DOI http://dx.doi.org/10.20947/S0102-3098a0241

ORIGINAL ARTICLE

A study on social contact rates relevant for the spread of infectious diseases in a Brazilian slum

Sílvio Segundo Salej Higgins[*]
Adrian Pablo Hinojosa Luna[**]
Reinaldo Onofre dos Santos[***]
Andreia Maria Pinto Rabelo[****]
Maíra Soalheiro[*****]
Vanessa Cardoso Ferreira[******]

Inspired by the POLYMOD study, an epidemiological survey was conducted in June 2021 in one of the most densely populated and socially vulnerable sectors of Belo Horizonte (Brazil). A sample of 1000 individuals allowed us to identify, within a 24-hour period, the rates of social contacts by age groups, the size and frequency of clique in which respondents participated, as well as other associated sociodemographic factors (number of household residents, location of contact, use of public transportation, among others). Data were analyzed in two phases. In the first one, results between two SIR models that simulated an eight-day pandemic process were compared. One included parameters adjusted from observed contact rates, the other operated with parameters adjusted from projected rates for Brazil. In the second phase, by means of a log-lin regression, we modeled the main social determinants of contact rates, using clique density as a proxy variable. The data analysis showed that family size, age, and social circles are the main covariates influencing the formation of cliques. It also demonstrated that compartmentalized epidemiological models, combined with social contact rates, have a better capacity to describe the epidemiological dynamics, providing a better basis for mitigation and control measures for diseases that cause acute respiratory syndromes.

Keywords: Epidemiological survey. POLYMOD. Social contact rate. Cliques.

[*] Department of Sociology, Federal University of Minas Gerais (UFMG), Belo Horizonte-MG, Brazil (sisahi@yahoo.com; https://orcid.org/0000-0002-3573-0578).

[**] Department of Statistics, Federal University of Minas Gerais (UFMG), Belo Horizonte-MG, Brazil (adrhnj@gmail.com; https://orcid.org/0000-0002-8844-3062).

[***] Federal University of Juiz de Fora (UFJF), Juiz de Fora-MG, Brazil (reinaldosantos80@gmail.com; https://orcid.org/0000-0001-6762-9100).

[****] Interdisciplinary Group on Social Network Analysis, Federal University of Minas Gerais (UFMG), Belo Horizonte-MG, Brazil (rabelo.andreiamp@gmail.com; https://orcid.org/0000-0002-4015-4026).

[*****] Department of Statistics, Federal University of Minas Gerais (UFMG), Belo Horizonte-MG, Brazil (mairasoalheeiro@gmail.com; https://orcid.org/0000-0003-4099-4532).

[******] Consultant for the Inter-American Development Bank (IDB) and the United Nations Population Fund (UNFPA), Belo Horizonte-MG, Brazil (va.cafes@gmail.com; https://orcid.org/0000-0001-7011-7755).

Introduction

Mathematical modeling of infectious processes is not recent, having been used since the late 17th century to understand the dynamics of contagion and to support control and mitigation strategies. However, the use of frequency of social contacts in these models, even today, is not usually present. This is largely explained by the lack of adequate estimates that reflect the reality of contacts in each population under study. Recent advances in epidemiological data collection have shown that the predictive and explanatory power of models is enhanced through the quantification of social contacts (PREM *et al.*, 2021).

In epidemiological studies, it is common to use systems of differential equations, called compartmental models, which includes the SIR model (susceptible, infected and recovered) (KEELING; ROHANI, 2008). The construction of these traditional epidemiological models requires the estimation of certain parameters of the system to adequately capture the dynamics of a disease in a population. One of these parameters is the daily contagion rate (β), which indicates how many secondary cases an infectious individual generates, per day, in a susceptible population and measures the rate of interaction between the susceptible and infected compartments of the model. Usually, when estimating this rate, the contexts in which the contacts are taking place are not measured,[1] using the assumption that the parameter is the same, or converges towards this, in all population subgroups or social contexts in which contacts occur. In this regard, it's possible to treat the case of groups of different age as compartments in a SIR like model, modifying the estimated daily contagion rate to construct an age specific rate using the product between the rate of social contacts and the rate of contagion of the pathogen, given the occurrence of contact. Some examples of the use of contact rates for compartmental epidemiological models are present in the work of Chin *et al.* (2021) and Prem *et al.* (2021).

The contact rate, understood as the average number of daily contacts of a population segment, is known to be a function of both individual attributes and the environment in which the contacts are made. In a pandemic scenario, the promotion of non-pharmacological actions, such as restrictions on the function of economic activities to promote greater social distancing, requires an understanding of how contacts develop, a *sine qua non* condition for the construction of reliable epidemiological scenarios and the evaluation of public policies.

The aim of this article is to present the significance of contact rates as a vital instrumental measure for epidemiological analysis. To this end, we present the general aspects of designing a contact data collection survey and then demonstrate how epidemiologically relevant contact rates can be used to improve traditional epidemiological models. Furthermore, we present an analysis of the socio-demographic constraints of epidemiologically relevant contact rates. These were obtained through field research carried out in a sector of the city of Belo Horizonte (Brazil) in June 2021. The contact rates collected

[1] In the case of COVID-19, see Yang *et al.* (2020) and Zhou *et al.* (2020).

also made it possible to simulate the dynamics of COVID-19 through the parameterization of a SIR model. Firstly, the paper describes the methodological field on epidemiological social surveys that aim to collect social contact rates, highlighting the main approaches with their challenges and limitations. Secondly, it presents the application of the former methodology in the universe of a slum community in Belo Horizonte Brazil, including sample size, strategy for data collection and post-stratification procedures. Thirdly, we present the statistics describing the social contact rates collected. Fourthly, aiming to test the heuristic power of social contact rates, we include a comparison of two SIR models, one informed with parameters that consider the social contact rates observed and another one using social contact rates projected for Brazil in international studies. Fifthly, via a log-lin model, we explore social determinants of social contact rate. Considering an epidemiological perspective, a proxy variable, density of cliques[2], was constructed to operationalize the social contact rate as a dependent variable. Finally, as for practical recommendations, we present the advantages of informing SIR models with social contact rates just like the identification of some relevant determinants of social contacts.

Development of survey methodology on social contacts

The first large-scale quantitative survey was carried out in 2008 on contact patterns relevant to respiratory and close-contact infections. The study Improving Public Health Policy in Europe through the Modeling and Economic Evaluation of Interventions for the Control of Infectious Diseases (POLYMOD) (MOSSON *et al.*, 2008) involved 7,290 people from eight European countries (Belgium, Germany, Finland, Great Britain, Italy, Luxembourg, Netherlands and Poland) and used the epidemiological diary to record participants' contacts in one day, providing data on different age groups and different interaction environments, such as school, home, work, among others.

Other smaller-scale research has explored patterns of social interaction to understand the transmission of infectious diseases. An example was conducted in the province of San Marcos, in the northern highlands of Peru, involving rural communities (GRIJALVA *et al.*, 2015). Another study carried out, called BBC Pandemic (KLEPAC *et al.*, 2020) was an innovative research experiment conducted through the Pandemic app, specially created to identify the human networks and behaviors that spread infectious diseases. Their data were used by researchers at the University of Cambridge and the London School of Hygiene and Tropical Medicine to build a map of social interactions in the UK. Recently, Chin *et al.* (2021) studied the contribution of age groups to the dynamics of SARS-COV 2 in the United States. To this end, they performed a longitudinal study with six-wave data from the Berkeley Interpersonal Contact Survey (BICS). They worked with social contact information collected between March 2020 and February 2021 in six metropolitan areas in the United States. Other

[2] In the relational sociological approach, cliques are social structures which actors are connected by cohesive ties (BURT, 1978).

studies regarding the action of subjects during a pandemic involve the measurement of mobility during the COVID-19 related to the dynamics of the subjects on spatial structures, whereas in this case we are interested in the patterns of contacts, especially the frequency of contacts between individuals belonging to different age groups. However, the structures of the patterns of mobility describe the dynamics of the epidemic on a mesoscopic scale, whereas with the contact process we see the phenomenon at the microscopic scale. The former approach has been preferred in several studies, one suggested by the anonymous referee (OLIVEIRA *et al.*, 2021) which uses Google COVID-19 Community Mobility data.

All these studies have contributed to filling the gap in the production of empirical data about social contacts relevant for modeling the dynamics of infectious disease transmission. In the same direction, the present work, a result of the research "Covid-19: epidemiological model that incorporates structures of social contacts" (funded by the Ministry of Health of Brazil, public notice MCTIC/CNPq/FNDCT/MS/SCTIE/Decit nº 07/2020), seeks to advance in the field of social contact research, and includes the main recommendations pointed out by Hoang *et al.* (2019) regarding sampling, instruments and data collection methods, especially with the elaboration of a complex sample design, taking into account, on the one hand, the relevant socio-demographic parameters and, on the other hand, the structural parameters of an unobserved network of contacts.

Sampling design and data collection

Data collection

For the estimation of social contacts, we implemented a survey in an impoverished sector of Belo Horizonte, capital of one the largest federated states of Brazil. Located in the center-south region of the city, the Aglomerado da Serra comprises a contiguous space of eight villages, located on the slopes of the Serra do Curral, an old urban occupation with a complex environmental degradation situation. It is an occupied area on the fringes of public planning, with a low-income population, in which the public power recognizes the need to organize the occupation, through housing programs, urbanization interventions and land regularization actions.

This population was chosen with the aim of observing the rates of social contact in a high vulnerability area. In this region, with a high population density, people share reduced housing units, rendering social distancing impractical. In addition, many houses present characteristics of precariousness and insalubrity, such as lack of adequate ventilation, poor sunlight and excess humidity. These factors increase housing insalubrity and reflect on people's health, especially children and the elderly (SILVEIRA, 2015).

The sample size calculation of the survey considered socio-demographic and network structure parameters, which, due to network sampling, resulted in a larger sample size than would be necessary for a conventional survey. This presented several challenges. Firstly, the unavailability of an updated demographic census (IBGE), as the last one available dates from 2010. According to this, the total population of Aglomerado da Serra was 38,405 inhabitants who lived in 10,900 households. Secondly, we estimate network parameters so we assume as a population target a large, but unobserved, network for this universe of people. To calculate the minimum size of nodes to get credible estimates for the network, we assume as a sampling target a complete, but unobserved, network for this universe of people. To do so, we simulated 500 networks of 5,000 and 10,000 nodes, using a test power of 8% and a significance level of 95%, which allowed us to estimate the average number of cliques (groups of people where all of them are in contact at the same time) of sizes 2, 3, 4 and 5 giving us a sample size of at least 1,000 nodes. We use the ergm and graphlets packages, built and made available in the R package. Both are part of the Statistical Network Analysis library of R (YAVEROGLU *et al.*, 2014; GJOKA *et al.*, 2014; HUNTER *et al.*, 2008). Given the total number of individuals in the sample and the plausible frequency of cliques, we designed a three-step stratified sample. At first, the sample was stratified, following a proportional estimate of households according to the neighborhood where they are located and the number of residents. In a second stage, the households were randomly selected in each neighborhood, according to a previously established systematic agenda. In the third stage, a respondent was drawn, at random, from each household. This provided a unit of information collection, which we call the observation unit. It consists of the individual drawn within the household, as well as several units of analysis that feed the explanatory models, some at the individual level, such as the contact rates aggregated by age groups, and others at the collective level, such as the size of the household, measured in number of inhabitants, and the social circles where relationships are held, inside and outside the home.

Based on a confidence level of 95% and a sampling error of 2%, we determined a first sample of households of size 1,000 to apply the instrument. To correct the demographic census lag and the availability bias at the time of collection, we returned to the field and collected a second sample, following the stratified design of the first survey, with 450 households. With this, we subsequently calibrated the data taking into account two basic variables: sex and age group. To obtain the values of the standard deviation and confidence intervals for the estimators, a simple resampling procedure of 100 copies of the original database was applied, with replacement, plus a column with the calibrated weights for each resampled observation (CHEN; SHEN, 2019).

An epidemiological diary (HOANG *et al.*, 2019), adapted from the instrument used by POLYMOD in the United Kingdom (MOSSONG *et al.*, 2008), was applied. Due to the application time and costs, the interview method was chosen. A self-administered questionnaire along several days in a week, as applied in Europe, would require a great follow-up effort, given the social conditions of the target population, while at the same time potentially compromise the response rate necessary to attain the optimal sample size. The questionnaire was designed in three blocks of questions. The first one identified the number of people living in the house and their sex. With this information the respondent was drawn by lot, then it was decided if the drawn respondent was qualified to give the information, asking if in the last twenty-four hours he/she talked face to face or had any physical contact (i.e. handshake, hugs, kisses, contact while doing sports) with one or more people at the same time. The second one asked about socio-demographic characteristics of respondents (age, sex, race, work condition, income and educational level). The third inquired, in detail, about the social circle where the contacts took place – in house, outside, neighborhood, church, work, school –, the number of people grouped together – which for the purpose of the analysis we call *cliques* – the sex and age of the *alteri*, the duration, the frequency among other characteristics.[3] This research considered qualified respondents as adults aged 18 or over, as well as children and young people who gave their consent under the guidance of a responsible adult, regarding the consent that the epidemiological diary assumes. The study was approved by the Research Ethics Committee of the Federal University of Minas Gerais (UFMG).

Results

Socio-demographic data

It was found that the adult population predominates in Aglomerado da Serra, in the range of 20 to 59 years (61.6%), with a significant presence of children and adolescents from 0 to 14 years (18.7%) and the elderly (12.1%), which comprise the so-called dependent population. The presence of young people between 15 and 19 years old (7.61%) is not very expressive. As expected, for the Brazilian case, the vast majority of the population declared to identify as black and brown (80.8%), this is due to the fact that poverty affects mainly these population segments (OSÓRIO, 2019).

In Aglomerado da Serra, the data indicate the greater presence of households with up to 3 residents (65.6% of households), within the limit of the average size of households projected for Brazil in 2020 – 3.0 residents (GIVISIEZ, 2018). Another 19.8% of households have up to 4 people and only 14.7% have more than 4 people, confirming the trend towards smaller households, as a result of demographic changes that have taken place in recent decades.

Household income in Aglomerado da Serra is low. 76.0% of households have a monthly family income of up to 2 minimum wages, and of these, almost 39.0% of households live with up to 1 minimum wage. However, only 36.9% of households received emergency aid in 2021, confirming the limited scope of social protection measures to reduce impact in times of health crisis.

[3] Sociometrically speaking, the *ego* is the person who indicates a contact and the *alteri* is the partner with whom he/she relates. *Clique* designates a grouping of people, represented as nodes, where all are in relation to each other. This last concept will be further developed in section six of the article.

Social contact rate and their characteristics

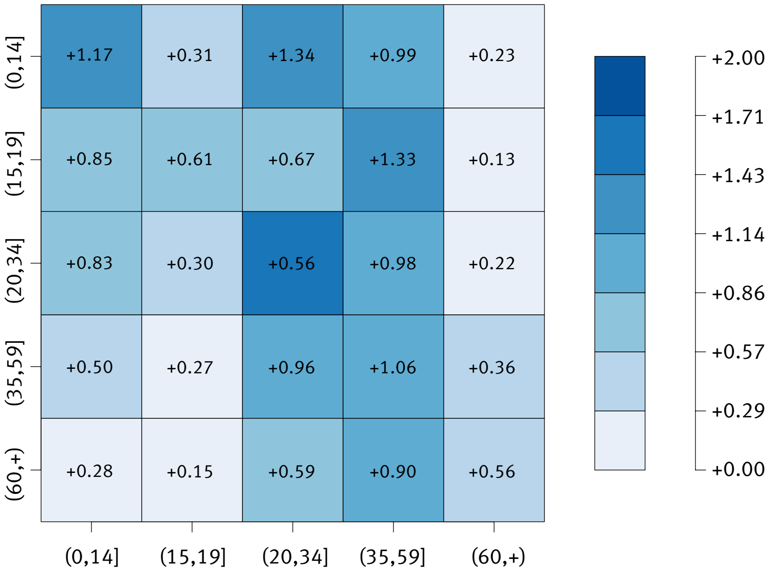
From this sample, it was possible to identify the social contact rate by age group, as well as the distribution of these contacts by meeting place (or social circle) and their duration. It was found that children and adolescents, from 0 to 14 years old, reported a higher average of contacts, compared to other age groups. Young people and adults between 20 and 34 years old were the second group to report more contacts, whose average also exceeds that of the other age groups. The age group of 60 years and over showed the lowest rate.

As for contacts through social circles, 62.3% of the reported contacts took place at home, followed by contacts made in the neighborhood, 19.3%. Contacts in other circles were greatly reduced. Only 0.82% of contacts were made in school environments. The data are consistent with the period in which the research was conducted (end of the 1st semester of 2021), when Belo Horizonte had measures to restrict circulation and ban school attendance.

The data also showed differences in the duration of each contact by social circles. Contacts were reported to last longer (more than 4 hours) at home (70.7%), at work (50.3%) and leisure (46.5%). Shorter contacts (less than 5 min and between 6 and 15 min) occurred mainly in commerce/services (38.3%) and neighborhood (33.8%).

It is important to note that in addition to the average number of reported contacts, location and duration, the dynamics of contagion also depend on the interaction between different age groups, more specifically, on knowing which age groups interact with each other, which corresponds to the rates of transmission, intra- and inter-band contacts. When we see the contact process as a Poisson process, then the contact rate of the process is estimated by the average number of contacts. The contact rates, or average number of contacts, between age groups can be seen in the Figure 1.

FIGURE 1
Matrix of social contact rates by age groups



Source: Research data.
Note: The matrix is read in the direction of the line to the column. For example, if we want to know the observed rate of contacts between people aged 0-14 and those aged 35-59, we look for the respective vertex, which indicates it is 0.99 contacts per person per day. In other words, on average, each child or adolescent reported one contact with an adult in the age group in question. We then looked at the line for the 35 to 59 age group to see how many contacts they indicated with a child or adolescent, and found that it was 0.5 on average. The rates do not match because, within the sample, the *alteri* (indicated) do not necessarily coincide with those indicating (ego). This makes it necessary to symmetrize the matrix, using the arithmetic mean, to include it in the SIR models.

SIR model by ages using contact rates

To simulate pandemic behavior, incorporating the effect of the structure of social contacts, a SIR epidemiological compartment model was applied – (S) susceptible, (I) infected and (R) recovered – by age (KEELING; ROHANI, 2008). Five age groups were used, corresponding to individuals with ages in the ranges, in complete years, from 0 to 14 (group 1), 15 to 19 (group 2), 20 to 34 (group 3), 35 to 59 (group 4) and 60 years or older (group 5). This set of intervals was obtained by grouping adjacent age groups in the contact rate matrix, until it was reduced to five age groups.

The system of differential equations describing the evolution of the number of individuals in each compartment, in the aged SIR model, is:

$$\frac{dS_i}{dt}(t) = -\beta S_i(t) \Sigma_j \frac{I_j(t)}{N_j} t_{i,j} \tag{1}$$

$$\frac{dI_i}{dt}(t) = \beta S_i(t) \Sigma_j \frac{I_j(t)}{N_j} t_{i,j} - \frac{1}{y} I_i(t) \tag{2}$$

$$\frac{dR_i}{dt}(t) = -\frac{1}{y} I_i(t) \tag{3}$$

Here β is the infection rate and $1/y$ is the recovery rate, $S_i(t)$, the number of susceptibles, $I_i(t)$ the number of infected e $R_i(t)$ the number of recovered at time t. If N is the total population being considered then $S_i(t) + l_i(t) + R_i(t) = N$.Since the formulas consider age groups, the subscripts i and j designate the ith and jth age group, respectively. Thus, $t_{i,j}$ means the contact rate between the i age group and the j age group.

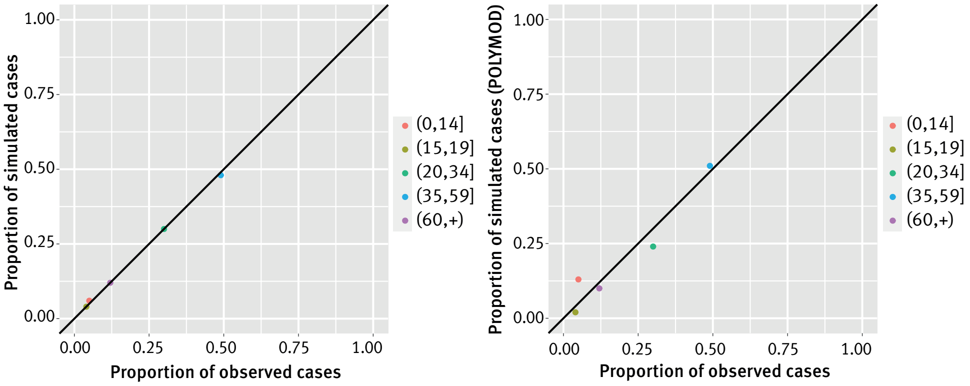
The COVID-19 infection rate at the time of the survey was $\beta_0 = 0{,}05$ (YANG *et al.*, 2020; ZHOU *et al.*, 2020) and, for the calibration of the model, we adjusted this rate taking into account the average of the contacts, which is also the average of rates, of the matrix of contact rates, using the estimated matrix for Brazil in POLYMOD, we obtain $\bar{t}_0 = 9.068888$ and then, to simulate the results in Belo Horizonte, we use the estimated contact matrix for Aglomerado da Serra, obtaining $\beta = \beta_0 / \bar{t}_0 = 0.0055$, this value corrects the effect of contacts in estimating the rate β_0. The recovery rate $1/y = 1/7$ corresponds to the inverse of the average recovery time for COVID-19. Three values were used for the rates, corresponding to the estimated average rates and the upper and lower limits of the confidence intervals (of 95%) for the rates. Rate matrices were symmetrized to reduce bias (HAMILTON *et al.*, 2022).[4] The system of equations implementation was done using the Epimodel package from R (JENNESS *et al.*, 2018).

To assess the consistency of the model proposed here, two simulations were carried out, using epidemiological parameters of the reference week of data collection as reference. One, using the adjusted contact rate data for Brazil, derived from the POLYMOD (PREM *et al.*, 2021;MOSSONG *et al.*, 2008) and the other with the empirical rates of Aglomerado da Serra/BH, both with a projection horizon of one week.[5] The results are shown in Figure 2, where we plot in a graph the simulated and observed proportions of infected by age group.

[4] The non-symmetry present on contact matrices was observed on several studies and evaluated recently on the work by Hamilton (2022). This study compared symmetric versus non-symmetric contact matrices, via simulation of SIR type models using POLYMOD estimates and comparing also with observed data. According to the study, models with non-symmetric matrices "underestimated the basic reproduction number, had delayed timing of peak infection incidence, and underestimated the magnitude of peak infection incidence". Non-symmetric matrices also "influenced cumulative infections observed per age group, and the projected impact of age- specific vaccination strategies".

[5] Prem *et al.* (2014) performs sophisticated demographic projection work to find social contact rates by age group in three social circles (home, school, and work) in 152 countries covering 95.9% of the world's population. They use three data sources: POLYMOD, Demographic and Health Survey and national data from different countries. The projection process starts with a hierarchical Bayesian model that estimates contact rates by age and social circles in each of the eight European countries covered by POLYMOD and for the whole set. This first exercise allows for the construction of three matrices with contact rates by age groups, one for each social circle. Subsequently, the contact rates of each matrix are projected for the countries that were not part of POLYMOD considering the following demographic parameters available in national databases: (a) population profile by age groups, (b) labor force participation, (c) student-teacher ratio, (d) school enrollment rates.

FIGURE 2
Comparative proportion of cases from proposed contact rate and the POLYMOD model



Source: COVID-19: research data.

The estimates produced by the model proposed in this research, from the survey of contact rates by age groups, offer a better approximation between predictions and observations than the approximation that uses estimated rates of the POLYMOD (Figure 2). In both simulations, the average contact rates for Brazil from POLYMOD were used as a correction factor.

Contact rates and their conditions: a clique approach

Next, we present the main determinants of contact rates collected in Aglomerado da Serra. Since this is the most epidemiologically relevant data from the perspective of contact structures, it is pertinent to explore it from the perspective of some socio-demographic factors that were raised at the time of collection.

First, it should be explained that we approach contact numbers using a proxy variable: the cliques or groupings declared by respondents. Each respondent was asked about the contacts he had in the last 24 hours, according to the specific place where they happened (house, neighborhood, business, etc.), but also asked to indicate how many other people they had been in contact with simultaneously, as well as the age and sex of these *alteri*. Given that we conducted a basic SIR model on diseases that are transmitted person-to-person, such as respiratory diseases, it is useful to understand which covariates are associated with these agglomerations, of variable size, where the contagions happen. To this end, we must highlight that we chose to name the variable of interest as “clique density”, due to the sociometric concept that defines a clique as a group where all its members are adjacent to each other, that is, where all are in contact with each other. We assume that a clique is a cluster with $k(k\text{-}1)$ contacts, where k is the number of vertices, which in this case corresponds to number of people in contact. Respondents declared cliques with a minimum size of 2 and a maximum of 11 people.[6] At this point, two clarifications are necessary. First, we use the concept of clique in the mathematical-formal sense that it has in graph theory, i.e., a grouping where all nodes are adjacent to each other. It does not have the substantive sense of a cohesive group by a common identity recognized by the members. Secondly, the contact rate we have discussed is nothing but an average of the relationships considering all the cliques in which a person takes part. For example, if a respondent declared that, in the last twenty-four hours, he or she was grouped into three cliques of size 3, 5 and 7 respectively, then his or her average number of contacts is $\Sigma(k\text{-}1) * 2/ \Sigma k$, which corresponds to 1.6.

The frequency distribution of the clique size variable presents a concentration in the smaller size cliques. Only contacts declared in the first contact situation were used for the purpose of this analysis, as the memory bias delivered a decreasing valid data frequency: 99.4% in the first situation, 30.5% in the second situation and 2.5% in the third situation. Since the number of contacts in the cliques of k size follows a geometric progression, the natural logarithm of this progression was used as a scale for our independent variable: clique density.[7] In this way, we tested, using a log-lin model, the associations between the response variable and its determinants with explanatory power.

We interrogated our data using two "log-lin" models, following the expression used by Gujarati and Porter (2012). The general equation is:

$$\ln Y = \beta_0 + \beta x_1 \ldots + r \tag{4}$$

In both models, the response variable is clique density. We use two treatment criteria, social circle, and age group, to see how a set of determinants impact the fact that an individual clusters in cliques with different numbers of social contacts. These are the main independent variables. Both variables were transformed from categorical into a set of binary ones where each category corresponds to a new variable. In the first model, only the social circle home was included while the last age group (› 60) was used as the omitted reference category. In the second model, home is the omitted reference category, keeping all other variables, including age group, in the same way. We should remember that at the time of our data collection, the Brazilian federal government had already implemented economic aid for low-income families to allow them to subsist during the social isolation measures.

The models assume the point of view of the interviewee.

[6] In sociometric terms, the number of contacts, or relationships, in a clique is a function, on the one hand of the number of members (size k), and, on the other hand, of the orientation of the relationship. The latter means that we can consider, or not, the direction of the relationship. For example, if two people are married, their relationship is not oriented, it does not make sense to say that A is married to B, but that B is not married to A. In the case of contagious relationships, we must consider that A can infect B, but that B does not necessarily infect A, or vice versa. If the non-orientation perspective of the relationship is adopted, the number of contacts will be $k(k\text{-}1)/2$, but if orientation is adopted, as we have done in this article, the number of contacts will be $k(k\text{-}1)$.

[7] Considering $k(k\text{-}1)$ as the number of contacts in the clique of size k, we have a variable with a geometric progression and with a nonlinear distribution. That is, in a clique of size 3, we have 6 contacts, if the size is 4, we have 12 contacts and so on up a clique of size 11 which has 110 contacts.

TABLE 1
Log-lin model – Dependent variable: Natural logarithm of the clique density (situation inside the house)

Variables	Type of variable	Unstandardized coefficient		Standardized coefficients	t	Sig
		B	Std. Error	Beta		
Constant		1,360	0,182		7,468	0,000**
Number of residents in household	Continuous	0,284	0,021	0,430	13,778	0,000**
Men	Binary	0,056	0,056	0,028	0,996	0,319
Home	Binary	-0,610	0,072	-0,246	-8,490	0,000**
Emergency aid	Binary	0,100	0,059	0,048	1,705	0,089
Public transport	Binary	-0,074	0,075	-0,028	-0,981	0,327
0 - 14 age	Binary	0,419	0,110	0,163	3,805	0,000**
15 - 19 age	Binary	0,246	0,132	0,065	1,866	0,062
20 - 34 age	Binary	0,395	0,097	0,178	4,085	0,000**
35 - 59 age	Binary	0,155	0,095	-0,050	-1,634	0,103
› 60 age	Omitted category					

Source: Research data.
**P ‹ 0,01; N=1,000; R^2= 0,254.

TABLE 2
Log-lin model – Dependent variable: Natural logarithm of the clique density (situations outside the house)

Variables	Type of variable	Unstandardized coefficient		Standardized coefficients	t	Sig.
		B	Std. Error	Beta		
Constant		0,914	0,176		5,201	0,000
Number of residents in household	Continuous	0,288	0,021	0,435	14,005	0,000**
Men	Binary	0,036	0,056	0,018	0,644	0,519
Commerce	Binary	0,059	0,194	0,009	0,307	0,759
Church	Binary	1,866	0,490	0,105	3,806	0,000**
Leisure	Binary	0,871	0,140	0,175	6,234	0,000**
Workplace	Binary	0,655	0,173	0,106	3,785	0,000**
Neighborhood	Binary	0,579	0,093	0,179	6,242	0,000**
School	Binary	0,583	0,383	0,042	1,522	0,128
Home	Omitted category					
Emergency aid	Binary	0,082	0,059	0,039	1,398	0,162
Public transport	Binary	-0,063	0,075	-0,023	-0,834	0,405
0 - 14 age	Binary	0,368	0,111	0,143	3,332	0,001**
15 - 19 age	Binary	0,218	0,131	0,057	1,664	0,097
20 - 34 age	Binary	0,345	0,097	0,155	3,563	0,000**
35 – 59 age	Binary	0,125	0,095	-0,041	-1,321	0,187
›60 age	Omitted category					

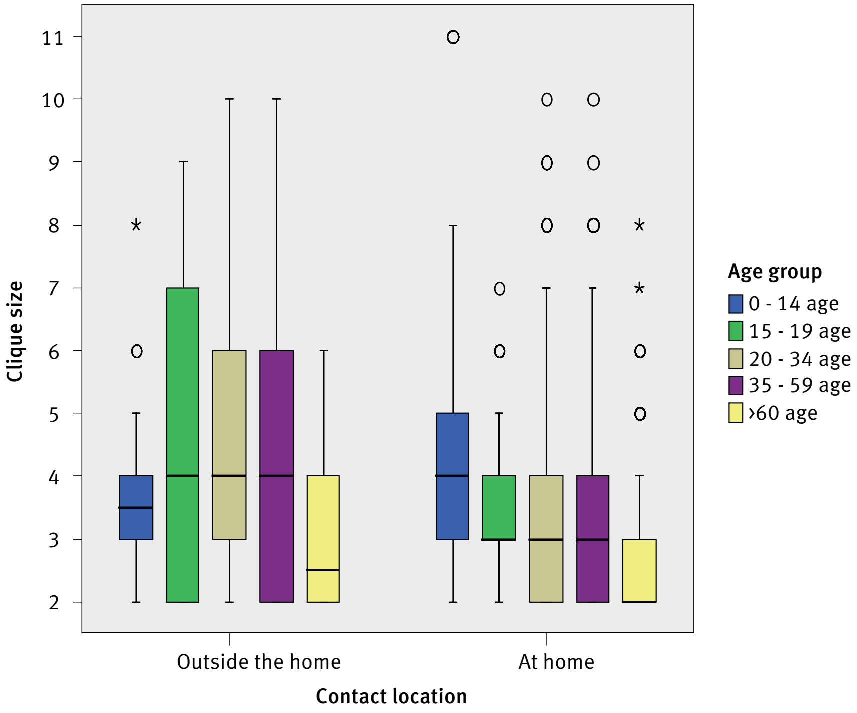
Source: Research data.
**P ‹ 0,01; N=1,000; R2= 0,268.

Since the scale of the dependent variable is the natural logarithm, the coefficients are percentage proportions of how much each explanatory variable increases the response variable (clique density). In the first model, where the contact situation is the house (Table 1), one more resident in the household increases the clique size by 28.4%. Gender has no significant effect. The contact situation within the household reduces the clique density by 60%. In the opposite case, when we invert the binarization, recording the situation outside the home, there is a 60% increase in clique density. Neither the emergency aid nor the fact of using public transport had significant effects. Children aged 0 to 14, and young adults aged 20 to 34 had 41%, and 39%, respectively, more contacts in the clique group than people aged 60 and over (≥ 60). The age groups from 15 to 19 and 35 to 59 did not present a significant coefficient at the conventional value of 5%.

On the second model, there are several contact situations that correspond to forms of socialization that take place outside the home. The number of residents maintains its aggregation effect on clique density. Gender, emergency aid and public transport, like in the first model, do not present a statistically significant effect. Among the contact situations, two were not significant, namely, commerce and school. It is important to remember that on the date of data collection, the Belo Horizonte school system had suspended activities due to the pandemic. It is possible that the contacts declared at school correspond to occasional visits to pick up some study material to work on at home. The other contact locations show important impacts on the size of the clicks compared to the home. Groupings at church generate 186% more contacts than those at home, followed by leisure (87%), workplace (65.5%) and neighborhood (58%). The only age groups with predictive power over clique size correspond to the same age groups as in the first model, children, and young adults. Taken together, the previous results demonstrate that the size of households, in terms of number of residents, is an important determinant in the formation of epidemiologically relevant clusters. However, the formation of cliques within the household does not mean that they took place in the respondent's household. In fact, cross-table analyses show only a 65% correspondence between the number of residents and the size of the declared cliques, a percentage obtained by dividing the total diagonal values – between values 2 and 10 for both the number of households and for the size of cliques – by the total of reported cliques. In the case of cliques located outside the home, the correspondence is only 10%.

The formation of intra-household cliques appears as a phenomenon formed by children and caregivers-parents between 20 and 34 years old, something to be expected when the school system was closed. Figure 3 explains the distribution of clique size according to age groups and contact location. In the latter, greater variability in the clique size can be observed in the situation outside the home, as evidenced by the interquartile range and the amplitude of the clique size. This is consistent with the results of the second model, which showed how church, leisure, work and neighborhood are spaces that encourage people to meet. This is something we can expect if: (1) people are outside the more controlled environment of the home; (2) physical spaces contain vital socialization circles for people and (3) we assume a margin of randomness in social encounters.

FIGURE 3
Clique size grouped by age group and contact location



Source: Research data.

Discussion

Following the current pandemic, triggered by SARS COV 2, a wide range of work has undertaken the challenge of monitoring the expanding course of the pandemic. Some global initiatives turned to technological devices to identify the mobility of human populations almost in real time. The Google platform provided data on the mobility of its users by making use of smartphones' geolocation. To this end, the Google COVID-19 Community Mobility Report (GCCMR) was made available with data from 131 countries. This initiative, with specific research purposes, ended in 2022-10-15. Oliveira *et al.* (2021) used the GCCMR data in ten Latin American countries to associate mobility indexes with the COVID-19 stringency index from Oxford COVID-19 Government Response Tracker (OxCGRT). Without undervaluing the advantages of the previous analysis strategy, it is important to point out several limitations imposed by the baseline data when working with conventional SIR models. As the authors themselves rightly acknowledge (OLIVEIRA *et al.*, 2021), the mobility data provided by GCCMR, in various social circles (parks, work, commerce, among others), constitute a digital proxy for face-to-face human interactions. The attribution of an individual to a place depends on whether the user has activated their phone's location history. Furthermore, Google reserves the right to provide data, about social circles, where there is low frequency of visits, as this may compromise the anonymity of the information. However, the two most serious limitations, marking a substantial difference with our strategy, are the absence of relational information between human beings and the no disaggregation of age groups. We know nothing about physical contacts, nor about one of the most important behavioral determinants, age. In summary, having to refine parameters in mathematical SIR models imposes the rigorous collection of primary data through surveys that provide social contact rates.

When comparing our work with other studies inspired by the POLYMOD strategy and carried out in developing countries, we found some important convergences and differences. Johnstone-Robertson *et al.* (2011) conducted a survey of social contacts in a Township of just under 20 thousand inhabitants, near Cape Town, a rural population with a well-defined census. This allowed for a random sampling of individuals considering age groups. The results accurately demonstrated that the young population between 5 and 19 years of age was at the highest risk of infection by respiratory diseases endemic to that community (tuberculosis and influenza), thus, confirming that, by disaggregating the data by age groups, an epidemiologically relevant determinant of social behavior is identified.

In turn, the work of Grijalba *et al.* (2015) highlights the difficulties of collecting data on social contacts in several population universes at the same time. Wanting to cover 54 rural communities in Peru, they had to give up probabilistic sampling to settle for convenience samples in which the members of at least two households per community were interviewed. Costs and logistics make probability sampling plans less feasible.

Furthermore, when estimating infection rates of pathogens, the incorporation of social context is extremely important to determine the evolution of the epidemic, especially in the case of airborne viruses such as SARS (YANG *et al.*, 2020; ZHOU *et al.*, 2020; LIU *et al.*, 2020). As already pointed out by other research (ZHOU *et al.*, 2020; PREM *et al.*, 2017; BARMPARIS; TSIRONIS, 2020), factors such as household size and age are intrinsically linked to SARS-like virus infection rates, and these factors are linked to socioeconomic conditions that need to be evaluated *in situ* to have a more realistic determination of what infection rates mean.

In turn, the exploration of the socio-demographic conditions of contact rates in this study, through two log-lin models, showed that research of this type is also useful for the determination of covariates associated with the formation of small agglomerations that result in epidemiologically relevant contacts. In this case, we saw how the demographic size of the household is a fundamental covariate when planning mitigation or epidemic control scenarios, as it increases the density of groups between people. Understanding social circles is important to understanding how forms of socialization increase the risk of contagion. In popular communities, places of worship, neighborhoods, and places of leisure, among others, are favorable scenarios for forms of socialization that substantially increase contact rates. The low coverage of emergency assistance provided by the Brazilian federal government was not associated with the reduction of social contacts. When we interpret this finding together with the statistically significant effect of the work circle, we can infer that government aid, in the case of the SARS COV 2 pandemic in urban areas, was useful for the survival of families and less effective in pandemic control. Families from popular sectors that survive on up to two minimum wages have no other option than to seek their livelihood in a job market with a high rate of informality. Beyond methodological divergences, this finding has been reinforced by research that digitally captured social mobility indexes (OLIVEIRA *et al.*, 2021).

This study also indicated, as already demonstrated in specialized literature, that compartmental epidemiological models combined with social contact rates have greater ability to describe epidemiological dynamics because they incorporate interaction between ages (GJOKA *et al.*, 2014; CHIN *et al.*, 2021; PREM *et al.*, 2017). In this regard, we observed the social contact rates collected in Aglomerado da Serra provided a better fit in the SIR model relative to the demographic projection made by Prem (PREM *et al.*, 2017) for Brazil.

Therefore, this study reveals the importance of investing in epidemiological diary research that provides information on the covariates associated with the formation of epidemiologically relevant clusters, and informs compartmental models better, improving their fit and allowing projecting the effect of mitigation processes, such as vaccines or isolation (KEELING; ROHANI, 2008; RAM; SCHAPOSNIK, 2021; COLOMBO; GARAVELLO, 2020), in different age groups, which increases the relevance of their use.

Conclusions

The crisis triggered by SARS COV 2 was a significant opportunity to adapt the technique of epidemiological dairy in the context of health surveillance in Brazil. This study demonstrates how the empirical, *in situ*, estimation of social contact rates improves the descriptive power of compartmental models widely used in epidemiology. In general, these models work at average levels of contact rates, disregarding the heterogeneity of contacts between social groups. In this work, we estimated the rate of social contact by age and the results are more sensitive to the reality of the pandemic.

The technique of epidemiological diaries, adapted as an interview, makes it possible to gather information on rates of social contacts as well as on factors of the socio-demographic structure that affect the rates of social contacts. With greater clarity, on one hand, about the morphological factors of social life, such as the demographic size of households and age composition of the social universe, and, on the other hand, about socialization circles, we can broaden our comprehension of the infectious processes in terms of the different structures of interaction between human beings.

References

BARMPARIS, G. D.; TSIRONIS, G. P. Estimating the infection horizon of COVID-19 in eight countries with a data-driven approach. **Chaos, Solitons & Fractals**, v. 135, p. 109842, 2020.

BURT, R. S. Cohesion versus structural equivalence as a basis for network subgroups. **Sociological Methods and Research**, v. 7, n. 2, p. 189-212, 1978.

CHEN, H.; SHEN, Q. R. Variance estimation for survey-weighted data using bootstrap resampling methods: 2013 methods-of-payment survey questionnaire. **Advances in Econometrics**, v. 39, p. 87-106, 2019.

CHIN, T. *et al.* Contact surveys reveal heterogeneities in age-group contributions to SARS-CoV-2 dynamics in the United States. **medRxiv**, 2021.

COLOMBO, R. M.; GARAVELLO, M. Optimizing vaccination strategies in an age structured SIR model. **Mathematical Biosciences and Engineering**, v. 17, n. 2, p. 1074-1089, 2020.

GIVISIEZ, G. H. N.; OLIVEIRA, E. L. de. **Demanda futura por moradias**: demografia, habitação e mercado. Niterói/RJ: Universidade Federal Fluminense/PROPPi, 2018.

GJOKA, M.; SMITH, E.; BUTTS, C. Estimating clique composition and size distributions from sampled network data. *In*: IEEE CONFERENCE ON COMPUTER COMMUNICATIONS WORKSHOPS (INFOCOM WKSHPS). **Proceedings** [...]. Toronto: IEEE, 2014.

GRIJALVA, C. G. *et al.* A household-based study of contact networks relevant for the spread of infectious diseases in the highlands of Peru. **PLOS ONE**, v. 10, n. 3, e0118457, 2015.

GUJARATI, D. N.; PORTER, D. C. **Econometria básica**. 5 ed. Porto Alegre: AMGH, 2012.

HAMILTON, M. A.; KNIGHT, J.; MISHRA, S. Failure to balance social contact matrices can bias models of infectious disease transmission. **medRxiv**, 2022.

HOANG, T. *et al.* A systematic review of social contact surveys to inform transmission models of close-contact infections. **Epidemiology**, v. 30, n. 5, p. 723-736, 2019.

HUNTER, D. R.; HANDCOCK, M. S.; BUTTS, C. T.; GOODREAU, S. M.; MORRIS, M. ergm: a package to fit, simulate and diagnose exponential-family models for networks. **Journal of Statistical Software**, v. 24, n. 3, p. 1-29, 2008. DOI: 10.18637/jss.v024.i03.

JENNESS, S. M.; GOODREAU, S. M.; MORRIS, M. EpiModel: an R package for mathematical modeling of infectious disease over networks. **Journal of Statistical Software**, v. 84, n. 8, 2018.

JOHNSTONE-ROBERTSON, S. P.; MARK, D.; MORROW, C.; MIDDELKOOP, K.; CHISWELL, M.; AQUINO, L. D.; BEKKER, L. G.; WOOD, R. Social mixing patterns within a South African township community: implications for respiratory disease transmission and control. **American Journal of Epidemiology**, v. 174, n. 11, p. 1246-1255, 2011. DOI: 10.1093/aje/kwr251.

KEELING, M.; ROHANI, P.; POURBOHLOUL, B. Modeling infectious diseases in humans and animals: modeling infectious diseases in humans and animals. **Clinical Infectious Diseases**, v. 47, n. 6, p. 864-865, 2008.

KLEPAC, P. *et al.* Contacts in context: large-scale setting-specific social mixing matrices from the BBC Pandemic project. **medRxiv**, 2020.

LIU, Y. *et al.* What are the underlying transmission patterns of COVID-19 outbreak? An age-specific social contact characterization. **eClinicalMedicine**, v. 22, 100354, 2020.

MOSSONG, J. *et al.* Social contacts and mixing patterns relevant to the spread of infectious diseases. **Medic PLoS**, v. 5, n. 3, e74, 2008.

OLIVEIRA, G. L. A.; LIMA, L.; SILVA, I.; RIBEIRO-DANTAS, M. C.; MONTEIRO, K. H.; ENDO, P. T. Evaluating social distancing measures and their association with the Covid-19 pandemic in South America**. International Journal of Geo-Information**, v. 10, n. 3, 121, 2021. https:// doi.org/10.3390/ijgi10030121.

OSORIO, R. G. **A desigualdade racial da pobreza no Brasil**. Rio de Janeiro: Ipea, 2019.

PREM, K.; COOK, A. R.; JIT, M. Projecting social contact matrices in 152 countries using contact surveys and demographic data. **PLOS Computational Biology**, v. 13, n. 9, e1005697, 2017.

RAM, V.; SCHAPOSNIK, L. P. A modified age-structured SIR model for COVID-19 type viruses. **Scientific Reports**, v. 11, n. 1, p. 1-15, 2021.

SILVEIRA, D. C. **A implantação do Programa Vila Viva em áreas de Belo Horizonte**: uma análise documental. 2015. 93 f. Dissertação (Mestrado em Saúde Coletiva com concentração em Ciências Humanas e Sociais em Saúde) – Programa de Pós-graduação em Saúde Coletiva, Centro de Pesquisas René Rachou, Fundação Oswaldo Cruz, Belo Horizonte, 2015.

YANG, Y. *et al*. The deadly coronaviruses: the 2003 SARS pandemic and the 2020 novel coronavirus epidemic in China. **Journal of Autoimmunity**, v. 109, 102434, 2020.

YAVEROGLU, O. N. *et al*. Ergm.graphlets: a package for ERG modeling based on graphlet statistics. **arXiv, 405.7348 [cs]**, 2014.

ZHOU, P. *et al*. A pneumonia outbreak associated with a new coronavirus of probable bat origin. **Nature**, v. 579, n. 7798, p. 270-273, 2020.

About the authors

Sílvio Segundo Salej Higgins holds a BA in Philosophy from Pontificia Universidad Javeriana and an MA in Political Sociology from Universidade Federal de Santa Catarina. PhD in Sociology from the University of Paris Dauphine (France) and in Political Sociology from the Federal University of Santa Catarina (Brazil) in the framework of the French-Brazilian Doctoral College – CAPES, Ministry of Education of Brazil and Ministère de l'Éducation National (France). He leads the Interdisciplinary Research Group on Social Network Analysis (GIARS) – UFMG – CNPq Certificate. Associate Professor of Sociology Department, Federal University of Minas Gerais (UFMG) – PQ 2 Productivity Fellow.

Adrian Pablo Hinojosa Luna holds a BA in Mathematics from the Universidad Central del Ecuador, an M.Sc. in Mathematics from the Asociación Instituto Nacional de Matemática Pura e Aplicada, and a Ph.D. in Mathematics from the Asociación Instituto Nacional de Matemática Pura e Aplicada. Adjunct professor IV at the Federal University of Minas Gerais.

Reinaldo Onofre dos Santos is graduated in Geography (IGC-UFMG), Master and PhD in Demography from Federal University of Minas Gerais (CEDEPLAR-UFMG). Adjunct professor at the Department of Geosciences of the Federal University of Juiz de Fora.

Andreia Maria Pinto Rabelo is PhD in Sociology from University of Minas Gerais (UFMG). Graduated in Social Sciences from Fundação Educacional de Divinópolis/UEMG and has a master's degree in Education, Culture and Social Organizations from the same institution. Participates in the Interdisciplinary Research Group in Social Network Analysis (GIARS-UFMG) and the Research Group Observatory of Innovations, Networks and Organizations (OIRO-UFOP).

Maíra Soalheiro has an undergraduate degree in Statistics from the Federal University of Minas Gerais, a master's degree in Statistics from the Federal University of Minas Gerais). She is a doctoral student in Statistics at the same institution.

Vanessa Cardoso Ferreira is PhD in Demography from the Center for Regional Development and Planning (CEDEPLAR) at the Federal University of Minas Gerais (UFMG). She holds a Master's degree in Demography and a Bachelor's degree in Economic Sciences from UFMG.

Contact address

Sílvio Segundo Salej Higgins
Universidade Federal de Minas Gerais, Faculdade de Filosofia e Ciências Humanas
Av. Presidente Antônio Carlos, 6.627, Pampulha
31270901 – Belo Horizonte-MG, Brasil
Caixa postal: 253

Adrian Pablo Hinojosa Luna
Universidade Federal de Minas Gerais, Instituto de Ciências Exatas
Av. Presidente Antônio Carlos, 6.627, Pampulha
31270901 – Belo Horizonte-MG, Brasil

Reinaldo Onofre dos Santos
Universidade Federal de Juiz de Fora, Instituto de Ciências Humanas
Rua José Lourenço Kelmer, s/n, São Pedro
36036900 – Juiz de Fora-MG, Brasil

Andreia Maria Pinto Rabelo
Rua Boston, 391, Davanuze
35500548 – Divinópolis-MG, Brasil

Maíra Soalheiro
Universidade Federal de Minas Gerais, Instituto de Ciências Exatas
Av. Presidente Antônio Carlos, 6.627, Pampulha
31270901 – Belo Horizonte-MG, Brasil

Vanessa Cardoso Ferreira
Rua Cecília Fonseca Coutinho, 458, ap. 302, Castelo
30840500 – Belo Horizonte-MG, Brasil

Resumo

Um estudo sobre as taxas de contatos sociais relevantes para a difusão de doenças infecciosas em um aglomerado brasileiro

Inspirado no estudo POLYMOD, foi realizado, em junho de 2021, um *survey* epidemiológico num dos setores de maior densidade populacional e vulnerabilidade social de Belo Horizonte (Brasil). Uma amostra de 1.000 domicílios permitiu identificar, num período de 24 horas, as taxas de contatos sociais por faixas etárias, o tamanho e a frequência de cliques do qual participou o respondente, assim como outros fatores sociodemográficos associados (número de moradores do domicílio, local do contato, uso do transporte público, entre outros). Os dados foram analisados em duas fases. Na primeira, foram comparados os resultados entre dois modelos SIR que simularam um processo pandêmico de oito dias. Um incluiu parâmetros ajustados a partir das taxas de contatos observadas. O outro operou com parâmetros ajustados a partir de taxas projetadas para o Brasil. Na segunda fase, mediante uma regressão *log-lin*, modelamos os principais determinantes sociais das taxas de contato, utilizando o adensamento de cliques como uma variável *proxy*. A análise dos dados mostrou que o tamanho da família, a idade e os círculos sociais são as principais covariáveis que influenciam a formação dos cliques. Também demonstrou que modelos epidemiológicos compartimentais, combinados com taxas de contato social, têm melhor capacidade de descrever a dinâmica epidemiológica, fornecendo uma melhor base para medidas de mitigação e controle de doenças que causam síndromes respiratórias agudas.

Palavras-chave: Survey epidemiológico. POLYMOD. Taxa de contato social. Cliques.

Resumen

Un estudio sobre las tasas de contactos sociales relevantes para la propagación de enfermedades infecciosas en un barrio popular del Brasil

Con inspiración en el estudio POLYMOD, se hizo una encuesta epidemiológica, en junio de 2021, en uno de los sectores más densamente poblados y socialmente vulnerables de Belo Horizonte (Brasil). Una muestra de mil hogares permitió identificar, en un período de 24 horas, el tamaño y la frecuencia de los cliques en los que participó el encuestado, las tasas de contactos sociales por grupos de edad, así como otros factores sociodemográficos asociados (número de residentes en el hogar, lugar de contacto, uso del transporte público, entre otros). Los datos se analizaron en dos fases. En la primera, se compararon los resultados entre dos modelos SIR que simularon un proceso pandémico de ocho días. Uno incluyó parámetros ajustados a partir de tasas de contacto observadas; el otro operó con parámetros ajustados a partir de tasas proyectadas para Brasil. En la segunda, mediante una regresión *log-lin*, se modelaron los principales determinantes sociales de las tasas de contacto, utilizando la densificación de cliques como una variable proxy. El análisis de los datos mostró que el tamaño de la familia, la edad y los círculos sociales son las principales covariables que influyen en la formación de camarillas. También demostró que los modelos epidemiológicos compartimentados, combinados con tasas de contacto social, son más capaces de describir la dinámica epidemiológica, proporcionando una mejor base para las medidas de mitigación y control de las enfermedades causantes de síndromes respiratorios agudos.

Palabras clave: Encuesta epidemiológica. POLYMOD. Tasa de contacto social. Cliques.

Received for publication in 02/12/2022
Approved for publication in 19/04/2023