Relationship between household attributes and contact patterns in urban and rural South Africa

Kausutua Tjikundi[1,2], Jackie Kleynhans[3,4], Stefano Tempia[3,4,5], Cheryl Cohen[3,4], Daniela Paolotti[1], Ciro Cattuto[1], Lorenzo Dall'Amico[1]*

1 ISI Foundation, Turin, Italy, 2 University of Turin, Department of Computer Science, Turin, Italy, 3 Centre for Respiratory Diseases and Meningitis, National Institute for Communicable Diseases of the National Health Laboratory Service, Johannesburg, South Africa, 4 School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa, 5 The Partnership for International Vaccine Initiative, Taskforce for Global Health, Atlanta, Georgia, United States of America

* lorenzo.dallamico@isi.it

Abstract

Households play a crucial role in the propagation of infectious diseases due to the frequent and prolonged interactions that typically occur between their members. Recent studies have emphasized the need to include socioeconomic variables in epidemic models to account for the heterogeneity induced by human behavior. While sub-Saharan Africa suffers the highest burden of infectious disease diffusion, few studies have investigated the mixing patterns in the countries and their relation with social indicators. This work analyzes household contact matrices measured with wearable proximity sensors in a rural and an urban village in South Africa. Leveraging a rich data collection describing additional individual and household attributes, we investigate how the household contact matrix varies according to the household type (whether it is composed only of a familiar nucleus or by a larger group), the gender of its head (the primary decision-maker), the rural or urban context, and the season in which it was measured. We show the household type and the gender of its head induce differences in the interaction patterns between household members, particularly regarding child caregiving, suggesting they are relevant attributes to include in epidemic modeling.

1 Introduction

Close-range proximity interactions are the main driver of infectious disease transmission [1,2]. Quantifying these interactions and understanding their determinants is thus necessary to design optimal epidemic mitigation strategies. Contact matrices describe detailed data on social mixing patterns between groups of individuals and inform epidemic models by accounting for diverse social contexts and age-specific infection risks [3]. Age is commonly adopted to define such groups [4–6] even if recent works called for a more holistic view, integrating socio-economic variables

into epidemic models [7–14]. Contact matrices are commonly defined within specific contexts where the interactions occurred [15,16] – *e.g.,* household, school, workplace – to account for the diverse importance each context has on the outcome of transmission. Households, in particular, play an important role in spreading the disease, as interactions within these settings are frequent and prolonged [17,18] and can also act as a bridge between different contexts, such as schools and workplaces.

Although most studies and data collections focus on high-income countries [19], sub-Saharan Africa experiences the highest burden of infectious disease epidemics [20]. The past decade has seen an increase in interest in studying social mixing across sub-Saharan African countries to evaluate the impact of infectious disease transmission, with diverse approaches ranging from contact diaries [21–25] to novel technology-driven approaches based on proximity sensors [26–29]. Findings show a high degree of heterogeneity in the observed contact patterns, with few other recurring features across studies. For instance, primary school children generally display higher contact rates compared to adults [21–23,30]. In Kenya [21,31] and Zambia [32], rural areas record higher contact rates compared to semi-urban regions. On the opposite, in Zimbabwe [22] peri-urban regions display higher contact rates compared to rural ones. In South Africa [25] and Mozambique [33], people living in rural areas experience almost double the contact rates as the ones in urban areas. In these settings, most interactions occur within households [26,27,34], and intergenerational contacts are driven by multigenerational households which are more common in rural areas, as reported by studies in Kenya [31], Zimbabwe [22], and Malawi [27]. Community-based social mixing, such as those observed in township settings in South Africa [35], and Zambia [32] are associated with high rates of tuberculosis transmission in the community.

The interaction dynamics within a household are not solely based on age, but also on the roles of its members and the overall organization of the household. In Africa, variations in this organization have been associated with poverty, urbanization, and the loosening of societal norms [36,37]. The household type (whether it is composed only of a familiar nucleus or by a larger group) and its composition (defined by age, gender, marital status, and relationships among the household members) may hence be needed to describe interactions accurately [38,39]. Children are regarded as primary introducers of infections like influenza due to high external exposure in schools, while mothers often drive within-household transmission due to frequent caregiving interactions [40,41]. Multigenerational households have high transmission risks for diseases like influenza and RSV, particularly affecting infants and the elderly [40,42]. In high-birth-rate settings like Kenya, larger households increase measles transmission [43], while communities composed of extended households may also be more vulnerable to larger, more severe outbreaks of infectious diseases [44]. Similarly, in low-income settings, household composition is more intergenerational and leads to increased intra-household transmission due to more frequent interactions between older adults and younger individuals compared to high-income settings [45]. Studies confirm that household composition, rather than size alone, influences infection risk [46–48], yet disease models often assume random household mixing despite

evidence of structured contact patterns [17]. Most of these studies focused on understanding how household structure and composition influence the transmission dynamics of infectious diseases, but little effort on how structure and composition shape contact patterns relevant to the transmission of infectious diseases.

In this work, we go beyond age-only contact pattern measurements and focus on additional social attributes that may be relevant for disease spread. We consider household contact matrices measured with high resolution proximity sensors during the PHIRST study, conducted in South Africa in 2018 [25,28,49]. These data include rich information related to the social role of individuals in the household. We investigate how household composition, headship, seasonality, and location affect the contact patterns, identifying socio-demographic attributes that may be relevant for epidemic modeling.
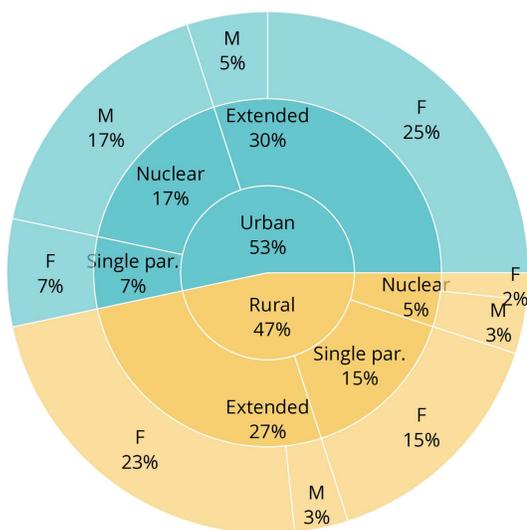
2 Results

The data collection took place in a rural site – Agincourt, Bushbuckridge Municipality (Mpumalanga Province) – and an urban area – Klerksdorp, Matlosana Municipality (North West Province) over three measurement waves with a duration between 10 and 14 days. The first wave was in February (South African summer), the second in April (South African autumn), and the last one in June (South African winter). In-household interactions were measured using the proximity sensors developed by the SocioPatterns collaboration (sociopatterns.org, [50]). Sect 4.1 describes the ethical aspects of the study, while Sect 4.2 describes the data collection and cleaning procedures.

We follow the study protocol that identifies households as groups of three or more people who regularly share at least two meals in the same residence at least two days per week, excluding residential institutions. In Africa, households are diverse and shaped by various factors such as cultural diversity, economic conditions, and historical influences. South Africa is no exception, with its diverse population and the lasting influence of apartheid [37]. To quantify the relation between the household structure and the contact patterns, we group households according to three features: i) the site, which can be urban or rural; ii) the gender of the household head, which can be male or female; iii) the household type, which can be nuclear, extended, or single-parent. The left plot of Fig 1 summarizes the composition of our dataset according to the classification above. Sect 4.3 provides additional details on the household classification.

We analyze how household attributes impact the interaction patterns of in-household interaction rates by focusing on contact matrices describing the interaction rates between groups of individuals categorized according to their age and gender. All participants disclosed a binary gender attribute. Age categories are defined as *children* for the age range $0-10$ years, *adolescents* for the age range $11-18$, and *adults* for individuals older than 18 years old. We classified individuals aged 18 as adolescents rather than adults, in line with the convention that individuals typically complete their final year of schooling in the year they turn 18. The table on the right frame of Fig 1 shows the population sizes disaggregated by age and gender for the different types of household classification. Fig 2 shows the household contact matrices for each aggregation level, reporting the daily average contact duration per individual. The matrices are obtained using bootstrap sampling, and the figure reports the average value and the 10% confidence interval. All matrices display a disassortative structure (assortativity index Q [51] ranging between -0.19 and -0.16), dominated by inter-generational interactions commonly observed within households [6,26,52]. Sect 4.4 provides further details on the definition of these contact matrices.

2.1 Comparison of the interaction rates

We compare the interaction rates per individual for different household attributes. We consider the interaction rates per household, by age and gender, and with children. The analysis is performed on a bootstrapped sample of the empirical contact matrices, which is detailed in Sect 4.4. Size effects are quantified in terms of the Cohen's d score, and larger scores imply higher separations between the distributions. To handle the dependencies between the household attributes (see Fig 1), we run a ridge regression to predict the interaction rates from the household attributes. This analysis allows us to identify statistically significant effects in predicting the interaction rates. The details of the ridge regression

Fig 1. Summary of household and population compositions disaggregated by age and gender and according to the household classification. *Left: fraction of households disaggregated by site, household type, and gender of the household head.* The inner level indicates the fraction of households in the rural and the urban sites. The second layer categorizes households by type, while the third layer categorizes households by the gender of the household head. *Right: number of individuals per age-gender for varying household grouping strategies.* The column "Classification" indicates the different ways of characterizing households considered in the main text. The rows that are not separated by a horizontal line indicate non-overlapping groups of households (*e.g.*, Rural and Urban), while overlaps exist across lines separated by horizontal lines (*e.g.*, Rural and Nuclear). We report, in parentheses, the number of households in each group. The subsequent columns indicate the number of individuals in each age group: "M" stands for male, "F" for female, "Ch" for children, "Ado" for adolescents, and "Adu" for adults. The column "Total" reports the total population in each household type (the sum over the row), while the row "Total" is the population per age-gender group. The rightmost bottom cell reports the total population.

implementation are provided in Sect 4.5. The statistical analysis is performed in Python, using the numpy, pandas, and scipy packages.

2.1.1 Household interaction rate. Fig 3 compares the distributions of the total interaction rate per household, defined as the sum of the upper-triangular elements of the matrices shown in Fig 2.

The first wave (summer) recorded a much shorter average interaction time – 3.7 hours per individual on average – with respect to the second (autumn) and the third (winter) – 9.1 and 10.2 hours per individual on average, respectively. The size effect is much larger between the first and second waves ($d = 13$) than it is between the second and third ($d = 1.4$). Higher average contact rates ($d = 3.1$) are observed in the urban setting – 8.9 hours per individual on average – compared to the rural setting – 6.8 hours per individual on average. Extended households have the highest average contact rates – 9.8 hours per individual on average –, followed by nuclear households – 4.2 hours per individual on average – (size effect between extended and nuclear households $d = 14.4$.), and single-parent households – 3.2 hours per individual on average – (size effect between nuclear and single-parent $d = 8.0$). Longer average contact rates ($d = 9.3$) are measured in male-headed households – 12.3 hours per individual on average – with respect to female-headed households – 5.8 hours per individual on average.

The bottom frame in Fig 3 shows the results of the ridge regression analysis, which accounts for the dependencies between variables. This analysis confirms the statistical significance of the trends described above, except for the urban/rural classification, which is not significant. The observed difference in the second frame of the first row in Fig 3 can be explained by the high correlation between household headship (which has a significant impact) and the site location.

2.1.2 Interaction rates by age and gender. We repeat the analysis of Sect 2.1.1, disaggregating the interaction times by age and gender. Fig 4 shows that, in all settings, male children experience the highest interaction rates. The smallest
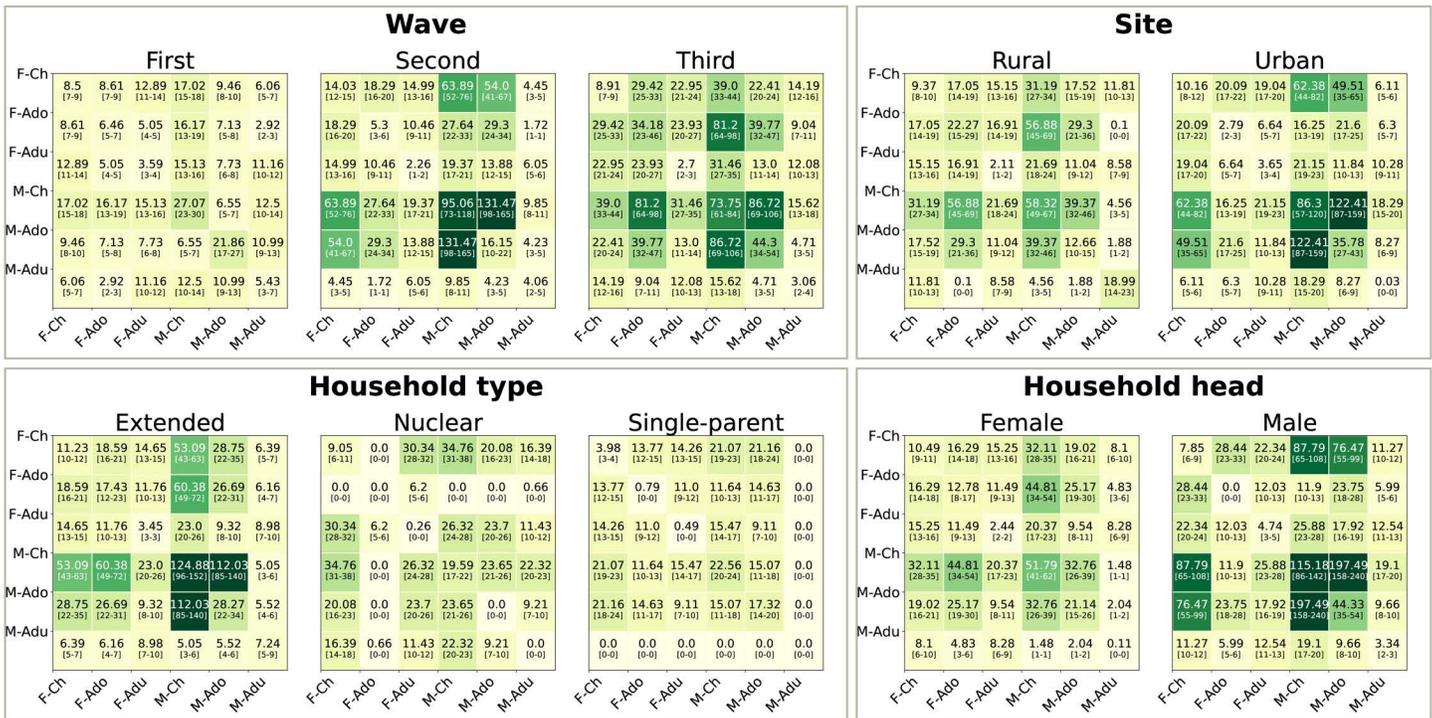
differences are recorded with male adolescents in the urban site ($d = 1.9$), in single-parent households ($d = 1.7$), and in male-headed households ($d = 1.9$). The involvement of female adolescents varies substantially in different contexts. It increases from summer to winter ($d = 7.6$ between first and second wave; $d = 6.9$ between second and third waves); it is larger in the rural context ($d = 5.5$), in extended ($d = 8.7$ with respect to single-parent households), and in female-headed households ($d = 3.4$). Adults, and in particular male adults, experience the lowest interaction rates in all contexts.

The ridge regression shows that men of all ages interact more often in male-headed households. Moreover, in extended households, female adolescents and male children experience higher interaction rates, while male adults have fewer contacts. Only female adolescents have more interactions in nuclear households.

2.1.3 Interactions with children. Fig 5 shows that, in all contexts, female adults have higher interaction with children compared to male adults. This observation provides evidence that the burden of child care is mostly carried by women. The separation is smaller ($d = 10.2$) in nuclear households compared to extended households ($d = 14.6$), suggesting a more equitable share of child care between mothers and fathers in nuclear households. This is also confirmed by the ridge regression that identifies the "Nuclear" attribute as a positive predictor of the interaction rates between male adults and children. In our dataset, all single-parent households are female-headed and, therefore, cannot be compared. Stark differences are observed in the interaction patterns of adolescents and children. As shown in Fig 4, the involvement of adolescents drastically increases from winter to summer, and so does their interaction rates with children. Male adolescents generally have higher interaction rates with children, with few exceptions in which female adolescents record higher interaction rates: the rural households ($d = 2.4$), and female-headed households
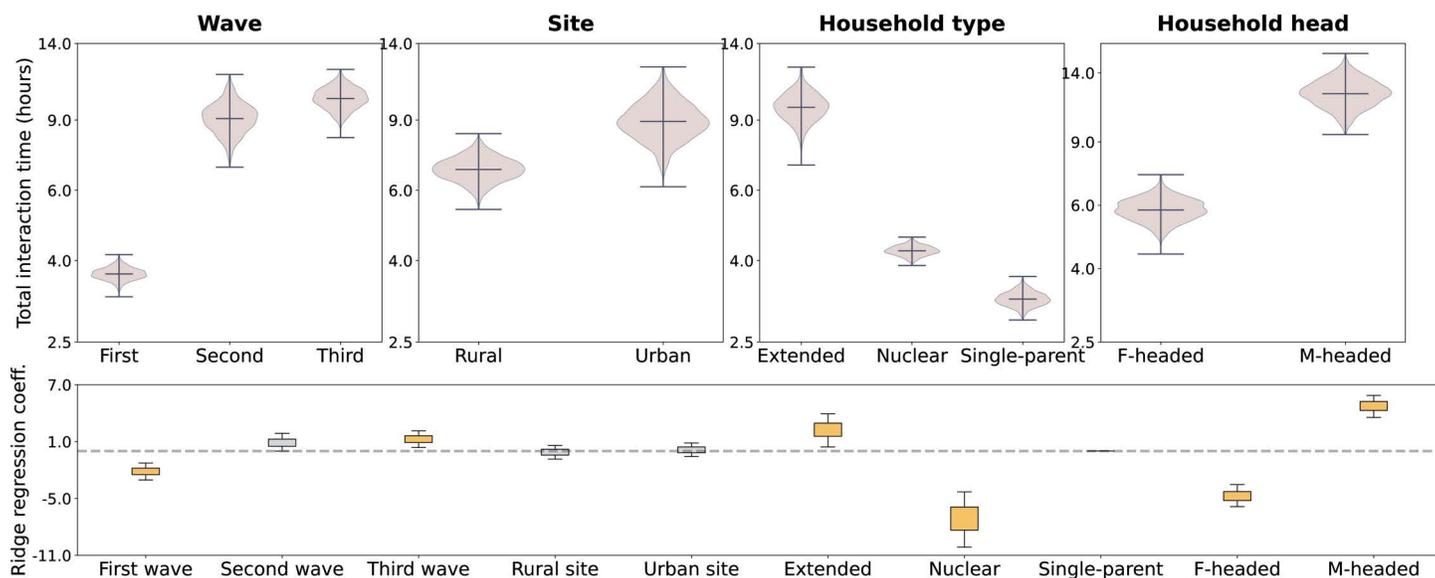
Fig 3. Household interaction time per individual. *Top row*: daily *household interaction times normalized by the population, disaggregated by wave, site, household type, and gender of the household head.* The *y*-axis indicates the sum of all entries on the upper diagonal of the normalized household contact matrices shown in Fig 2. The distributions are obtained from a bootstrap sample, as described in Sect 4.4. *Bottom row: ridge regression.* Box plot of the ridge regression coefficients obtained from several bootstrap randomizations. The whiskers show the $5 - 95$ confidence interval. The face color is gray if the confidence interval crosses zero, and it is orange otherwise.

(d = 1.5). The regression shows that the "Nuclear" and the "Extended" household types positively predict interaction rates between female adolescents and children, while the "Nuclear" type negatively predicts interactions with male adolescents. Moreover, the gender of the household head is a statistically significant predictor of the interactions between male adolescents and children.

2.2 Epidemiological impact of age and gender stratification according to the household classification

Multi-class compartmental models, such as the SIR (Susceptible-Infected-Recovered) model, are widely used in epidemiology [53]. In this model, susceptible individuals (S) become infected (I) when entering into contact with already infected individuals. After some time, infected individuals recover (R). By including contact matrices in this model, one can account for across-group heterogeneities in the social mixing behavior. In this model, two parameters – β and μ – quantify the probability per unit time of infection and recovery, respectively. These parameters are disease-dependent and determine the basic reproductive number R_0, which is the average number of secondary infections caused by a single infected person in a fully susceptible population. Its value crucially determines the epidemic outcome since if $R_0 < 1$, the epidemic likely dies quickly, while if $R_0 > 1$, it is expected to spread rapidly through the population. For the SIR model, the R_0 has a known expression depending on the disease parameters β, μ and on the contact matrix R [54]:

$$R_0 = \frac{\beta \rho(RN^{-1})}{\mu} \, ,$$

(1)

where ρ denotes the largest eigenvalue of a matrix, and N is the diagonal matrix containing the number of people in each group, specifically $N_{aa} = n_a$. For a given infectious disease, the parameters β, μ are fixed, and variations to
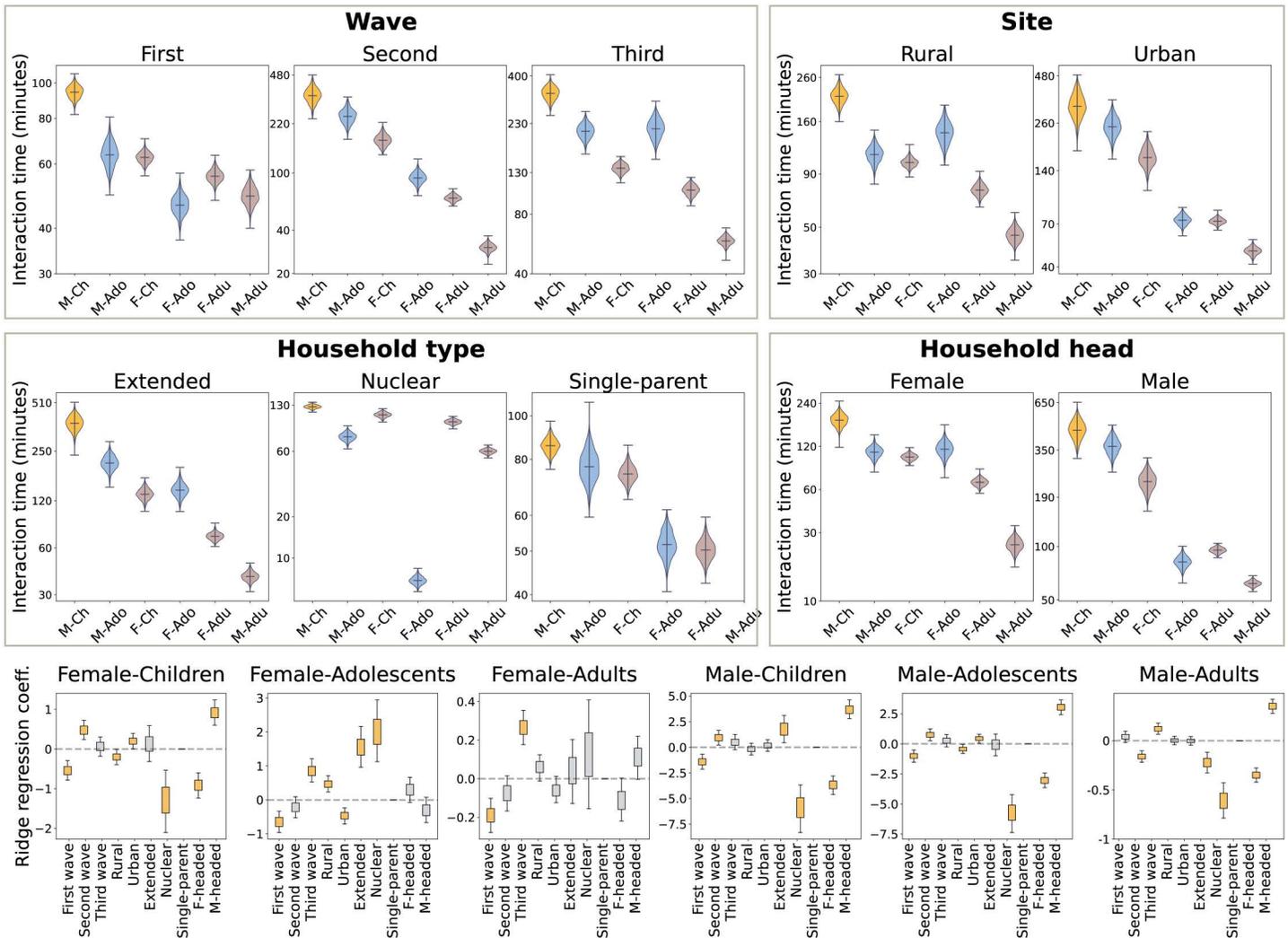
Fig 4. Interaction rates disaggregated by age and gender. *Top row: daily interaction times by age and gender, normalized by the population and disaggregated by wave, site, household type, and gender of the household head.* The y-axis indicates the sum of the normalized household contact matrices rows. The distributions are obtained from a bootstrap sample, as described in Sect 4.4, and aggregated according to the measurement wave (first frame), the measurement site (second frame), the household type (third frame), and the gender of the household head (fourth frame). We highlight in orange the interactions involving male children *M-Ch* and in light blue those involving adolescents *M/F-Adu*. *Bottom row: ridge regression.* Box plot of the ridge regression coefficients obtained from several bootstrap randomizations. The whiskers show the $5-95$ confidence interval. The face color is gray if the confidence interval crosses zero, and it is orange otherwise.

$\rho(RN^{-1})$ across household groups imply differences in the epidemic outcome. As recently shown in [13], when combining multiple levels of stratification to define groups (for example, age-gender instead of only age), Equation (1) still holds, but the spectral radius ρ – *i.e.,* the largest eigenvalue – of the normalized contact matrix RN^{-1} (and, consequently, R_0) cannot decrease. Therefore, single-attribute stratifications may be insufficient for capturing relevant heterogeneity patterns.

The matrices shown in Fig 2 utilize a two-level stratification that combines age and gender. To evaluate the significance of this stratification, Fig 6 compares the spectral radius of three contact matrices for each of the household classifications introduced in Sect 2: one using only gender for indexing groups, another using only age, and a third combining both age and
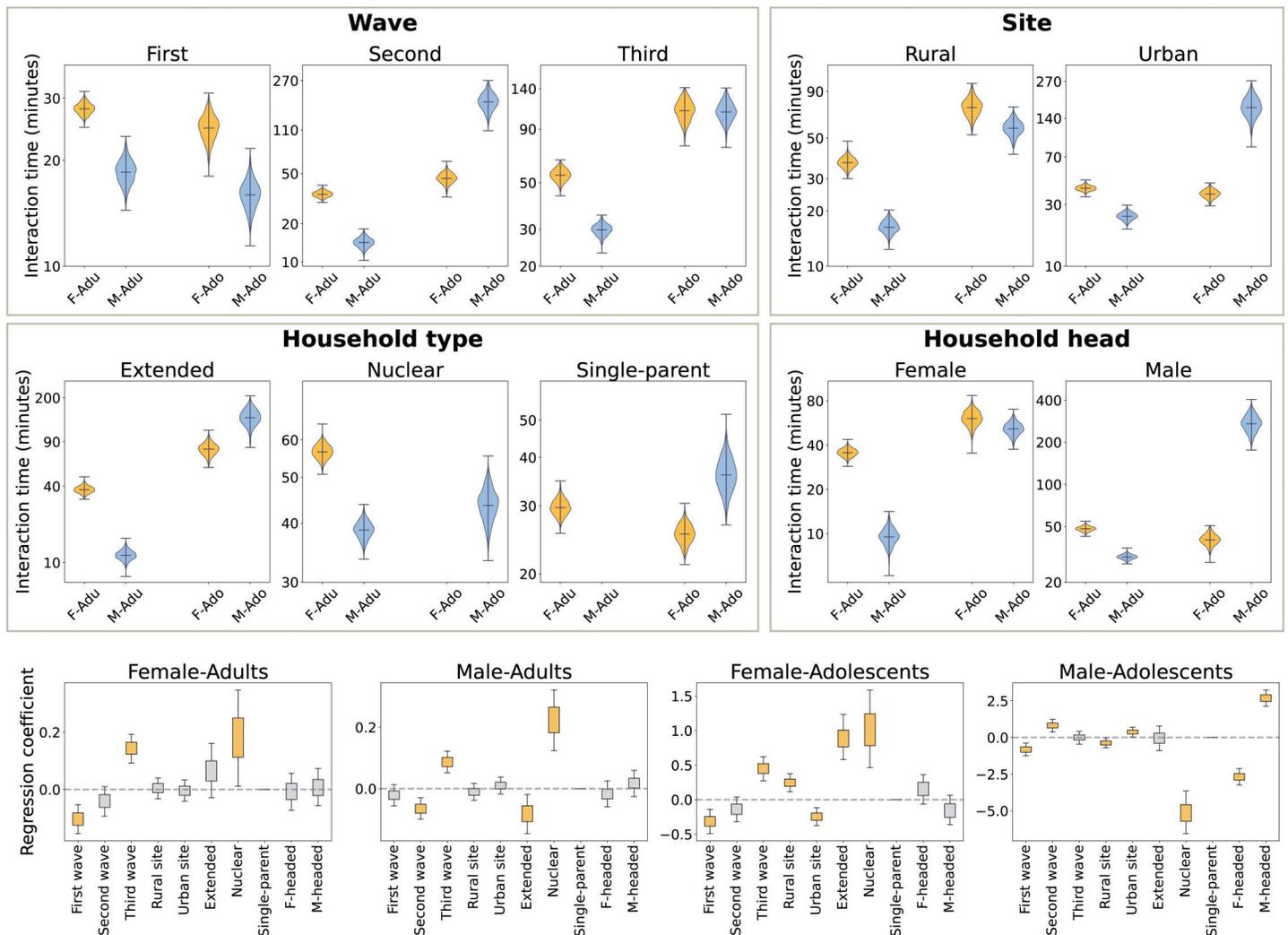
Fig 5. Interactions with children. *Top row: daily interaction rates by age and gender normalized by the population, disaggregated by wave, site, household type, and gender of the household head.* The plot shows the histograms of the interaction rates with children (male and female) per individual, for adults and adolescents of both genders. The histograms are obtained from the bootstrap sample described in Sect 4.4 section and aggregated according to the measurement wave (top left frame), the measurement site (top right frame), the household type (bottom left frame), and the gender of the household head (bottom right frame). *Bottom row: ridge regression.* Box plot of the ridge regression coefficients obtained from several bootstrap randomizations. The whiskers show the $5-95$ confidence interval. The face color is gray if the confidence interval crosses zero, and it is orange otherwise.

https://doi.org/10.1371/journal.pone.0344732.g005

gender. The results are obtained for several bootstrap randomizations, as detailed in Sect 4.4. As expected, age accounts for a higher level of heterogeneity than gender, but less than the combination of both attributes. Table 1 reports the size effects between the distributions of R_0, comparing "Gender"-only and "Age"-only contact matrices with "Age-Gender" matrices, testifying the importance of including both variables to explain the contact patterns in extended households where the complex patterns of interactions described in the previous sections lead to the largest increase in R_0 when using both variables at once.

3 Discussion

We conducted an analysis of the relation between in-household contacts and household attributes. Our findings show the frequency of in-household interactions increases during winter, in agreement with other studies in
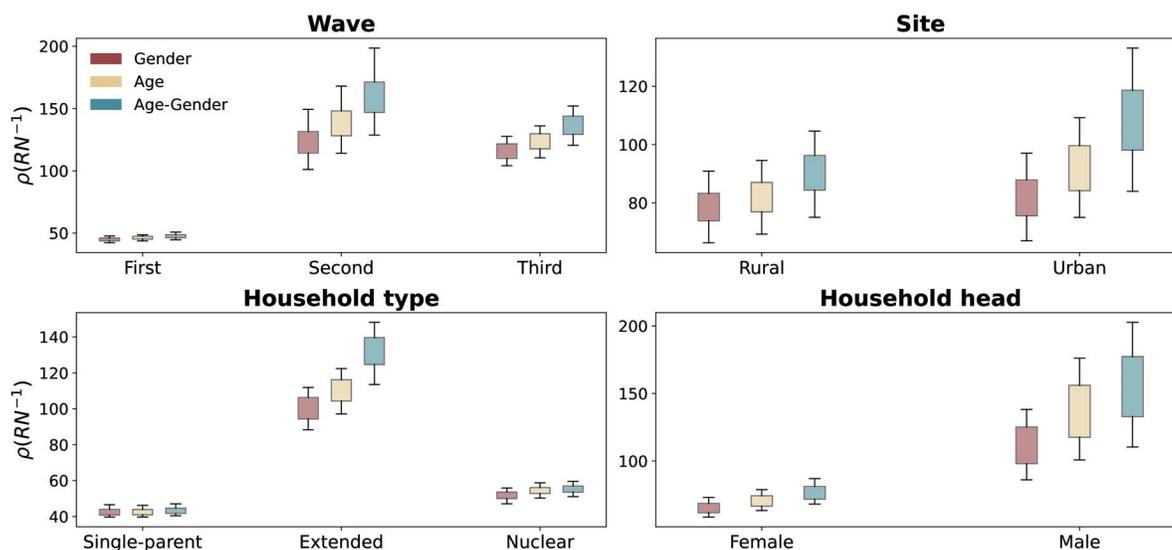
Fig 6. Largest eigenvalue of the contact matrix for different aggregation and indexing strategies. Each panel shows an aggregation based on a specific attribute indicated by the title. We consider three indexing strategies (shown on the x-axis) for each matrix: age, gender, and age-gender. The y-axis shows $\rho(RN^{-1})$, the largest eigenvalue of contact matrices the normalized contact matrix, which is proportional to R_0. The box plots are obtained for several bootstrap randomizations, as detailed in Sect 4.4. The whiskers indicate the $5-95$ confidence interval.

https://doi.org/10.1371/journal.pone.0344732.g006

Table 1. Effect size comparison between contact matrix specifications on the basic reproductive number R_0.

Wave	First	Second	Third
Gender	1.52	2.05	2.34
Age	0.69	1.14	1.38
Site	**Urban**	**Rural**	
Gender	1.44	2.13	
Age	0.97	1.25	
Household type	**Single-parent**	**Extended**	**Nuclear**
Gender	0.37	3.31	1.32
Age	0.39	2.16	0.32
Household head	**Female**	**Male**	
Gender	1.91	2.19	
Age	1.0	0.83	

We compare three contact matrix types: "Gender"-only (red in Fig 6) "Age"-only (yellow), and "Age-Gender" (blue). For each measurement wave, site, household type, and gender of household head, we compute R_0 using Equation (1) with each matrix type. The table reports the d score between "Age-Gender" and the two simpler specifications ("Gender"-only vs. "Age-Gender", and "Age"-only vs. "Age-Gender"). Higher d' values indicate larger differences in the R_0 distributions.

https://doi.org/10.1371/journal.pone.0344732.t001

Western countries that observed people spending more time outside the household during summer [55,56]. Children, and in particular male children, are responsible for the highest contact rates, supporting the known fact that children are the primary introducers of influenza-like illnesses in households [40,41]. However, accounting for the household composition is necessary to interpret the observed contact patterns better. Extended households are generally larger than nuclear and single-parent households and have higher interaction rates on average. Most of the activity involving male children is recorded in these households, possibly due to the presence of siblings

and cousins of similar age. In these households, adult men interact less than women (including adolescents). This observation agrees with the fact that in many African cultures, including South Africa, women are identified as the primary caregivers for children, even when they are not the biological mothers [57]. In extended households, women are often in charge of caregiving also for the elderly [58]. From an epidemiological perspective, these observations might provide support to the known fact that large and multigenerational households experience higher risks of transmission of diseases such as influenza, RSV, and measles [40,42–44]. In nuclear households, adult males experience higher interaction rates with children than in extended households. Fathers and mothers seem to undertake similar roles in child-rearing in this type of household. All single-parent households in our dataset are female-headed. This agrees with a global picture with 84% of single-parent households run by women [59]. The most common reasons explaining the gender gap are widowhood, divorce, the presence of children outside marriage, and migrant workers [60].

Female-headed households are more prevalent in rural settings and display a generally higher involvement of women. Nonetheless, we observed that male-headed have higher activity rates than female-headed households. This is a non-obvious result because female-headed households tend to be larger than male-headed households as they mostly fall under the extended category.

The differences between the rural and urban settings are mainly ascribed to a greater diffusion in rural settings of extended and female-headed households. The higher contact rates we observed in rural areas are consistent with previous results in Zimbabwe [22], but in contrast to other studies in similar contexts from low-resource settings in the sub-Saharan region [21,25,31–33], suggesting that the rural-urban context might not be as influential as the household composition or gender head. The higher involvement of women in rural contexts agrees with a societal norm that heavily influences household roles. The relatively better balance between gender roles in the urban setting can be attributed to relaxed societal gender norms [61]. Evidence from Zambia [62] also suggests that in urban areas, caregiving responsibilities are becoming more balanced as men and male adolescents are increasingly involved in care duties, explaining the high involvement of males in interactions in the urban area. We remark that differences in results between our studies and other published studies could be, in part, as a result of different data collection methods.

We evaluated the epidemiological implications of observed differences in contact patterns by computing a proxy of the basic reproductive number associated with contact matrices collected in various social contexts and different seasons. The R_0 is mainly driven by age-based heterogeneities, even though the full degree of heterogeneity becomes apparent when accounting for both age and gender. Notably, the most significant variations are observed with changes in seasonality and in extended households, where complex relations are observed between age and gender attributes.

Let us now comment on the main limitations of our work. Our analysis is based on a small number of households. In some cases, only a few households are considered when disaggregating the data by site, type, and headship. To cope with this model and attribute statistical significance to our results, we performed bootstrap sampling. Given the small size of our dataset, however, our results are not representative of the population, and some household types (for instance, extended households) are over-represented in our dataset compared to the South-African population. Given the biases present in our dataset, our analysis cannot be interpreted as hypothesis-generating and conclusive. Moreover, the dataset used for this study contains only in-household interactions, and our conclusions do not apply to the across-household interactions that are fundamental in epidemic modeling. Despite these limitations, our results show the important role played by the household structure and composition in shaping face-to-face interactions. These results contribute to addressing the limited data gap on face-to-face interactions in sub-Saharan Africa, where the prevalence and re-emergence of infectious diseases pose a significant public health challenge. Our results encourage future data collection on face-to-face interactions to include rich metadata information related to the social role of individuals in the household.

4 Methods

4.1 Ethics

Ethics permission to conduct the experiment was received from the Wits Human Research Ethics Committee (Medical) (ethics reference no. 150808). The project and its implementation were also discussed with community advisory boards at both study sites. The study was designed and implemented by South Africans. In each study site, the collaborators implementing the study had a long-standing relationship with the community. Informed consent documents were translated into the local languages and discussed with community advisory boards. Written informed consent/assent was obtained from all study participants.

4.2 Data description

The PHIRST study [25,28,49] was a perspective household cohort study, conducted in South Africa during $2016 - 2018$ to investigate the burden and transmission dynamics of influenza and respiratory syncytial virus (RSV). Randomly selected households were approached for enrolment, and households with more than 2 members and where at least 80% of household members consented, were included in the study. A new cohort was enrolled each year, with enrolment occurring in the November through to December the prior year, and intensive follow up occurring from February – October with twice weekly home visits and nasopharyngeal swab collection for testing of influenza and (RSV), irrespective of symptoms. Recruitment occurred at the end of the preceding year to ensure all households were enrolled before the start of the RSV season which starts in February/March in South Africa. A household contact survey was nested in PHIRST in 2018, in which all households included in the 2018 cohort were approached to partake. During this data collections, contact patterns were measured with the proximity sensors developed by the `SocioPatterns` collaboration (sociopatterns.org, [50]). These sensors are non-obtrusive devices worn on the chest of the study participants during the data collection. They emit low-power radio signals, exchanging approximately 40 information packets per second when facing one another. They record a proximity interaction if two sensors exchange at least one packet in a 20-second period. In this case, each sensor records the unique identifier of another interacting sensor, the interaction time, and the power attenuation from the sender to the receiver. Proximity interactions within approximately 1.5 meters are identified by filtering the measurements based on their signal attenuation. These sensors provide high-resolution measurements of human interactions and have been deployed in several contexts, including schools [63,64], hospitals [65], and conferences [66], among others. The proximity sensors were packaged in PVC pouches and worn as a badge pinned to either the inside or outside of the clothing with a safety pin, or on a lanyard around the neck, depending on the preference of the participant. Field workers visited the household at least two times during the deployment and reminded participants to wear the sensors. We followed the data cleaning procedure described in [28], which we here briefly summarize.

1. Only interactions between members of the same households are kept, as they are the only ones needed to define household contact matrices.

2. The field team performing the data collection reported the dates of beginning and end of each deployment for every household. All the interactions recorded beyond the official measurement window are dropped. The compliance with the experiment protocol was extremely variable in those days, making these measurements unreliable.

3. We removed the interactions occurring during the first and the last days of each deployment. During these days, we observed anomalous activity patterns due to the sensors' dispatching.

4. Some households were excluded because of the impossibility of correctly mapping the sensor data with some (or all) of the household members.

5. All households with fewer than days of valid measurements per wave were excluded. We recall that in each measurement wave we collected data for at least *10* days.

6. Households with non-circadian activity patterns (amenable to a misuse of the sensors) were excluded.

   Our cleaned dataset is composed of *307* individuals and *60* households.

4.3 Household classification

Our results are based on a household classification made according to the site in which household are located, its headship and its type. Below we provide additional details describing and motivating this classification.

• **Site (rural or urban).** Socio-demographic characteristics of individuals and households differ across rural and urban settings. Hence, it is important to capture differences and similarities in contact patterns driven by these attributes. The rural site is composed of *28* households, and the urban by *32*.

• **Gender of the household head.** The household head does not have a unique definition, but it generally refers to the person who holds the economic responsibility (for example being the primary breadwinner) or who has the most decision-making power in the household [67]. In our study, the household head is the person identified by the household members as the primary decision-maker. Especially in South Africa, this role is commonly held by the oldest individuals who typically also have the highest income [68]. Regardless of the actual dynamics within the household, men are commonly more likely to be designated as household heads [67], while women-headed households are more susceptible to economic hardships [68]. Our dataset comprises *43* female-headed households (*24* in Agincourt, *19* in Klerksdorp) and *17* male-headed households (*4* in Agincourt and *13* in Klerksdorp). This agrees with the general trend in South Africa that sees female-headed households more prevalent in rural areas [69].

• **Household type.** For this classification, we follow [36], focusing only on the household types observed in our cohort. These include: (i) *nuclear households* (*12* households), formed by a couple (married or cohabiting) and their biological or adopted children; (ii) *single-parent* (*13* households) in which the household head lives with his/her children; (iii) *extended households* (*33* households), composed by relatives beyond nuclear family such as cousins, aunts, uncles, and grandparents. Our dataset further comprises two households classified as *complex* as they also contain non-relative members. Given that only two households belong to this class, we chose to absorb them into the extended and nuclear households, according to their composition. In our dataset, extended households are larger on average with 5.7 individuals against 3.9 and 4.2 observed in nuclear and single-parent households, respectively. According to the General Household Survey [70], nuclear households are the most common in South Africa, reaching 40.1% of the total. In our dataset, extended households are over-represented in both contexts: 57% in Agincourt versus 45% observed in rural areas; 53% in Klerksdorp versus 30% observed in urban areas. On the opposite, nuclear households are less represented than expected but, in agreement with the national trend, they are more frequent in the urban area (28% in Klerksdorp) than in the rural one (11% in Agincourt).

4.4 Contact matrices definition

For each household h in the dataset, and a pair of age-gender indices a, b (for instance a = male-children, b = female adults), we denote with $R_{ab}^{(h)}$ the daily average interaction duration aggregated between the groups a and b. We further let $n_a^{(h)}$ be the number of individuals in age-gender group a in household h. By letting α be one of the household attributes described in Sect 4.2 (*e.g.*, "male-headed", or "rural"), we denote with χ_α a random sample with repetition of 1000 households having the attribute a. The corresponding household contact matrix (HCM) is defined by the following equation

$$\tilde{R}_{a,b}(\chi_\alpha) = \frac{\langle R_{ab}\rangle_{\chi_\alpha}}{\langle n_a\rangle_{\chi_\alpha}\langle n_b\rangle_{\chi_\alpha}} \ ,$$

(2)

where $\langle M\rangle_{\chi_\alpha} = \frac{1}{|\chi_\alpha|}\sum_{h\in\chi_\alpha} M^{(h)}$ is the empirical average of the attribute M over the set χ_α. For instance, let a denote male-headed households, then the matrix $\tilde{R}(\chi_\alpha)$ is the empirical average of all household contact matrices collected in male-headed households, and normalized by the average population sizes in these households. Note that each household can appear more than once in the set χ_α. The bootstrap distribution is obtained over 1000 random samples of the set χ_α.

4.5 Ridge regression

To disentangle the dependencies of the household attributes in shaping the interaction patterns, we run a ridge regression in which we predict the contact rates from a list of household attributes. First, we define a sample χ_i of 100 (households-deployment) pairs drawn with repetitions from the whole dataset, composed of 180 instances. Using Eq (2), we obtain the corresponding matrix $\tilde{R}_{\chi_i}$. We then extract a scalar quantity y_i from R_{χ_i} (*e.g.*, the total interaction time), that needs to be predicted from the regression model. Finally, we define a feature vector $\boldsymbol{x}_i$ whose entries indicate the number of sampled instances per attribute. The attributes are first, second, and third wave, nuclear, single-parent, and extended households, male and female-headed households. By repeating the process for different randomizations $i = 1, \ldots, 1000$, we obtain a target vector $\boldsymbol{y}$ and a feature matrix X that we use as inputs for the ridge regression. We then repeat the process *1000* times and obtain a distribution of the fitted parameters, which we consider to be significant only if the $5 - 95$ confidence interval does not cross zero.

Author contributions

Conceptualization: Stefano Tempia, Cheryl Cohen, Daniela Paolotti, Ciro Cattuto, Lorenzo Dall'Amico.

Data curation: Jackie Kleynhans.

Formal analysis: Kausutua Tjikundi.

Funding acquisition: Cheryl Cohen, Ciro Cattuto.

Investigation: Kausutua Tjikundi.

Software: Kausutua Tjikundi, Lorenzo Dall'Amico.

Supervision: Daniela Paolotti, Ciro Cattuto, Lorenzo Dall'Amico.

Writing – original draft: Kausutua Tjikundi.

Writing – review & editing: Jackie Kleynhans, Stefano Tempia, Cheryl Cohen, Daniela Paolotti, Ciro Cattuto, Lorenzo Dall'Amico.

References

1. Wallinga J, Edmunds WJ, Kretzschmar M. Perspective: human contact patterns and the spread of airborne infectious diseases. Trends Microbiol. 1999;7(9):372–7. https://doi.org/10.1016/s0966-842x(99)01546-2 PMID: 10470046

2. Anderson RM, May RM. Infectious diseases of humans: dynamics and control. Oxford University Press; 1992.

3. Rohani P, Zhong X, King AA. Contact network structure explains the changing epidemiology of pertussis. Science. 2010;330(6006):982–5. https://doi.org/10.1126/science.1194134 PMID: 21071671

4. Edmunds WJ, O'Callaghan CJ, Nokes DJ. Who mixes with whom? A method to determine the contact patterns of adults that may lead to the spread of airborne infections. Proc Biol Sci. 1997;264(1384):949–57. https://doi.org/10.1098/rspb.1997.0131 PMID: 9263464

5. Wallinga J, Teunis P, Kretzschmar M. Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents. Am J Epidemiol. 2006;164(10):936–44. https://doi.org/10.1093/aje/kwj317 PMID: 16968863

6. Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. PLoS Med. 2008;5(3):e74. https://doi.org/10.1371/journal.pmed.0050074 PMID: 18366252

7. Gozzi N, Tizzoni M, Chinazzi M, Ferres L, Vespignani A, Perra N. Estimating the effect of social inequalities on the mitigation of COVID-19 across communities in Santiago de Chile. Nat Commun. 2021;12(1):2429. https://doi.org/10.1038/s41467-021-22601-6 PMID: 33893279

8. Gauvin L, Bajardi P, Pepe E, Lake B, Privitera F, Tizzoni M. Socio-economic determinants of mobility responses during the first wave of COVID-19 in Italy: from provinces to neighbourhoods. J R Soc Interface. 2021;18(181):20210092. https://doi.org/10.1098/rsif.2021.0092 PMID: 34343450

9. Buckee C, Noor A, Sattenspiel L. Thinking clearly about social aspects of infectious disease transmission. Nature. 2021;595(7866):205–13. https://doi.org/10.1038/s41586-021-03694-x PMID: 34194045

10. Tizzoni M, Nsoesie EO, Gauvin L, Karsai M, Perra N, Bansal S. Addressing the socioeconomic divide in computational modeling for infectious diseases. Nat Commun. 2022;13(1):2897. https://doi.org/10.1038/s41467-022-30688-8 PMID: 35610237

11. Zelner J, Masters NB, Naraharisetti R, Mojola SA, Chowkwanyun M, Malosh R. There are no equal opportunity infectors: Epidemiological modelers must rethink our approach to inequality in infection risk. PLoS Comput Biol. 2022;18(2):e1009795. https://doi.org/10.1371/journal.pcbi.1009795 PMID: 35139067

12. Manna A, Karsai M, Perra N. Social inequalities in vaccine coverage and their effects on epidemic spreading. PLoS Comput Biol. 2025;21(10):e1013585. https://doi.org/10.1371/journal.pcbi.1013585 PMID: 41082566

13. Manna A, Dall'Amico L, Tizzoni M, Karsai M, Perra N. Generalized contact matrices allow integrating socioeconomic variables into epidemic models. Sci Adv. 2024;10(41):eadk4606. https://doi.org/10.1126/sciadv.adk4606 PMID: 39392883

14. Manna A, Koltai J, Karsai M. Importance of social inequalities to contact patterns, vaccine uptake, and epidemic dynamics. Nat Commun. 2024;15(1):4137. https://doi.org/10.1038/s41467-024-48332-y PMID: 38755162

15. Ajelli M, Poletti P, Melegaro A, Merler S. The role of different social contexts in shaping influenza transmission during the 2009 pandemic. Sci Rep. 2014;4:7218. https://doi.org/10.1038/srep07218 PMID: 25427621

16. Mistry D, Litvinova M, Pastore Y Piontti A, Chinazzi M, Fumanelli L, Gomes MFC, et al. Inferring high-resolution human mixing patterns for disease modeling. Nat Commun. 2021;12(1):323. https://doi.org/10.1038/s41467-020-20544-y PMID: 33436609

17. Goeyvaerts N, Santermans E, Potter G, Torneri A, Van Kerckhove K, Willem L, et al. Household members do not contact each other at random: implications for infectious disease modelling. Proc Biol Sci. 2018;285(1893):20182201. https://doi.org/10.1098/rspb.2018.2201 PMID: 30963910

18. Longini IM Jr, Koopman JS, Monto AS, Fox JP. Estimating household and community transmission parameters for influenza. Am J Epidemiol. 1982;115(5):736–51. https://doi.org/10.1093/oxfordjournals.aje.a113356 PMID: 7081204

19. Hoang T, Coletti P, Melegaro A, Wallinga J, Grijalva CG, Edmunds JW, et al. A Systematic Review of Social Contact Surveys to Inform Transmission Models of Close-contact Infections. Epidemiology. 2019;30(5):723–36. https://doi.org/10.1097/EDE.0000000000001047 PMID: 31274572

20. Organization WH. World health statistics 2024: monitoring health for the SDGs, sustainable development goals. World Health Organization; 2024.

21. Kiti MC, Kinyanjui TM, Koech DC, Munywoki PK, Medley GF, Nokes DJ. Quantifying age-related rates of social contact using diaries in a rural coastal population of Kenya. PLoS One. 2014;9(8):e104786. https://doi.org/10.1371/journal.pone.0104786 PMID: 25127257

22. Melegaro A, Del Fava E, Poletti P, Merler S, Nyamukapa C, Williams J, et al. Social Contact Structures and Time Use Patterns in the Manicaland Province of Zimbabwe. PLoS One. 2017;12(1):e0170459. https://doi.org/10.1371/journal.pone.0170459 PMID: 28099479

23. le Polain de Waroux O, Cohuet S, Ndazima D, Kucharski AJ, Juan-Giner A, Flasche S, et al. Characteristics of human encounters and social mixing patterns relevant to infectious diseases spread by close contact: a survey in Southwest Uganda. BMC Infect Dis. 2018;18(1):172. https://doi.org/10.1186/s12879-018-3073-1 PMID: 29642869

24. le Polain de Waroux O, Flasche S, Kucharski AJ, Langendorf C, Ndazima D, Mwanga-Amumpaire J, et al. Identifying human encounters that shape the transmission of Streptococcus pneumoniae and other acute respiratory infections. Epidemics. 2018;25:72–9. https://doi.org/10.1016/j.epidem.2018.05.008 PMID: 30054196

25. Kleynhans J, Tempia S, McMorrow ML, von Gottberg A, Martinson NA, Kahn K. A cross-sectional study measuring contact patterns using diaries in an urban and a rural community in South Africa, 2018. BMC Public Health. 2021;21:1–10.

26. Kiti MC, Tizzoni M, Kinyanjui TM, Koech DC, Munywoki PK, Meriac M, et al. Quantifying social contacts in a household setting of rural Kenya using wearable proximity sensors. EPJ Data Sci. 2016;5:21. https://doi.org/10.1140/epjds/s13688-016-0084-2 PMID: 27471661

27. Ozella L, Paolotti D, Lichand G, Rodríguez JP, Haenni S, Phuka J, et al. Using wearable proximity sensors to characterize social contact patterns in a village of rural Malawi. EPJ Data Sci. 2021;10(1). https://doi.org/10.1140/epjds/s13688-021-00302-w

28. Dall'Amico L, Kleynhans J, Gauvin L, Tizzoni M, Ozella L, Makhasi M, et al. Estimating household contact matrices structure from easily collectable metadata. PLoS One. 2024;19(3):e0296810. https://doi.org/10.1371/journal.pone.0296810 PMID: 38483886

29. Kleynhans J, Dall'Amico L, Gauvin L, Tizzoni M, Maloma L, Walaza S, et al. Association of close-range contact patterns with SARS-CoV-2: a household transmission study. Elife. 2023;12:e84753. https://doi.org/10.7554/eLife.84753 PMID: 37461328

30. Osei I, Mendy E, van Zandvoort K, Jobe O, Sarwar G, Wutor BM, et al. Directly observed social contact patterns among school children in rural Gambia. Epidemics. 2024;49:100790. https://doi.org/10.1016/j.epidem.2024.100790 PMID: 39270441

31. Del Fava E, Adema I, Kiti MC, Poletti P, Merler S, Nokes DJ, et al. Individual's daily behaviour and intergenerational mixing in different social contexts of Kenya. Sci Rep. 2021;11(1):21589. https://doi.org/10.1038/s41598-021-00799-1 PMID: 34732732

32. Dodd PJ, Looker C, Plumb ID, Bond V, Schaap A, Shanaube K, et al. Age- and Sex-Specific Social Contact Patterns and Incidence of Mycobacterium tuberculosis Infection. Am J Epidemiol. 2016;183(2):156–66. https://doi.org/10.1093/aje/kwv160 PMID: 26646292

33. Kiti MC, Sacoor C, Aguolu OG, Zelaya A, Chen H, Kim SS, et al. Social Contact Patterns in Rural and Urban Settings, Mozambique, 2021–2022. Emerg Infect Dis. 2025;31(1):94.

34. van Zandvoort K, Bobe MO, Hassan AI, Abdi MI, Ahmed MS, Soleman SM, et al. Social contacts and other risk factors for respiratory infections among internally displaced people in Somaliland. Epidemics. 2022;41:100625. https://doi.org/10.1016/j.epidem.2022.100625 PMID: 36103782

35. Johnstone-Robertson SP, Mark D, Morrow C, Middelkoop K, Chiswell M, Aquino LDH, et al. Social mixing patterns within a South African township community: implications for respiratory disease transmission and control. Am J Epidemiol. 2011;174(11):1246–55. https://doi.org/10.1093/aje/kwr251 PMID: 22071585

36. Ibisomi LD, De Wet N. The dynamics of household structure in sub-Saharan Africa. Continuity and change in sub-Saharan African demography. 2014. 173 p.

37. Noumbissi A, Bawah AA. Parental survival and residential patterns. The demography of South Africa. 2016. 160 p.

38. Madhavan S, Myroniuk TW, Kuhn R, Collinson MA. Household structure vs. composition: Understanding gendered effects on educational progress in rural South Africa. Demogr Res. 2017;37:1891–916. https://doi.org/10.4054/DemRes.2017.37.59 PMID: 29270077

39. Menashe-Oren A, Compaoré Y, Bocquier P, Ginsburg C. Dynamic household structure and composition: a manual for longitudinal analysis of living arrangements. BMC Res Notes. 2023;16(1):223. https://doi.org/10.1186/s13104-023-06485-x PMID: 37726762

40. Endo A, Uchida M, Kucharski AJ, Funk S. Fine-scale family structure shapes influenza transmission risk in households: Insights from primary schools in Matsumoto city, 2014/15. PLoS Comput Biol. 2019;15(12):e1007589. https://doi.org/10.1371/journal.pcbi.1007589 PMID: 31877122

41. Ozella L, Gesualdo F, Tizzoni M, Rizzo C, Pandolfi E, Campagna I, et al. Close encounters between infants and household members measured through wearable proximity sensors. PLoS One. 2018;13(6):e0198733. https://doi.org/10.1371/journal.pone.0198733 PMID: 29879196

42. Mahikul W, J White L, Poovorawan K, Soonthornworasiri N, Sukontamarn P, Chanthavilay P, et al. Modeling household dynamics on Respiratory Syncytial Virus (RSV). PLoS One. 2019;14(7):e0219323. https://doi.org/10.1371/journal.pone.0219323 PMID: 31287832

43. Hilton J, Keeling MJ. Incorporating household structure and demography into models of endemic disease. J R Soc Interface. 2019;16(157):20190317. https://doi.org/10.1098/rsif.2019.0317 PMID: 31387486

44. Chisholm RH, Crammond B, Wu Y, Bowen AC, Campbell PT, Tong SYC, et al. A model of population dynamics with complex household structure and mobility: implications for transmission and control of communicable diseases. PeerJ. 2020;8:e10203. https://doi.org/10.7717/peerj.10203 PMID: 33194407

45. Mousa A, Winskill P, Watson OJ, Ratmann O, Monod M, Ajelli M, et al. Social contact patterns and implications for infectious disease transmission - a systematic review and meta-analysis of contact surveys. Elife. 2021;10:e70294. https://doi.org/10.7554/eLife.70294 PMID: 34821551

46. House T, Keeling MJ. Household structure and infectious disease transmission. Epidemiol Infect. 2009;137(5):654–61. https://doi.org/10.1017/S0950268808001416 PMID: 18840319

47. Wing K, Grint DJ, Mathur R, Gibbs HP, Hickman G, Nightingale E, et al. Association between household composition and severe COVID-19 outcomes in older people by ethnicity: an observational cohort study using the OpenSAFELY platform. Int J Epidemiol. 2022;51(6):1745–60. https://doi.org/10.1093/ije/dyac158 PMID: 35962974

48. Møgelmose S, Vijnck L, Neven F, Neels K, Beutels P, Hens N. Population age and household structures shape transmission dynamics of emerging infectious diseases: a longitudinal microsimulation approach. J R Soc Interface. 2023;20(209):20230087. https://doi.org/10.1098/rsif.2023.0087 PMID: 38053386

49. Cohen C, McMorrow ML, Martinson NA, Kahn K, Treurnicht FK, Moyes J, et al. Cohort profile: A Prospective Household cohort study of Influenza, Respiratory syncytial virus and other respiratory pathogens community burden and Transmission dynamics in South Africa, 2016-2018. Influenza Other Respir Viruses. 2021;15(6):789–803. https://doi.org/10.1111/irv.12881 PMID: 34296810

50. Cattuto C, Van den Broeck W, Barrat A, Colizza V, Pinton J-F, Vespignani A. Dynamics of person-to-person interactions from distributed RFID sensor networks. PLoS One. 2010;5(7):e11596. https://doi.org/10.1371/journal.pone.0011596 PMID: 20657651

51. Gupta S, Anderson RM, May RM. Networks of sexual contacts: implications for the pattern of spread of HIV. AIDS. 1989;3(12):807–17. PMID: 2517202

52. Prem K, Cook AR, Jit M. Projecting social contact matrices in 152 countries using contact surveys and demographic data. PLoS Comput Biol. 2017;13(9):e1005697. https://doi.org/10.1371/journal.pcbi.1005697 PMID: 28898249

53. Keeling MJ, Rohani P. Modeling infectious diseases in humans and animals. Princeton University Press; 2008.

54. Diekmann O, Heesterbeek JAP, Roberts MG. The construction of next-generation matrices for compartmental epidemic models. J R Soc Interface. 2010;7(47):873–85. https://doi.org/10.1098/rsif.2009.0386 PMID: 19892718

55. Leech JA, Nelson WC, Burnett RT, Aaron S, Raizenne ME. It's about time: a comparison of Canadian and American time-activity patterns. J Expo Anal Environ Epidemiol. 2002;12(6):427–32. https://doi.org/10.1038/sj.jea.7500244 PMID: 12415491

56. Hussein T, Paasonen P, Kulmala M. Activity pattern of a selected group of school occupants and their family members in Helsinki—Finland. Sci Total Environ. 2012;425:289–92.

57. Hatch M, Posel D. Who cares for children? A quantitative study of childcare in South Africa. Dev South Afr. 2018;35(2):267–82. https://doi.org/10.1080/0376835x.2018.1452716

58. Chant S. Female household headship and the feminisation of poverty: facts, fictions and forward strategies. Gender Institute, London School of Economics and Political Science; 2003.

59. Equality UNEFG, the Empowerment of Women. Progress of the world's women 2019-2020: Families in a changing world. Un Women; 2019.

60. Desai S, Ahmad S. Female-headed households. In: Nelly PS, editor. Women in the Third World: An Encyclopedia of Contemporary Issues. New York: Garland Pub; 1998. p. 227–35.

61. Pozarny P. Gender roles and opportunities for women in urban environments. Governance and Social Development Resource Centre (GSDRC) Helpdesk Research Report. Birmingham: GSDRC and University of Birmingham; 2016.

62. Evans A. 'Women Can Do What Men Can Do': The Causes and Consequences of Growing Flexibility in Gender Divisions of Labour in Kitwe, Zambia. J South Afr Stud. 2014;40(5):981–98. https://doi.org/10.1080/03057070.2014.946214

63. Stehlé J, Voirin N, Barrat A, Cattuto C, Isella L, Pinton J-F, et al. High-resolution measurements of face-to-face contact patterns in a primary school. PLoS One. 2011;6(8):e23176. https://doi.org/10.1371/journal.pone.0023176 PMID: 21858018

64. Fournet J, Barrat A. Contact patterns among high school students. PLoS One. 2014;9(9):e107878. https://doi.org/10.1371/journal.pone.0107878 PMID: 25226026

65. Vanhems P, Barrat A, Cattuto C, Pinton J-F, Khanafer N, Régis C, et al. Estimating potential infection transmission routes in hospital wards using wearable proximity sensors. PLoS One. 2013;8(9):e73970. https://doi.org/10.1371/journal.pone.0073970 PMID: 24040129

66. Génois M, Barrat A. Can co-location be used as a proxy for face-to-face contacts? EPJ Data Science. 2018;7(1):1–18.

67. Chant S. Household Organisation and Survival in Developing Countries. In: Smith SJ, Elsinga M, Fox O'Mahony L, Ong SE, Wachter S, editors. International Encyclopedia of Housing and Home. Oxford, UK: Elsevier; 2011. p. 223–9.

68. Posel DR. Who are the heads of household, what do they do, and is the concept of headship useful? An analysis of headship in South Africa. Dev South Afr. 2001;18(5):651–70. https://doi.org/10.1080/03768350120097487

69. Dungumaro EW. Gender Differentials in Household Structure and Socioeconomic Characteristics in South Africa. J Comp Fam Stud. 2008;39(4):429–51. https://doi.org/10.3138/jcfs.39.4.429

70. Statistics South Africa. General Household Survey 2022. Pretoria, South Africa: Statistics South Africa; 2022.