11] [6], Zimbabwe [15], Kenya [12] and Uganda [5] but none was conducted in Democratic Republic of Congo situated in Central Africa. These studies have shown on average a higher number of contacts, also evidence of highly assortative mixing among young age groups but also inter-generational mixing more than observed in high-income countries. In addition, only three of these four studies were performed in rural areas, where people have limited access to medical care. This information gap negatively affect disease control optimization. Therefore, identifying social mixing pattern related to the spread of SARS-COV-2 and other respiratory diseases in rural Sub-Saharan Africa as well as tracking changes in mixing patterns across different disease waves, is critical.

Study objectives

The aim of our study is to gain an understanding on the transmission dynamics of SARS-CoV-2 in the rural area of DRC:

1. To evaluate factors influencing reported contacts in rural Sub-Saharan Africa.

2. To compute contact patterns explaining the mixing behaviour the population.

Through these objectives, the comparison will be made for two waves when different restriction measures were implemented, in order to assess the relationship between restriction measures and changes in mixing patterns for rural Sub-Saharan population.
The subsequent part of the thesis is organised as follows: Section 2 describes the study design, Section 3 summarises the data used in this study, Section 3 describes statistical methods that were implemented. Section 4 and 5, give an outline of the results and discussion respectively, Section 7 discusses the strength and limitation of this study, Section 8 provides recommendations to future studies, Section 9 provides the conclusion. References and appendices are presented thereafter.

2 Study design

The study was conducted in the Health Demographic Surveillance Site of Kimpese located in district of Central Kongo, Republic Democratic of Congo. The study includes two components: a sero-prevalence and social mixing, but we will focus on the social mixing component.

The sero-prevalence study began in October 2020 with a cohort of 800 individuals aged 18 to 50 years, assuming they are the most mobile sector of the population, and hence the most likely to come into contact with the virus. By this time, there were about 41 cases, all limited to the Cité. Therefore, there were no reported cases in the rural area. Once the transmission of SARS-CoV-2 was confirmed in Kimpese Health Zone, children and elderly individuals were also included in the survey. Blood samples were collected every two months for SARS-CoV-2 serology. In the case of a positive SARS-CoV-2 test, household members of the infected individual, were also tested. This sample would allow the estimation of any sero-prevalence below 5% with a precision in the order of 2%, and allow us to exclude with 98% certainty a prevalence of $\geq 1\%$ if no positive samples are found.

The social mixing survey was organized, with a target sample of 2400 individuals, resulted from sampling 800 persons in each age group of the three age categories, young (0-17years), adults (18-50years) and elderly (50+). They were asked to fill in a questionnaire regarding their social contacts in the 24 hours preceding the study. Additional social mixing surveys would then be planned during the two-month follow up sero survey to track the mixing behaviour changes.

3 Data description

3.1 Sero-prevalence survey

Sero-prevalence data, consisted of information regarding the presence of antibodies(IgM and IgG) in each participant(present/absent). The presence of both IgM and IgG antibodies would indicate recent infection while the presence of IgM would indicate old infection. This information was available for 1709 participants.

3.2 Social mixing survey

Two social mixing surveys were conducted during two waves, when different restriction measures were implemented. For the first wave, data were collected from 24/09/2020 to 23/12/2020 and for the second wave, from 22/01/2021 to 24/02/2021 as illustrated in figure 1 below. DRC government provided access to information on restriction measure by time (*http://cmr-covid19.cd/*).

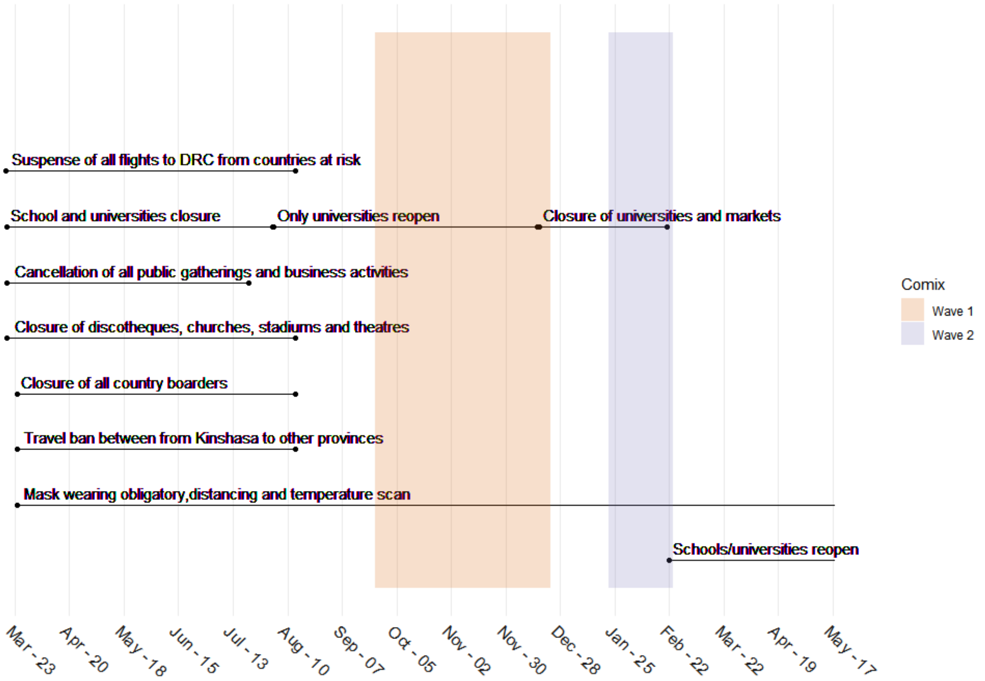

Figure 1: Restriction measures

During the first wave, measures were lifted. Public spaces, bars, restaurants, churches and places of worship reopened, and Universities and high institutions were open. However, during the second wave, measures were re-installed, with closure of public markets, gatherings limited to 10 people, a curfew between 10pm and 5 am and closure of universities and high institutions. During two waves, primary, secondary schools were closed.

Participant information

About 2400 residents of the Kimpese HDSS demographic surveillance area, were invited from registered households to participate in this study. A household was defined as a socio-economic unit within which different members are related or not. They live together in the same plot, all contribute to the same budget (especially for food) or take their main food from the same granaries and share the same meals. Here it is indeed individuals currently residing in households for a period of 3 months preceding the survey. Once participants were selected and agreed to participate in the study, their socio-demographic characteristics (described in table 1), were extracted from the central HDSS database. Participants were asked to record if they follow restrictive measures, such as wearing mask outside their household(field, work, transport, gatherings) and also if they would stay home when they have fever and dry cold.

Table 1: Participant's characteristics

Variables	Categories
Gender	Male ,Female
Age category	[0-17] ,[18-49] , 50+
Religion	Christians, Non Christians
Ethnicity	Ndibu ,Ngombe , Nianga, others
Employment status	Office worker, Farm worker,Student Unemployed , Other manual worker
Schooling level	Primary , Secondary University, None
Day of contact	Monday, Tuesday, Wednesday, Thursday Friday, Saturday, Sunday
Mask wearing	Yes, No
Number of contacts	Numeric

Contact information

All participants were asked to report all direct physical and non-physical contacts made prior to the survey day, between the time of wake-up the day before the survey, and the time of wake-up the day of the survey. A contact was defined as either of the following encounters:

- A two-way conversation with three or more words in the physical presence of another person.
- Physical skin-to-skin contact. Physical contact includes hand shaking, sharing a bike, kissing, embracing, and also sharing a glass or other utensils passed directly from mouth to mouth. Non-physical contact happens when a participant haven't touched the person.

Contacts that a participant had encountered with multiple times, were asked to be registered once if in the same place, with the total time spent with that person in that place recorded. In case participants did not know the age of the person they had contact with, they were asked to provide an estimate of the age range. For children participants, their parents were asked to provide contacts characteristics in their behalf. To construct contact matrices, participants data were combined with the contact data.

Table 2: Contact's characteristics

Variables	**Categories**
Gender	Male ,Female
Age category	[0-17] ,[18-49] , 50+
Location	Household , Other household, Field, School River , Place of worship, Shop, Work Transport, Place of leisure, Others
Vicinity	Within village, Outside village, Outside kimpese
Relation	Household member, Other relative, neighbour, Friend Collegue, Schoolmate, Boy/girlfriend, Others
Duration	<5mins, 5-14mins, 15min-1h, 1-4h, >4h
Type of contact	Physical, Non-physical

4 Methods

4.1 Study population

We analysed the frequency distribution of contacts from participants, with complete data available during two waves for a set of covariates, including age category, gender, employment status, education level, day of contact, religion, and ethnicity.

4.2 Participants contacts

Missing data were imputed based on the age distribution of observed data using probability of contacts between age categories. Frequency tables were used to describe and compute the mean number of contacts and Inter-quartile range(IQR) with respect to different characteristics of participants, during two waves. We used negative binomial regression to estimate the mean contacts as a function of the different covariates of interest.

Negative binomial was preferred over Poisson regression to account of over-dispersion, that is, the sample variance ($s^2 = \sum_{i=1}^{n}(y_i - \bar{y})^2/(n-1)$) is larger than the sample mean ($\bar{y} = \sum_{i=1}^{n} y_i/n$). The negative binomial distribution can be viewed as a Poisson distribution where the Poisson parameter is itself a random variable, distributed according to a Gamma distribution [9].

Let Y representing the number of daily contacts, follows a Poisson distribution with parameter θ. Thus the parameter θ itself, is assumed to vary according to the gamma distribution, with the scale parameter α and shape parameter β. The following is the conditional distribution of the random variable Y (conditional on θ):

$$P(Y|\theta) = \frac{e^{-\theta}\theta^{y}}{y!} \ , \ y = 0, 1, 2, ... \tag{1}$$

The probability density function of θ is defined as:

$$g(\theta) = \frac{\alpha^{\beta}}{\Gamma(\beta)}\theta^{\beta-1}e^{-\alpha\beta} \ , \ \theta > 0 \tag{2}$$

Then the joint distribution of Y and θ is the product of $P(Y|\theta)$ and $g(\theta)$ as follows:

$$P(Y, \theta) = \frac{e^{-\theta}\theta^{y}}{y!} * \frac{\alpha^{\beta}}{\Gamma(\beta)}\theta^{\beta-1}e^{-\alpha\beta} \tag{3}$$

By summing out θ in (3), the unconditional distribution of Y is obtained as:

$$P(Y = y) = \binom{y + \beta - 1}{y}\left(\frac{\alpha}{\alpha + 1}\right)^{\beta}\left(\frac{1}{\alpha + 1}\right)^{y} \ , \ y = 0, 1, 2, ... \tag{4}$$

Which is of the form of negative binomial function, and with the re-parametrization, it can be re-written as follows:

$$P(Y = y_i|X_i) = \frac{\Gamma(y_i + 1/\alpha)}{\Gamma(y_i + 1)\Gamma(1/\alpha)}\left(\frac{1}{1 + \alpha\mu}\right)^{\frac{1}{\alpha}}\left(\frac{\alpha\mu}{1 + \alpha\mu}\right)^{y_i} \tag{5}$$

With X_i representing the vector of explanatory variables, where $\mu = exp(X_i\beta)$, β being the vector of coefficients and α is the over-dispersion parameter. Where μ represent the average number of reported contacts, also represents the Poisson variance and α representing the extra-dispersion parameter in Negative binomial model. Usually α is unknown, so estimating it helps to summarize the extend of extra-dispersion in our data.

Therefore the statistical model is given by:

$$log(\mu_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip} \ , \ i = 1, 2, .., N \tag{6}$$

The maximum likelihood estimation gives:

$$E[Y] = \mu = \exp(x_i'\beta) \ , \quad var[Y] = \mu + \frac{1}{\alpha}\mu^2$$

Negative binomial regression assumes the independence of individual observations, the multiplicative effects of independent variables, and that the conditional means are not equal to the conditional variances.

The model was fitted using glm.nb() function from MASS package. We considered variables associated with reported contacts at $p < 0.05$ and retained them in the models if they resulted in a reduction of the Alkaike Information Criterion (AIC).

We also used frequency tables and graphical representation to have insight on the distribution of contacts reported by 1043 and 945 participants during two waves, by different characteristics: Age category and gender of contacts, location of contact, vicinity of contacts, relationship with the contact, and contact duration. We computed mean, and IQR to compare the distribution of contacts during two waves.

4.3 Age-specific contact patterns

We analysed the age specific mixing pattern by constructing matrix **C** representing the average number of reported contacts between participants in age group j and individuals in age group i.

Let y_{ij} denote the total number of contacts with individual in age group i reported by participants in age group j, and let n_j denote the sample size of age group j, then the average number of daily contact per participant was calculated as follow:

$$m_{ij} = \frac{y_{ij}}{n_j}$$

To account for the sampling variability, 100 matrices were generated using bootstrap. The age groups used were 0–17, 18–49 and 50 and over. Separate social contact matrices were also derived for all contacts, type of contact(physical vs non-physical contacts), and location (household and outside household).

4.4 Software

Data analysis was performed using R version 4.0.3. The MASS package were used to fit the negative binomial model and social contact matrices were derived using SocialMixr package and the Socrates tool.

5 Results

5.1 Data

Figure 2 illustrate the process of data accessibility of the social mixing survey. The target sample was 2400 individuals. During the first wave, contact information recorded by 1720 participants were available, however only 1043 participants were able to be matched, with their recorded information about age and gender in the HDSS database, due to data management difficulties that will be rectified in the near future. These were then used in the description of contact characteristics and social mixing pattern. Out of these 1043 participants, only 244 participants were able to be matched with their recorded information regarding social-demographic characteristics. The 224 participants were therefore used in modeling average number of contacts. The same process was performed during the second wave as illustrated in the figure.

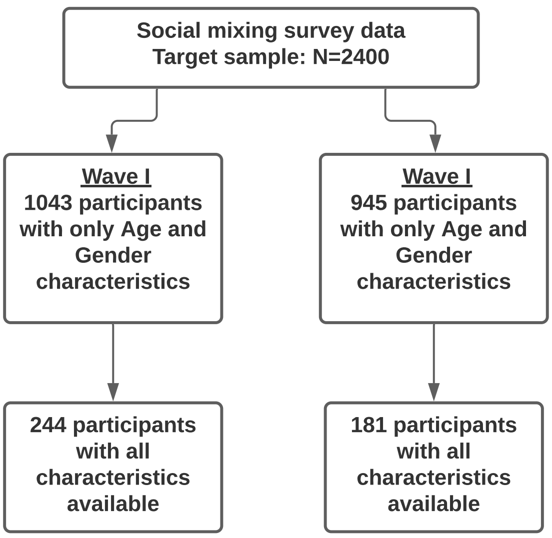

Figure 2: Data accessibility

5.2 Study population

During the first wave, out of 1043 participants who participated in the study, demographic characteristics was available for 244 participants, and for 181 out of 945 participants during the second wave (table 3). These were distributed among 105(46.9%) females and 119(53.1%) males during the fist wave, 89(49.2%) females and 92(50.8%) males during the second wave. During the first wave, more than half of participants were farm workers (59.4%, 133 participants), 25.4%(57 participants) were students, 4.5%(10 participants) were office worker, 0.9%(2 participants) were manual worker and 9.8%(22 participants) were unemployed. Overall 27.2% (61 participants) reported having primary level of education, 50.5% (113 participants) secondary level of education, 2.2%(5 participants) had university level of education and 20.1%(45 participants) reported having no education level. Christian participation made up 57.1%(128 participants) of the total, and Nianga ethnicity accounting for 63.4%(142 participants) of the total.

Table 3: Participant characteristics and mean numbe of contact(IQR)

Variables		Wave I		Wave II	
		N = 224	**Mean(IQR)**	**N = 181**	**Mean(IQR)**
Gender	Female	105(46.9%)	4.56(3-6)	89(49.2%)	4.72(3-6)
	Male	119(53.1%)	4.07(2-6)	92(50.8%)	4.38(2-6)
Age category	0-17	71(31.7%)	4.62(2-6)	51(28.2%)	4.69(3-6)
	18-49	78(34.8%)	4.56(3-6)	65(35.9%)	4.78(3-6)
	50+	75(33.5%)	3.72(2-5)	65(35.9%)	4.20(2-6)
Day of contact	Monday	8(3.6%)	3.87(2-5)	25(13.8%)	4.60(2-7)
	Tuesday	31(13.8%)	4.77(2-7)	13(7.2%)	4.15(2-6)
	Wednesday	60(26.8%)	4.18(2-6)	23(12.7%)	4.13(2-6)
	Thursday	56(25.0%)	3.89(2-5)	36(19.9%)	4.47(3-6)
	Friday	36(16.1%)	5.00(2-7)	47(25.9%)	5.0(4-6)
	Saturday	26(11.6%)	4.04(3-6)	11(6.1%)	3.18(2-5)
	Sunday	7(3.1%)	4.28(3-4)	26(14.4%)	4.92(3-6)
Occupation	Farm worker	133(59.4%)	4.14(2-6)	118(65.2%)	4.20(3-6)
	Manual worker	2(0.9%)	4.00(3-4)	-	-
	Office worker	10(4.5%)	5.10(1-8)	10(5.5%)	6.4(5-6)
	Student	57(25.4%)	4.59(2-6)	38(21.0%)	5.0(4-7)
	Unemployed	22(9.8%)	4.16(2-5)	15(8.3%)	4.86(4-5)
Level of Schooling	Primary	61(27.2%)	4.03(3-5)	49(27.1%)	4.18(2-6)
	Secondary	113(50.5%)	4.66(3-6)	92(50.8%)	4.89(3-6)
	University	5(2.2%)	1.4(1-2)	4(2.2%)	4.50(3-6)
	None	45(20.1%)	4.07(4-6)	36(19.9%)	4.17(3-5)
Religion	Christians	128(57.1%)	4.23(2-6)	107(59.1%)	4.34(2-6)
	Non-christians	96(42.9%)	4.38(3-6)	74(40.9%)	4.83(3-7)
Ethnie	Ndibu	52(23.1%)	3.83(2.85)	42(2-5%)	4.47(2-6)
	Ngombe	15(6.7%)	3.80(2-5)	11(6.1%)	4.09(2-5)
	Nianga	142(63.4%)	4.43(3-6)	116(64.1%)	4.67(3-7)
	Other	15(6.7%)	5.20(3-5)	12(6.6%)	4.00(3-5)

5.3 Participants contacts

Average number of contacts

The average number of contacts and IQR during the first and the second wave are displayed in table 3. During the first wave, elderly participants reported the lowest average number of contacts of 3.72(IQR:2-5) compared to 4.62(2-6) for young and 4.56(3-6) for adults. Elderly contacts increased to 4.20(2-6) contacts during the second wave while young and adults had 4.69(3-6) contacts and 4.78(3-6) contacts on average. Participants who reported contacts on Friday, reported more contacts on average(5.00) compared to those reported having contacts on other days in both waves. During the first and the second waves, office workers reported a higher number of contacts than others. Participants with university as their level of education reported the lowest average of 1.4(1-2) contact during the first wave and 4.50(3-6) contacts during the second wave.

	Estimate	Exp(Estimate)	Std. Error	P-value
Intercept	1.43	4.18	0.11	<2e-16
Young	0.08	1.08	0.10	0.44
Adult	Ref	-	-	-
Elderly	-0.16	0.85	0.10	0.10
Primary	-0.05	0.95	0.12	0.65
Secondary	0.15	1.16	0.11	0.17
University	-0.97	0.38	0.42	0.02 *
None	Ref	-	-	-

Table 4: Parameter estimates

The negative binomial regression was used to estimate the mean contacts as a function of the different covariates of interest. The first model included all variables with AIC equal to 1046, which later was simplified by considering variables that significantly reduce the AIC of the model. Only education level was found to be significant in lowering the AIC of the model, and individuals who claimed having a university level of education reported significantly lower contacts on average than those who did report not having any level of education. Although age category did not show a statistical effect to the model, it was included in the model, because it has been established to be an important variable that has an influence on the number of contacts a participant makes every day. The final model included age category and education level of participant. The estimates of the final model are displayed in table 4. This model estimated the over-dispersion parameter to be 9.6. (Std. Err.: 2.86), and has the lowest AIC equal to 1033.9.

Contacts characteristics

The distribution of contacts during the first and the second wave is presented in table 5. A total of 4341 contacts were reported by 1043 participants during the fist wave while 945 participants reported 2363 during the second wave. During the first and second wave, respectively, 67.9% and 63.3% of all contacts were physical(thus involving skin-to-skin contact or touch). Contacts within households accounted for 74.7% and 80.9% of all contacts during wave1 and wave2 respectively. Outside households, contacts mostly occurred in the field, river and other households. Also more than 86% of all interactions, were established with household members, 6% with other relative and 3.7% with friends. Apart from the reduction of the sample size during the second wave, due to drop-outs, overall the distribution of contacts across different characteristics is consistent in the two waves.

Table 5: Distribution of reported contact and proportion in different factors

		Wave1		**Wave2**	
Variables	**Categories**	**N=4341**	**Prop**	**N=2363**	**Prop**
Gender	Male	2219	51.1%	1125	48.6%
	Female	2122	48.9%	1238	52.4%
Age category	[0-17]	2210	50.9%	1213	51.3%
	[18-49]	1596	36.8%	855	36.2%
	50+	535	12.3%	295	12.5%
Location type	Household	3244	74.7%	1912	80.9%
	Other household	170	3.9%	42	1.8%
	Field	626	14.4%	229	9.7%
	River	163	3.8%	79	3.3%
	Shop	33	0.8%	3	0.1%
	Work	30	0.7%	13	0.6%
	School	3	0.1%	-	-
	Transport Wewa	4	0.1%	2	0.1%
	Transport bus	5	0.1%	-	-
	Place of worship	28	0.6%	31	1.3%
	Place of leisure	6	0.1%	7	0.3%
	Other	29	0.7%	45	1.9%
Type of Relation	Household member	3771	86.9%	2073	87.7%
	Other relative	268	6.2%	129	6.0%
	Colleague	14	0.4%	10	0.4%
	Schoolmate	1	0.0%	-	-
	Friend	161	3.7%	88	3.7%
	Neighbour	75	1.7%	27	1.1%
	Boy/girlfriend	2	0.0%	2	0.1%
	0ther	49	1.1%	34	1.4%
Vicinity	Within village	4206	96.9%	2292	97%
	Outside village	127	2.9%	66	2.8%
	Outside kimpese	8	0.2%	5	0.2%
Duration	<5min	723	16.7%	325	13.8%
	5-14mins	85	2.0%	85	3.6%
	15mins-1hr	170	3.9%	119	5.0%
	1-4hrs	444	10.2%	174	7.4%
	>4hrs	2919	67.2%	1660	70.2%
Type of Contact	Physical	2949	67.9%	1496	63.3%
	Non-physical	1392	32.1%	867	36.7%

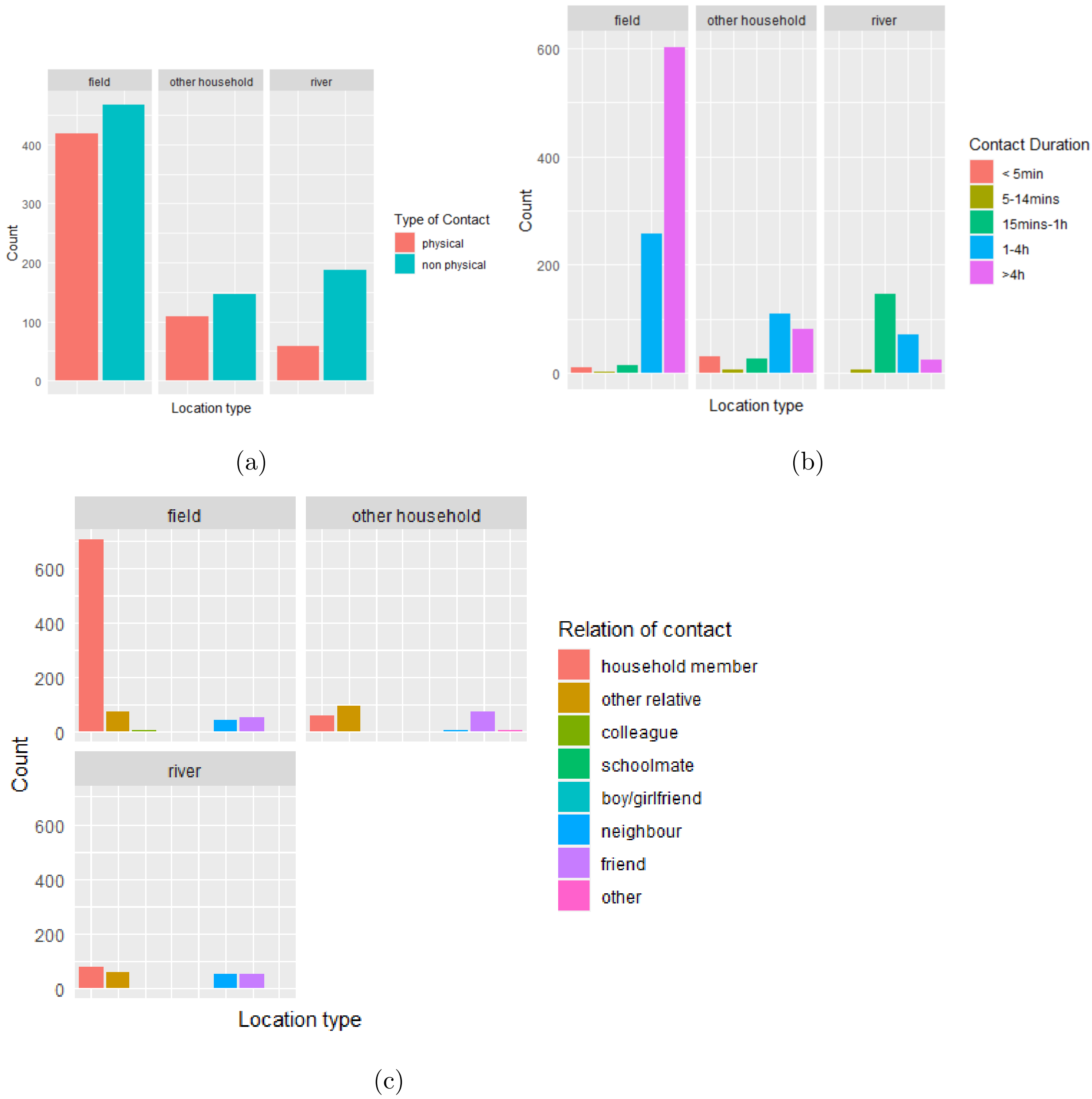

Figure 3: Distribution of reported contact number in top three outside household locations by: (a) Type of contact, (b) Duration of contact and (c) Relation with contact.

During the second wave, when measures were re-installed, there was a decrease in the number of contacts outside household such as field(from 14.4%

to 9.7%), other household(from 3.9% to 1.8%) and shops(from 0.8% to 0.1%) but increase in number of household contacs(from 74.7% to 80.9%). Since the majority of contacts outside households occurred in the field, river, and other household, we focused on contacts reported in these locations. Figure 3 illustrate the type of contacts, duration of those contacts, and with whom these contacts were made at these top three outside household locations during the first and second wave. The majority of contacts conducted in the field were non-physical, and lasted more than an hour. Also most of contacts in the field were made with household members.

At River places, more contacts were non-physical, with 15 minutes to 1 hour duration, and participants mostly interacted with household members and other relatives. At other household locations, participants interacted with other relatives, friends, and household members. It is an interesting finding that the majority of interactions were established with household members even while they were outside household locations. These findings were also observed during the second wave(table 7 and figure 8).

Table 6: Summary statistics of reported contacts

	Wave1		**Wave2**	
	Mean	**IQR**	**Mean**	**IQR**
All contacts	4.19	2-6	4.41	3-6
Physical	3.18	1-5	3.25	1-5
Non physical	3.05	2-4	3.219	1-4
Household	3.21	2-4	3.66	2-5
Outside Household	2.31	1-3	2.13	1-3
Within village	4.09	2-6	4.33	3-6
Outside village	2.56	2-3	2.62	2-3

Table 6 displays the mean and IQR of contact distribution during two waves.

Overall, the mean of reported contacts was 4.083 and 4.207 during the first and the second wave respectively. The average number of reported contacts within household was 3.207 and 3.656 with an avearage of 2.307 and 2.127 outside household for wave 1 and wave 2 respectively. Contacts outside village(2.56 and 2.62) were on average smaller than contacts occurred within village(4.09 and 4.33).

5.4 Age-specific contact patterns

Figure 4, 5, 6 and figure 7 shows the age-specific contact matrix of all contacts, physical contacts, household contacts and outside household contacts respectively during first wave(left) and second wave(right).

(a) Wave 1 (b) Wave 2

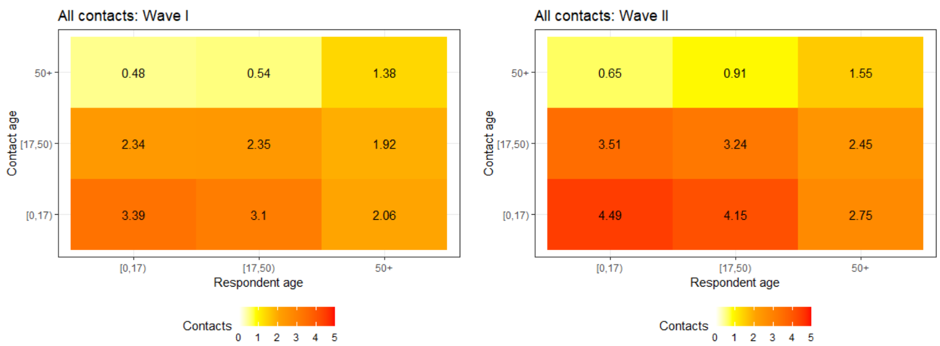

Figure 4: All contacts

(a) Wave 1 (b) Wave 2

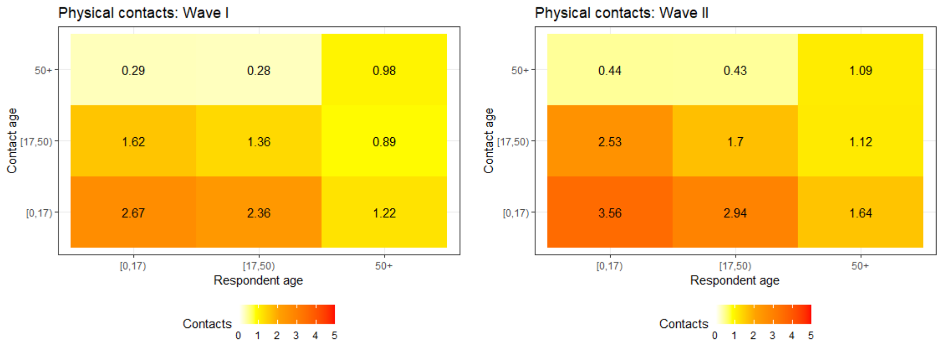

Figure 5: Physical contacts

For all contacts, during two waves, we observed the inter-generational mixing between young and adult, with adults interacting on average with 3.305([2.95,3.59] 95%CI) young individuals during the first wave, which significantly increased to 4.256([3.77,4.732] 95%CI) during the second wave. This inter-generational mixing was also observed in physical contacts and contacts occurred within households. This might be due to the high proportion of contacts occurred home which are mostly physical contacts.

(a) Wave 1 (b) Wave 2

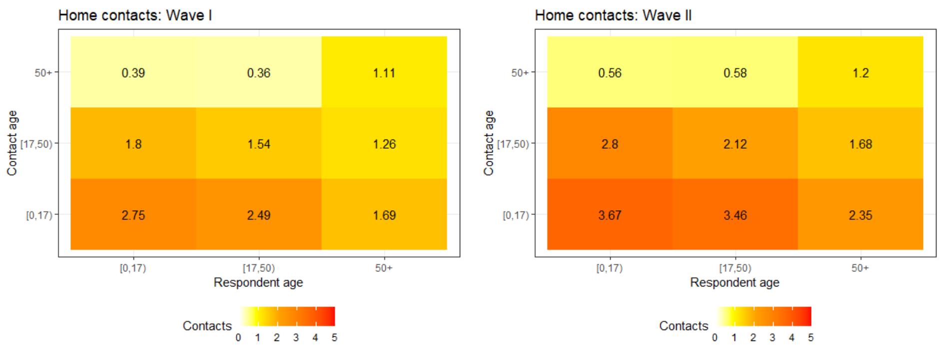

Figure 6: Household contacts

(a) Wave 1 (b) Wave 2

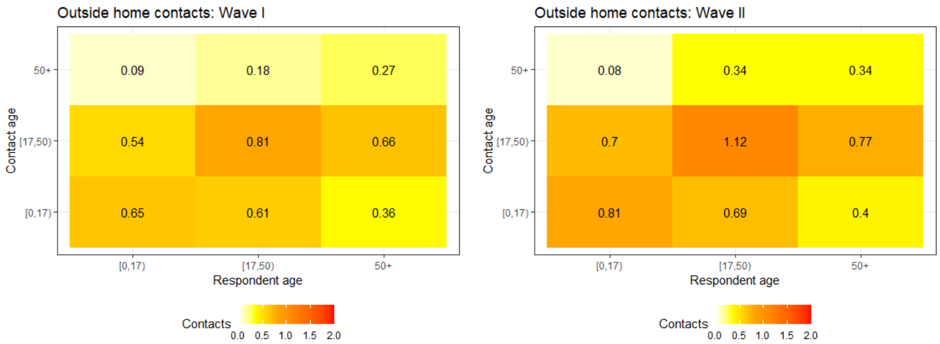

Figure 7: Outside household contacts

The mixing pattern outside household, was observed to be the mixture of inter-generational mixing and a moderate assortative mixing, mostly pronounced in adult population. The average number of contacts between and within age groups outside household, were relatively lower to those occurred within households. For instance during the first wave, adults interacted on average with 0.68[0.54,0.86] young individual outside household, but 2.64[2.31,2.93] young individual within household. Along with the social mixing survey, we estimated the sero-prevalence of COVID-19 from sero-survey data obtained during the first wave. Out of 1709 tested individuals, the test confirmed the presence of antibodies in 32 participants, from which the prevalence of COVID-19 disease was estimated to be 2.1%.

6 Discussion

A sero-survey and social mixing survey were conducted in Kimpese Health demographic area, to understand the social mixing behaviour in rural sub-saharan area, relevant to the transmission of COVID-19 and other respiratory disease, and their association with the sero-prevalence of COVID-19. This study reports the analysis results for two waves during the lifting and re-installement of restrictive measures. The negative binomial model was used to evaluate the relationship between the average number of reported contacts and participants socio-demographic characteristics. Age-specific matrices were constructed to describe the contact patterns in this population during two waves.

There was little but significant increase in average number of contacts, with adults interacting on average with 3.31([2.95,3.59] 95%CI) young individuals during the first wave, which increased to 4.26([3.77,4.73] 95%CI) during the second wave. During the second wave, when measures were re-installed, there was a decrease in the number of contacts outside household but an increase in number of household contacts, which is expected during lockdown [20]. This could explain the increase of reported number of contacts even during the re-installment of restrictive measures.

Most contacts occurred within households, which accounted for 75% and 81% of all contacts during the first and the second wave respectively, followed by the field(14.4% and 9%) and river(3.8% and 3%). Overall, the majority of contacts(86.9% in wave 1 and 87.7% in wave 2), were made with household members. Participants interacted with household members within, but also outside their household(in the field and river). This could explain the inter-generational mixing observed, highly observed between young and adult participants. This mixing was also observed in rural area of Zimbabwe [15] and

Kenya [12]. In Kenya, the study was conducted in 2014, in semi-urban and rural areas, with 568 participants from the Kilifi Health and Demographic Surveillance System, where the seasonal economic activities are fishing, farming, agriculture and tourism. In Zimbabwe, the study was conducted in 2013 with 554 participants from the peri-urban site and 691 participants from the rural site which is a predominantly farming area. Outside household, we observed inter-generational mixing and high mixing among adults.
Contacts which occurred outside villages(but within Kimpese area) accounted for 2.9% and 2.8% of all contacts during the first and the second wave respectively, whereas contacts that occurred outside Kimpese area accounted for 0.2% during the first and the second wave. Additionally to this low number of contacts, Kimpese area is not directly connected to Kinshasa(193km from Kimpese to Kinshasa) which is currently the epicenter of COVID-19 outbreak. Therefore, this might help to explain the low sero prevalence(2.1%) observed in this area.

Two activities are dominant in this area: Agriculture(59.4% are farm workers) and school(25.4% are students). It is also uncommon to find places of leisure, bars, restaurants and shops in this area. This remoteness of Kimpese area could be the reason why the average number of contacts(4.19) when restrictive measures in DRC were lifted(first wave), was similar to the average number of contacts estimated (4.41) in European setting(Belgian population), when lockdown was implemented [4]. However, this was relatively lower than the average number of reported contacts(mean = 18, IQR 7–23) in Kenya during lockdown, in 2020 among a sample of adults from five informal settlements in urban and peri-urban areas around Nairobi [18]. This is probably because in urban and peri-urban areas, there is more interactions compared to rural areas.

Participant gender, occupation, religion, ethnicity and day of contact did not show a significant relation with the average number of contacts. The negative binomial model that explains the variability in the average number contacts, included age category and level of education variables. Age of participant was also found to be statistically significant for the mean reported contacts in studies conducted in similar settings(rural sub-saharan areas) [15] [12] [5]. School size was observed to also affect the reported number of contacts, however in our study primary and secondary schools were closed.

7 Strength and limitation of the study

This study added knowledge to our understanding of mixing behavior in rural Sub-Saharan Africa, which is useful for the disease transmission modeling and intervention optimization. It also provided data on how mixing behavior changes in this population when various restriction measures were adopted. One of the limitations of our study was data availability. First, socio-demographic characteristics were available for only 224 participants living in the Northern region of Kimpese, out of 1043 participants. Therefore the analysis of the relationship between the average number of reported contacts, and participants socio-demographic characteristics, was performed on a portion of the data. This is not expected to have a significant impact on our findings, since we have observed the comparability of the average number of contacts between data from all participants and data from chosen participants (9). Nonetheless, once full data become available in the future, the analysis of these data should follow.

Second, our analysis did not incorporate sampling weights based on participant age distribution , because DRC population density from World Population Prospects (WPP) is unreliable as it has not been updated in a long time, and sampling weights from HDSS data were not available at the time of analysis. These might be taken into account in future analyses, using observations from the full HDSS data. Lastly, the simulation of COVID-19 disease dynamic using the observed mixing patterns could not be accomplished due to the low COVID-19 incidence. This goal may be pursued in the future if there is sufficient COVID-19 occurrence.

8 Recommendation and Conclusion

Future studies may compare the mixing behavior in DRC's rural, semi-urban, and urban areas, since their daily activities and cultures may differ, and the disease transmission is likely to spread more in semi-urban and urban regions than in rural areas. We also recommend an additional social mixing survey to be conducted when schools reopen, to evaluate the effect of schools on reported number of contacts, which was regarded to be the second main activity in this rural area, and expand the age bands of individuals for instance to pre-school(0-6years), Primary(7-13years), secondary(14-20years), adults(21-50years) and elderly(50+ years), to capture in-depth information on age-specific mixing behaviour.

We have conducted a study to explore the social mixing patterns relevant to the transmission of SARS-COV-2 in rural sub-saharan area. To our knowledge, this is the first of its kind conducted in DRC and one of the few conducted in rural sub-saharan area. Having this data fills the gap of knowledge about the mixing behaviour in sub-saharan countries. The findings suggest the high proportion of contacts occurring within households and the inter-generational mixing, highly observed between adults and young generations, and a relative low average number of contacts, compared to other rural sub-saharan countries. This study's conclusions can be applied to other places with comparable situations.

References

[1] Covid-19 coronavirus pandemic. `https://www.worldometers.info/coronavirus/`. Accessed: 2021-04-15.

[2] Naming the coronavirus disease (COVID-19) and the virus that causes it. `https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance`. Accessed: 2021-04-15.

[3] Roy M Anderson and Robert M May. *Infectious diseases of humans: dynamics and control*. Oxford university press, 1992.

[4] Pietro Coletti, James Wambua, Amy Gimma, Lander Willem, Sarah Vercruysse, Bieke Vanhoutte, Christopher I Jarvis, Kevin Van Zandvoort, John Edmunds, Philippe Beutels, et al. Comix: comparing mixing patterns in the belgian population during and after lockdown. *Scientific reports*, 10(1):1–10, 2020.

[5] O le Polain de Waroux, Sandra Cohuet, Donny Ndazima, AJ Kucharski, Aitana Juan-Giner, Stefan Flasche, Elioda Tumwesigye, Rinah Arinaitwe, Juliet Mwanga-Amumpaire, Yap Boum, et al. Characteristics of human encounters and social mixing patterns relevant to infectious diseases spread by close contact: a survey in southwest uganda. *BMC infectious diseases*, 18(1):172, 2018.

[6] Peter J Dodd, Clare Looker, Ian D Plumb, Virginia Bond, Ab Schaap, Kwame Shanaube, Monde Muyoyeta, Emilia Vynnycky, Peter Godfrey-Faussett, Elizabeth L Corbett, et al. Age-and sex-specific social contact patterns and incidence of mycobacterium tuberculosis infection. *American journal of epidemiology*, 183(2):156–166, 2016.

[7] S. Falahi and A. Kenarkoohi. Transmission routes for sars-cov-2 infection: review of evidence. *New Microbes and New Infections*, 38:100778, 2020.

[8] Seth Flaxman, Swapnil Mishra, Axel Gandy, H Juliette T Unwin, Thomas A Mellan, Helen Coupland, Charles Whittaker, Harrison Zhu, Tresnia Berah, Jeffrey W Eaton, et al. Estimating the effects of non-pharmaceutical interventions on covid-19 in europe. *Nature*, 584(7820):257–261, 2020.

[9] Joseph M Hilbe. *Negative binomial regression.* Cambridge University Press, 2011.

[10] Christopher I Jarvis, Kevin Van Zandvoort, Amy Gimma, Kiesha Prem, Petra Klepac, G James Rubin, and W John Edmunds. Quantifying the impact of physical distance measures on the transmission of covid-19 in the uk. *BMC medicine*, 18:1–10, 2020.

[11] Simon P Johnstone-Robertson, Daniella Mark, Carl Morrow, Keren Middelkoop, Melika Chiswell, Lisa DH Aquino, Linda-Gail Bekker, and Robin Wood. Social mixing patterns within a south african township community: implications for respiratory disease transmission and control. *American journal of epidemiology*, 174(11):1246–1255, 2011.

[12] Moses Chapa Kiti, Timothy Muiruri Kinyanjui, Dorothy Chelagat Koech, Patrick Kiio Munywoki, Graham Francis Medley, and David James Nokes. Quantifying age-related rates of social contact using diaries in a rural coastal population of kenya. *PloS one*, 9(8):e104786, 2014.

[13] Adam J Kucharski, Timothy W Russell, Charlie Diamond, Yang Liu, John Edmunds, Sebastian Funk, Rosalind M Eggo, Fiona Sun, Mark Jit, James D Munday, et al. Early dynamics of transmission and control of covid-19: a mathematical modelling study. *The lancet infectious diseases*, 20(5):553–558, 2020.

[14] You Li, Harry Campbell, Durga Kulkarni, Alice Harpur, Madhurima Nundy, Xin Wang, Harish Nair, Usher Network for COVID, et al. The temporal association of introducing and lifting non-pharmaceutical interventions with the time-varying reproduction number (r) of sars-cov-2: a modelling study across 131 countries. *The Lancet Infectious Diseases*, 21(2):193–202, 2021.

[15] Alessia Melegaro, Emanuele Del Fava, Piero Poletti, Stefano Merler, Constance Nyamukapa, John Williams, Simon Gregson, and Piero Manfredi. Social contact structures and time use patterns in the manicaland province of zimbabwe. *PloS one*, 12(1):e0170459, 2017.

[16] Joël Mossong, Niel Hens, Mark Jit, Philippe Beutels, Kari Auranen, Rafael Mikolajczyk, Marco Massari, Stefania Salmaso, Gianpaolo Scalia Tomba, Jacco Wallinga, et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med*, 5(3):e74, 2008.

[17] Maria Nicola, Zaid Alsafi, Catrin Sohrabi, Ahmed Kerwan, Ahmed Al-Jabir, Christos Iosifidis, Maliha Agha, and Riaz Agha. The socio-economic implications of the coronavirus and covid-19 pandemic: a review. *International journal of surgery*, 2020.

[18] Matthew Quaife, Kevin Van Zandvoort, Amy Gimma, Kashvi Shah, Nicky McCreesh, Kiesha Prem, Edwine Barasa, Daniel Mwanga, Beth

Kangwana, Jessie Pinchoff, et al. The impact of covid-19 control measures on social contacts and transmission in kenyan informal settlements. *BMC medicine*, 18(1):1–11, 2020.

[19] Erigene Rutayisire, Gerard Nkundimana, Honore K Mitonga, Alex Boye, and Solange Nikwigize. What works and what does not work in response to covid-19 prevention and control in africa. *International Journal of Infectious Diseases*, 97:267–269, 2020.

[20] Juanjuan Zhang, Maria Litvinova, Yuxia Liang, Yan Wang, Wei Wang, Shanlu Zhao, Qianhui Wu, Stefano Merler, Cécile Viboud, Alessandro Vespignani, et al. Changes in contact patterns shape the dynamics of the covid-19 outbreak in china. *Science*, 368(6498):1481–1486, 2020.

[21] Na Zhu, Dingyu Zhang, Wenling Wang, Xingwang Li, Bo Yang, Jingdong Song, Xiang Zhao, Baoying Huang, Weifeng Shi, Roujian Lu, et al. A novel coronavirus from patients with pneumonia in china, 2019. *New England journal of medicine*, 2020.

Appendix

Tables and Figures

Table 7: Number of Contacts, Outside Household for the two waves

		WAVE I		WAVE II	
Variable	**Categories**	**Tot. contacts**	**Prop**	**Tot. contacts**	**Prop**
Sex	Female	530	0.48	240	0.53
	Male	567	0.52	211	0.47
Age category	[0-17]	453	0.41	164	0.36
	[18-49]	528	0.48	206	0.46
	50+	116	0.11	81	0.18
Location type	Other household	170	0.15	42	0.09
	Field	626	0.57	229	0.51
	River	163	0.15	79	0.17
	Shop	33	0.03	3	0.01
	Work	30	0.03	13	0.03
	School	3	0.00		
	Transport Wewa	4	0.00	2	0.00
	Transport bus	5	0.00		
	Place of worship	28	0.03	31	0.07
	Place of leisure	6	0.01	7	0.02
	Others	29	0.03	45	0.10
Type of relation	Household member	619	0.56	228	0.50
	Other relative	185	0.17	82	0.18
	Collegue	13	0.01	10	0.02
	Schoolmate	1	0.00		
	Friend	159	0.15	84	0.19
	Neighbour	72	0.07	26	0.06
	Boy/girlfriend	2	0.00	0	0.00
	Other	46	0.04	21	0.05
Duration	<5mins	39	0.04	9	0.02
	5-14mins	16	0.01	13	0.03
	15mins-1h	164	0.15	69	0.15
	1-4h	378	0.34	124	0.28
	>4h	500	0.46	236	0.52
Type of contact	Physical	462	0.42	150	0.33
	Non physical	635	0.58	301	0.67

Table 8: Parameter estimates for the Negative binomial full model

	Estimate(exp(estimate))	Std. Error	Pr(> $\|z\|$)
Intercept	1.72343(5.603)	0.30498	$1.6e-08$
Male	-0.206065(0.814)	0.08342	0.0135
Young	0.22253(1.249)	0.12711	0.0800
Adult	0.19297(1.213)	0.09775	0.0484
Monday	-0.08153(0.922)	0.30094	0.7864
Tuesday	0.20063(1.222)	0.23945	0.4021
Wednesday	-0.03252(0.968)	0.22971	0.8874
Thursday	0.01019(1.010)	0.23005	0.9647
Friday	0.08357(1.087)	0.23991	0.7276
Saturday	-0.02088(0.979)	0.24743	0.9327
Primary	-0.03870(0.9620)	0.14510	0.7897
Secondary	0.16668(1.181)	0.12485	0.1819
University	-0.99524(0.369)	0.42472	0.0191
Farm worker	-0.07246(0.930)	0.16069	0.6521
Manual worker	-0.03731(0.963)	0.44455	0.9331
Office worker	0.41273(1.511)	0.24440	0.0913
Student	-0.05697(0.945)	0.18087	0.7528
Ndibu	-0.39202(0.676)	0.17675	0.0266
Ngombe	-0.54555(0.579)	0.22220	0.0141
Nianga	-0.27669(0.758)	0.15670	0.0774
Christians	-0.11037(0.895)	0.08170	0.1767

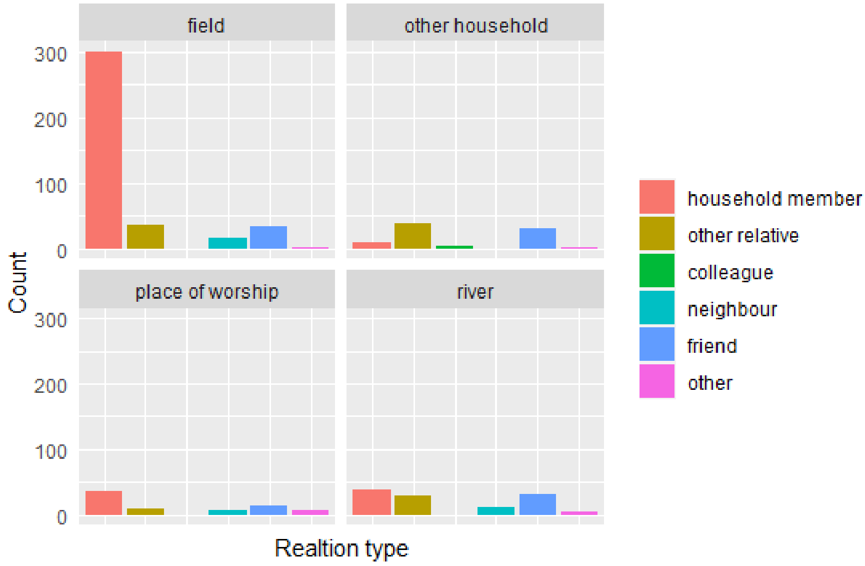

Figure 8: Distribution of contacts and their relation with participants in the top four location with high number of contacts

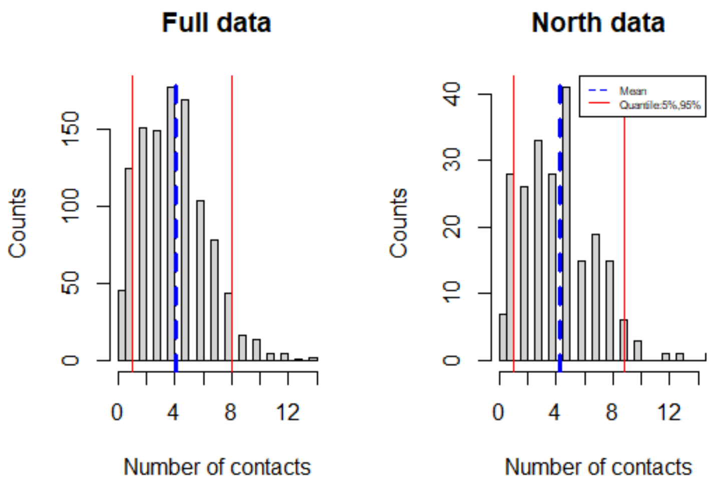

Figure 9: Comparison of full data and data with participants from Notrh of Kimpese

R codes

```
# Fitting negative binomial model

full_model <- glm.nb(contact_number ~ sex + age + occupation +
                                      education + ethnicity +
                                      religion + day_of_contact,
                                      data = data_north)

summary(full_model)

# Constructing age-specific matrices

age_breaks <- c(0,17,50,100)

surv_obj <- get_survey_object(country="Congo (Preliminary)",
                              daytype="All contacts",
                              touch="All contacts",
                              duration="All contacts",
                              gender="All",
                              cnt_location=opt_location,
                              wave="All waves",
                              quiet = TRUE)

matrix_Congo <- contact_matrix(survey=surv_obj,age.limits = age_breaks,
                               n=1,
                               missing.contact.age="sample",
                               estimated.contact.age="sample",
```

```
                                  weigh.age=FALSE,
                                  weigh.dayofweek=TRUE,
                                  weight.threshold = 3)

plot_cnt_matrix(matrix_Congo$matrix)
```