scientific reports

OPEN

Individual's daily behaviour and intergenerational mixing in different social contexts of Kenya

Emanuele Del Fava[1,2,8], Irene Adema[3,8], Moses C. Kiti[3], Piero Poletti[4], Stefano Merler[4], D. James Nokes[3,5], Piero Manfredi[6] & Alessia Melegaro[1,7✉]

We investigated contact patterns in diverse social contexts in Kenya and the daily behaviours that may play a pivotal role in infection transmission to the most vulnerable leveraging novel data from a 2-day survey on social contacts and time use (TU) from a sample of 1407 individuals (for a total of 2705 person days) from rural, urban formal, and informal settings. We used TU data to build six profiles of daily behaviour based on the main reported activities, i.e., Homestayers (71.1% of person days), Workers (9.3%), Schoolers (7.8%), or locations at increasing distance from home, i.e., Walkers (6.6%), Commuters (4.6%), Travelers (0.6%). In the rural setting, we observed higher daily contact numbers (11.56, SD 0.23) and percentages of intergenerational mixing with older adults (7.5% of contacts reported by those younger than 60 years vs. less than 4% in the urban settings). Overall, intergenerational mixing with older adults was higher for Walkers (7.3% of their reported contacts), Commuters (8.7%), and Homestayers (5.1%) than for Workers (1.5%) or Schoolers (3.6%). These results could be instrumental in defining effective interventions that acknowledge the heterogeneity in social contexts and daily routines, either in Kenya or other demographically and culturally similar sub-Saharan African settings.

Individual social interactions are critical drivers of transmission processes underlying the spread of infectious diseases such as measles, influenza, and the novel coronavirus infection (SARS-CoV-2).

The availability of age-specific data on social contact patterns is fundamental to properly estimate the age-specific transmission parameters of an infectious disease transmitted directly from person to person, and to mimic the potential age heterogeneity characterising different infection episodes[1,2]. Furthermore, these data could be used to estimate critical quantities such as age-specific matrices of secondary cases, i.e., the next generation matrices, which allow to identify the transmission routes bringing the disease into the most vulnerable population segments, to estimate reproduction numbers[3], and to evaluate the impact of health policies aimed at mitigating the infection spread[4,5].

The need for information on social contact patterns becomes crucial when considering resource-poor settings, where the quality and availability of health care requires special efforts to define feasible and sustainable strategies to prevent transmission and reduce the disease burden. Evidence from social contact surveys has shown that, in low- and middle-income countries (LMICs), contact patterns are generally characterised by (i) a marked age assortativeness (i.e., individuals tend to interact more often with peers of similar age) concentrated among school-age children and young adults, and (ii) higher rates of intergenerational mixing compared to high-income settings. Common features from previous studies are that the largest daily number of contacts is experienced by school-aged children between 6 and 18 years old and working-age adults, while intergenerational contacts are mainly reported between household members[6–19].

[1]Carlo F. Dondena Centre for Research on Social Dynamics and Public Policy, Bocconi University, Milan, Italy. [2]Max Planck Institute for Demographic Research, Rostock, Germany. [3]KEMRI-Wellcome Trust Research Programme, Kilifi, Kenya. [4]Fondazione Bruno Kessler, Trento, Italy. [5]School of Life Sciences and Zeeman Institute for Systems Biology and Infectious Disease Epidemiology Research (SBIDER), University of Warwick, Coventry, UK. [6]Department of Economics, Pisa University, Pisa, Italy. [7]Department of Social and Political Sciences, Bocconi University, Milan, Italy. [8]These authors contributed equally: Emanuele Del Fava and Irene Adema. ✉email: alessia.melegaro@unibocconi.it

The heterogeneity in the observed contact patterns appears to be strongly related to the social contexts in which these surveys were conducted. These include the geographical area—with studies being carried out in sub-Saharan Africa (Ethiopia, Kenya, South Africa, Uganda, Zambia, and Zimbabwe), East Asia and Pacific (China, Fiji, Japan, Taiwan, Thailand, and Vietnam), and Latin America (Peru); the urbanization level—with conflicting evidence for the difference in contact numbers between rural and urban settings (Zambia and South Africa[6], Kenya[10], Zimbabwe[12], and China[13]); the spatial distribution—with age-assortativeness and distance from home being positively correlated (China[13]); the activity status of individuals—with the number of contacts reported in a given location being positively correlated to the time spent in that location (Zimbabwe[12]); the ethnicity—with minimal mixing found between people of different ethnic background (Fiji[15]).

These findings remark the importance of gathering further data to shed light on the differences in mixing patterns between different social contexts, especially in countries characterised by large socio-economic differentials at varying levels of urbanization. Indeed, individuals living in diverse settings along the rural–urban gradient are shown to suffer diverging morbidity and mortality paths because of the interplay of environmental and socio-economic factors[19–21].

Following the seminal contact studies in Europe within the POLYMOD project[2], the quantification of social contact patterns in different countries and social contexts has led to the development of more refined computational models to evaluate public health policies aimed at preventing, containing and mitigating the disease spread, at defining appropriate vaccination campaigns, or at designing and projecting the impact of non-pharmaceutical interventions (NPIs) as case isolation, social distancing, school closures or reduction of economic activities[22–24].

The clinical outcomes experienced at different ages of infection and the role of different age groups in shaping the transmission strongly depend on the infection considered. For instance, infections by respiratory syncytial virus (RSV) mainly affect children below 5 years of age[25], with school-age children as the most likely source of infection for infants, therefore both representing the best targets of vaccination campaigns[26,27]. Conversely, evidence from the current COVID-19 pandemic strongly suggests that older adults, along with people affected by chronic conditions, are at higher risk of experiencing severe disease and death[28–31], while children are less susceptible to the infection and young adults are considered silent spreaders as they are more likely to be asymptomatic or paucisymptomatic compared to older ages[32,33]. To add to the clinical challenges, also the contextual complexities, especially in LMICs, have to be considered, as the huge heterogeneity characterizing people mobility and access to work opportunities, and the daily routine of individuals living in different socio-economic conditions are critical determinants of the epidemiological outcomes[12,34,35].

For this purpose, we conducted a population-based survey on the southern coast of Kenya—a country whose population, in 2014, was mostly rural (74.8%), and whose urban population was mostly living in informal settlements (56%), corresponding to around 6.4 million people[36]—and gathered data on social contact patterns and individuals' daily routines from a sample of 1407 individuals. We sought to assess whether differences existed between different social contexts and how these differences affected age-specific social mixing patterns. For this reason, the study was carried out in three settings along the rural/urban gradient, namely, a rural area, an urban informal settlement, and a formal urban area. In particular, our study population consisted of (i) one rural village within the Kilifi Health and Demographic Surveillance System (KHDSS) in Kilifi county[37], and (ii) two urban areas in Mombasa county, the second largest Kenyan city, namely, an informal settlement ("slum") and an urban mixed area, characterised by a mix of slum and non-slum elements.

As part of the study, carried out through face-to-face interviews with trained fieldworkers, participants reported the number of different persons encountered in two randomly assigned consecutive days (the first randomly sampled from the whole week), the type and location of such interactions, the age of contacted individuals, and the socio-demographic characteristics of households, schools, and workplaces. From the same individuals, we also collected time-use information on daily routines, quantifying the proportion of time spent during each day at home, school, work, and in the general community.

Our study innovates in two main directions: (i) we performed a comparison of social contact and mixing patterns in three extremely different socio-demographic settings within a LMIC; (ii) we identified which groups of subjects and behaviours can play a pivotal role in infection transmission across different age groups, as a consequence of the interplay of their age mixing and daily patterns. Additionally, since our data were collected before the COVID-19 pandemic in Kenya, we furnished a valuable benchmark to assess the impact of NPIs on the COVID-19 spread and to evaluate possible vaccination strategies in the country. As the three settings considered in our study cover a wide part of the rural–urban gradient, they may be also used to evaluate control policies in those sub-Saharan countries for which contact data are still lacking.

Results

Socio-demographic characteristics of the study population.

The sample included 1407 individuals who provided data on both social contact patterns and time use. As some participants dropped out after the first survey day, we ended up with a total of 2705 person-days (p.d.) observed. The characteristics of survey's participants and their mean reported contacts are summarised in Table 1.

To assess the characteristics of widely diverse settings in terms of socio-economic status (SES), we compared the rural setting (512 individuals, 1019 p.d.), which presented the typical young age structure of a growing population, with the two urban settings, namely, the slum (345 individuals, 645 p.d.), and the combined mixed and non-slum setting (550 individuals, 1041 p.d.), which were characterised by a lower proportion of young people and a larger proportion of people in the work-age category (Supplementary Fig. S1), as expected for a population experiencing different stages of the demographic transition[38].

After adjusting for survey design with sampling weights[12], relevant differences between settings were found in terms of sex, age, household (HH) size and living arrangements, SES, and type of contact day (weekday or

Variable	Category	Participants		Number of contacts		Relative number of contacts	
		N	%	Mean	SD	IRR	95% CI
Setting	Rural	512	36.4	11.56	0.23	1	[1.00, 1.00]
	Urban slum	345	24.5	6.73	0.19	0.66	[0.61, 0.72]
	Urban mixed/non-slum	550	39.1	7.72	0.22	0.76	[0.68, 0.84]
Sex	Female	858	61	8.8	0.17	1	[1.00, 1.00]
	Male	549	39	9.1	0.25	0.97	[0.92, 1.03]
Age group	00–01 y	131	9.3	5.53	0.25	0.67	[0.59, 0.76]
	01–05 y	192	13.6	8.7	0.31	0.98	[0.87, 1.09]
	06–14 y	232	16.5	11.28	0.33	1.2	[1.10, 1.32]
	15–18 y	139	9.9	10.49	0.52	1.14	[1.03, 1.26]
	19–34 y	393	27.9	7.89	0.25	1	[1.00, 1.00]
	35–59 y	248	17.6	8.6	0.34	1.04	[0.95, 1.14]
	60+ y	72	5.1	8.13	0.41	0.99	[0.87, 1.13]
HH size	1–2p	203	14.4	6.78	0.36	1	[1.00, 1.00]
	3p	250	17.8	7.35	0.29	1.12	[0.99, 1.27]
	4p	240	17.1	8.36	0.33	1.21	[1.05, 1.39]
	5–6p	364	25.9	9.23	0.25	1.25	[1.09, 1.43]
	7+p	316	22.5	11.83	0.29	1.41	[1.22, 1.63]
	Missing	34	2.4	5.94	0.56	0.94	[0.52, 1.68]
Living arrangements	One generation	455	32.3	7.36	0.24	1	[1.00, 1.00]
	Two generations	810	57.6	9.59	0.18	0.98	[0.90, 1.07]
	Three generations	101	7.2	11.01	0.48	0.95	[0.84, 1.07]
	Missing	41	2.9	6.71	0.77	1.05	[0.61, 1.80]
SES	Low	446	31.7	10.67	0.25	1	[1.00, 1.00]
	Medium	471	33.5	8.46	0.24	1.02	[0.96, 1.09]
	High	490	34.8	7.75	0.22	0.99	[0.89, 1.11]
Education level	Primary	498	35.4	10.11	0.25	1	[1.00, 1.00]
	Secondary	274	19.5	8.15	0.29	0.98	[0.90, 1.06]
	College/university	127	9	6.93	0.44	0.87	[0.77, 0.99]
	None	483	34.3	8.6	0.22	0.99	[0.91, 1.08]
	Missing	25	1.8	5.23	0.44	0.7	[0.59, 0.83]
TU profile	Homestayers	995	70.7	8.45	0.15	1	[1.00, 1.00]
	Workers	127	9	8.96	0.41	1.38	[1.25, 1.52]
	Schoolers	104	7.4	11.56	0.58	1.23	[1.14, 1.34]
	Walkers	108	7.7	10.36	0.52	1.1	[1.02, 1.19]
	Commuters	63	4.5	9.86	0.68	1.21	[1.08, 1.36]
	Travelers	8	0.6	6.01	1.51	1.09	[0.81, 1.48]
	Missing	2	0.1	4.05	1.68	0.56	[0.30, 1.04]
Contact day type	Weekday	894	63.5	8.9	0.16	1	[1.00, 1.00]
	Weekend	513	36.5	8.94	0.23	1	[0.96, 1.04]

Table 1. Characteristics of participants and their daily contacts. The relative numbers of daily contacts (incidence rate ratio, IRR) were obtained from a GEE model with negative binomial distribution and overdispersion parameter equal to 0.14, 95% CI [0.13, 0.16].

weekend) (Supplementary Fig. S2). Specifically, people living in the rural setting were on average younger, lived in larger HHs, and had a lower socio-economic status, both in terms of SES index and current education level.

Social contact patterns by setting. A total of 23,532 contacts were reported over the two survey days. The overall mean number of contacts per person per day was 8.92 (median 8, IQR 5–12) and it was higher in the rural setting (mean 11.56) than in the two urban ones. After adjusting for the other participants' characteristics, we found that the average number of contacts was 34% and 24% lower in the slum and the mixed urban area, respectively (Table 1). Across all setting, most of contacts were non-physical, of long duration (more than 1 h) and with people outside the family, although with percentages much higher in the urban settings than in the rural one. In the rural site, we found that the percentages of physical contacts as well as contacts with relative (likely belonging to the extended family) were around 30% of the total (Supplementary Fig. S3).
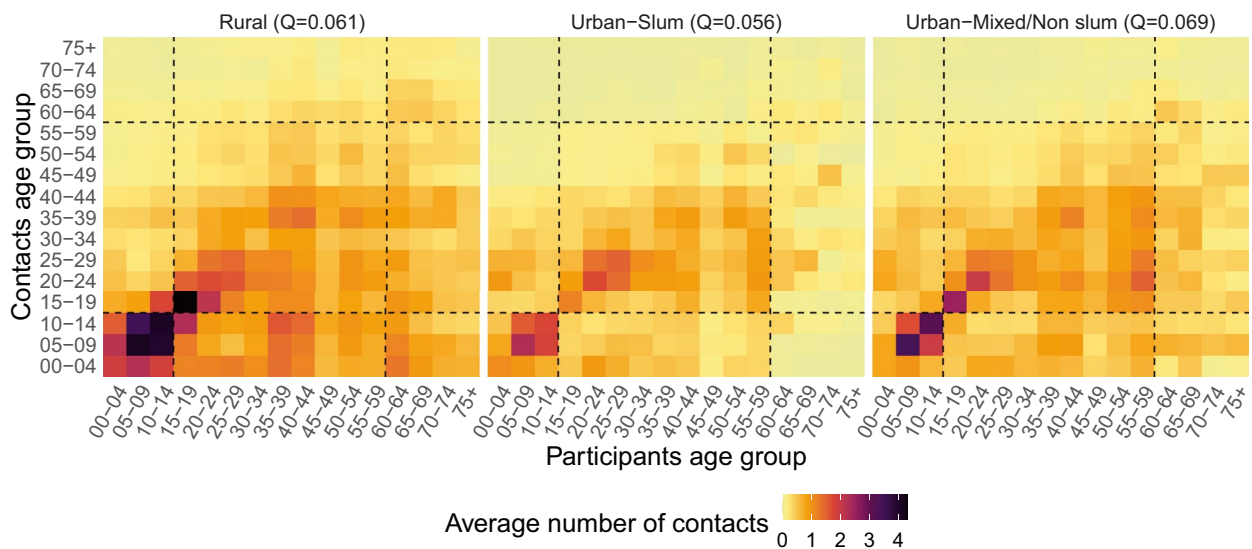
Figure 1. Average number of overall contacts between participants in age class *i* and contacted individuals in age class *j*, adjusting for reciprocity of contact at the population level, in each of the three settings. The dashed lines help identifying the contacts of children (0–14 years), teens and adults (15–59 years) and older adults (60+ years). At the top of each matrix, we report its assortativeness index Q.

For each of the three settings, we built contact matrices by 5-year age groups defining the intensity of interactions between the age groups (Fig. 1). In each setting, we found predominantly age-assortative contact patterns among younger individuals, with a larger number of contacts occurring among school-age children and adults up to 30 years. To quantify the importance of contacts occurring between individuals of the same age group in the overall contact matrix, we calculated the assortativeness index Q[39]. The closer the value of Q to zero, the lower the assortativeness and the higher the mixing between different age groups (Fig. 1). Overall, we found little difference between the three settings, as Q was equal to 0.061 for the rural setting, 0.056 for the urban slum setting, and 0.069 for the urban mixed/non-slum setting. However, a comparison with the estimates of Q for other LMICs countries (e.g., 0.051 in Zimbabwe[12], and 0.075 in Vietnam[8] and Peru[7], own calculations), and for a group of high-income countries in Europe (ranging from 0.14 in the United Kingdom[2], own calculation, to 0.21 in Belgium[40] and 0.22 in Russia[41]) clearly showed the higher degree of mixing between different age groups (and the lower assortativeness) in Kenya and in other LMICs compared to high-income countries. Although contacts with children (age group 0–14) were reported by participants of all ages, their intensity was higher among child participants. Higher contacts with children across all age groups were especially reported in the rural setting (Fig. 2C), possibly because of the higher rate of intergenerational residence and the larger HH size (Supplementary Fig. S4). A strong mixing between adults older than 20 years and younger than 60 years was also found across all settings.

Overall, the contact distribution was right-skewed, especially in the rural setting, where individuals in the last decile of the distribution reported between 18 and 42 contacts per day, while only between 12 and 21 in the slum (Fig. 2A). Regardless of the setting, most contacts were reported at home (from 59.7% in the mixed area to 67.9% in the slum) and in the general community (between 20.1% in the slum and 30% in the rural area) (Fig. 2B). Indeed, contacts at home and in the general community together represented 92.5% of all contacts in the rural setting, a share that was significantly larger than in both urban settings (87.9% in the slum and 86.6% in the mixed area), even after adjusting for participants' socio-demographic characteristics in a statistical model (sex, age group, education level, HH size, living arrangements, SES status, TU profile, and contact day type, see Supplementary Fig. S5).

Finally, across the settings, intergenerational mixing, defined as the mixing between different generations—children (aged 0–14 years), teens and adults (aged 15–59 years), and older adults (aged 60 years or more)—was predominant for older participants. Indeed, 90.3% of contacts reported by older adults were intergenerational, of which 23.3% (95% CI 20.7–26.2%) were with children. Lower rates of intergenerational mixing were reported by children (37% of their contacts, of which 1.7%, 95% CI 1.5–2.0%, with older adults) and by teens and adults (26.9% of their contacts, of which 3.2%, 95% CI 2.9–3.5%, with older adults). Moreover, likely because of the age distribution of the different settings, we found that the percentage of intergenerational contact reported with both older adults and children was higher in the rural setting, while the percentage of intergenerational contact reported with teens and adults was higher in the urban settings (Fig. 2C). These results were robust with respect to whether the day of the reported contacts was a weekday or a weekend day, as we did not find any relevant difference in the percentages of contacts between the two types of day (Supplementary Fig. S6).

Differences in participants' daily routines. To understand the behavioural determinants of the number of social contacts, we identified six profiles based on individuals' daily routine behaviour, which were subsequently used to investigate differences in contact patterns between different groups of individuals. To this
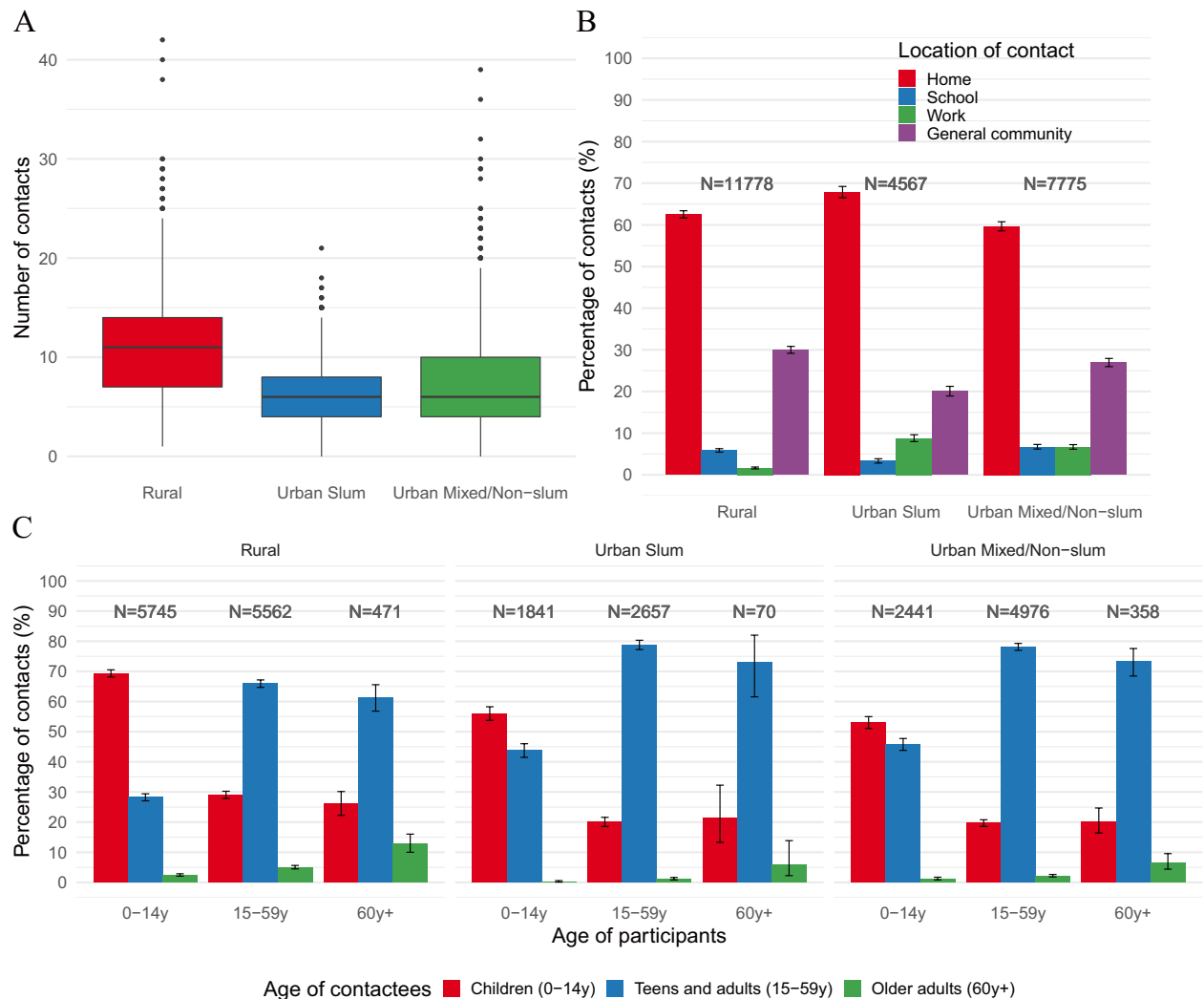
Figure 2. Distribution of daily social contacts by setting (**A**), percentage (with 95% CI) of daily social contacts by location of contact (**B**), and percentage (with 95% CI) of daily social contacts by setting and age of participants with children (0–14 years), teens and adults (15–59 years) and older adults (60+ years) (**C**). For each bar, the respective denominator (N) is reported. All CIs are computed as Wilson CIs for a binomial proportion[42].

purpose, we leveraged on the combination of a principal component analysis (PCA)—which we used to reduce the data dimensionality while retaining most of the information (transferred into newly created independent variables called "principal components")—and a hierarchical cluster analysis (HCA)—used to create groups of individuals with similar characteristics based on the principal components[43]. In our analysis, we first applied PCA to the time use (TU) data—a set of indicators for the time spent in different locations/activities within specific time slots during the day—and retrieved 11 principal components that summarized 79% of the total information. Next, we applied HCA to the principal components to group individuals into meaningful profiles that can be described based on the most frequently visited location during the day, in terms of time slots, by individuals included therein (Fig. 3).

Most individuals (92.3%) reported the same TU profile over the two survey days, especially in the rural site (99%). The largest group by far (*Homestayers*) accounted for 71.1% of p.d. and identified people *staying at home or within the village* the whole day (respectively 90.8% and 7% of their daily time). The remaining TU profiles were characterised by a specific location, other than the home. The second largest group (*Workers*: 9.3% of total p.d.) identified those spending over half of their day *at work* (56%); this profile was more prevalent in the urban settings, especially the slum, and only for the group of adults. The third profile group (*Schoolers*: 7.8% of p.d.) gathered those spending a large share of the day *at school* (48.6%), which was uniformly present in all settings, and mostly populated by children.

The last three groups encompassed individuals spending a relatively larger share of their time in the general community, and were gradually more prevalent in the urban settings and among teens and adults as the distance from home increased. The fourth group (*Walkers*: 6.6% of total p.d.) included those spending, outside the home, 37.5% of their time in locations, likely at walking distance, *within the study site* (i.e., within the rural site of
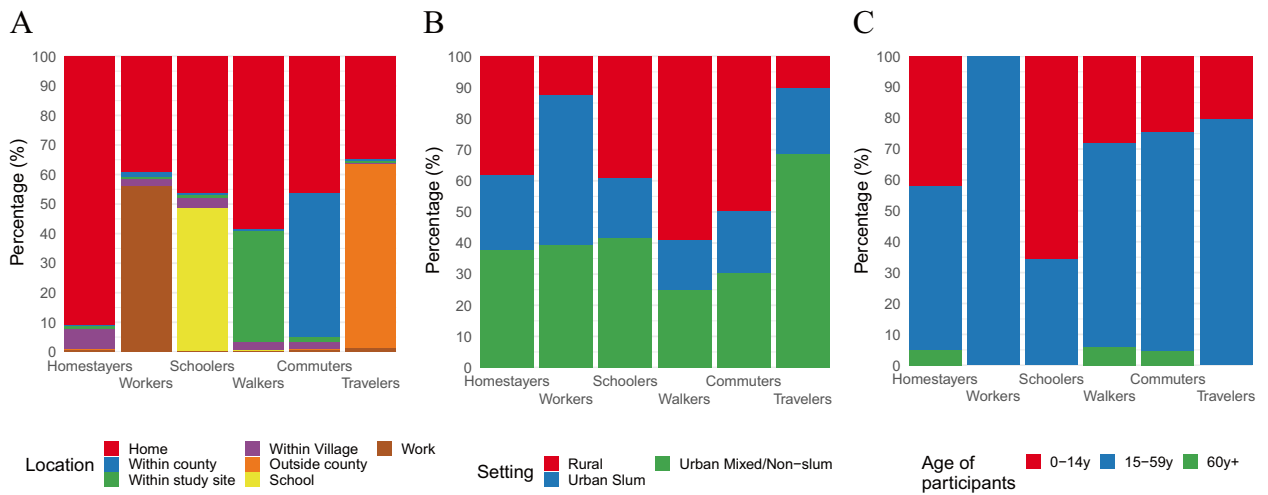
Figure 3. Description of the six TU profile groups by location (**A**), setting (**B**), and age group (**C**).

Banda-ra Salama or the urban site of Tudor). The fifth group (*Commuters*: 4.6% of the total p.d.) encompassed individuals spending, outside the home, 49.1% of their time in locations, possibly reached using transportation, *within the county* (i.e., in Kilifi County outside of Banda-ra Salama, or in Mombasa County outside of Tudor); finally, the last group (*Travelers*: 0.6% of the total p.d.) identified individuals spending a relevant part of their day *outside the county* (62.2%).

Social contact patterns by TU profile.   The TU profiles described above were used to assess differences in contact patterns between people characterised by different daily routines. No matter the TU profile, most of contacts were non-physical, of long duration (especially those of Homestayers, Schoolers, and Travelers), and with people outside the family (especially those of Workers and Schoolers, but also of Walkers and Commuters) (Supplementary Fig. S3).

With regards to age mixing, we found that the degree of age-assortativeness varied across TU profiles (Fig. 4). This was lower for Workers ($Q = 0.0030$), Travelers ($Q = 0.023$), and Commuters ($Q = 0.046$), for which age mixing, apart from contacts between parents and their children, was scattered among teens and adults of different age. The Q index increased to 0.057 for Homestayers and 0.066 for Walkers, reflecting a moderate level of age-assortative mixing (therefore, suggesting marked intergenerational mixing) among children and adults up to the age of 30 years. Finally, the highest level of age-assortativeness (and thus the lowest level of intergenerational mixing) was found among Schoolers ($Q = 0.10$), where the intensity of contact among children reached the peak.

The contact distribution under each TU profile was generally right-skewed, especially for Schoolers, who also reported a higher median number of contacts (Fig. 5A). Workers reported 43.3% of their contacts at work, Schoolers 48.5% of contacts at school, while Walkers, Commuters and Travelers reported between 55% and 64.4% of contacts in the general community, showing a positive relationship between the time spent and the reported number of contacts in each location (Fig. 5B). Contacts at home usually represented the largest share, except for Commuters and Travelers, who spent a larger share of time in the general community at increasing distance from home (Figs. 3A, 5B) and reporting a higher number of contacts as occurring in this location.

TU profiles also varied in terms of intergenerational mixing. Contacts with children were mainly reported by Schoolers (55.4%), followed by Homestayers (43.2%) and Walkers (34%) (Fig. 5C). Within these three profiles, contact with children was mainly reported by child participants; however, we also observed a high share of intergenerational contact reported by older adults with a Homestayer profile (26%). Conversely, contacts of Workers with children remained below 16%. Contacts with older adults represented a small fraction for all TU profiles, although we found significant differences between Workers and Schoolers (less than 2% of contacts), Homestayers (almost 3% of contacts), and Walkers and Commuters (almost 5% of contacts). Such contacts with older adults were mainly reported by participants aged 60 years or more; however, we also observed a substantial level of intergenerational contact reported by children with a Walker profile (3.2% of their reported contacts) or a Commuter profile (3.5%) and by teens and adults with the same profiles (4% of contacts for Walkers and 5.1% for Commuters), suggesting that interactions with the older adults occurred mostly with those individuals spending a high proportion of their time in the general community (Fig. 5C).

After adjusting for participants' socio-demographic characteristics in a statistical model, we estimated that Workers and Schoolers reported overall 38% and 23% more contacts than Homestayers, respectively (Table 1). We did not find any significant difference in contact numbers between Workers and Homestayers at home and in the general community (Supplementary Fig. S7). Conversely, for Schoolers we estimated a lower relative number of contacts than Homestayers both at home and in the general community, supporting the finding that most of their contacts occurred at school (Supplementary Fig. S7). Finally, Walkers and Commuters (for whom we estimated 10% and 21% more contacts than Homestayers, respectively) had less contacts at home and much more in the general community.
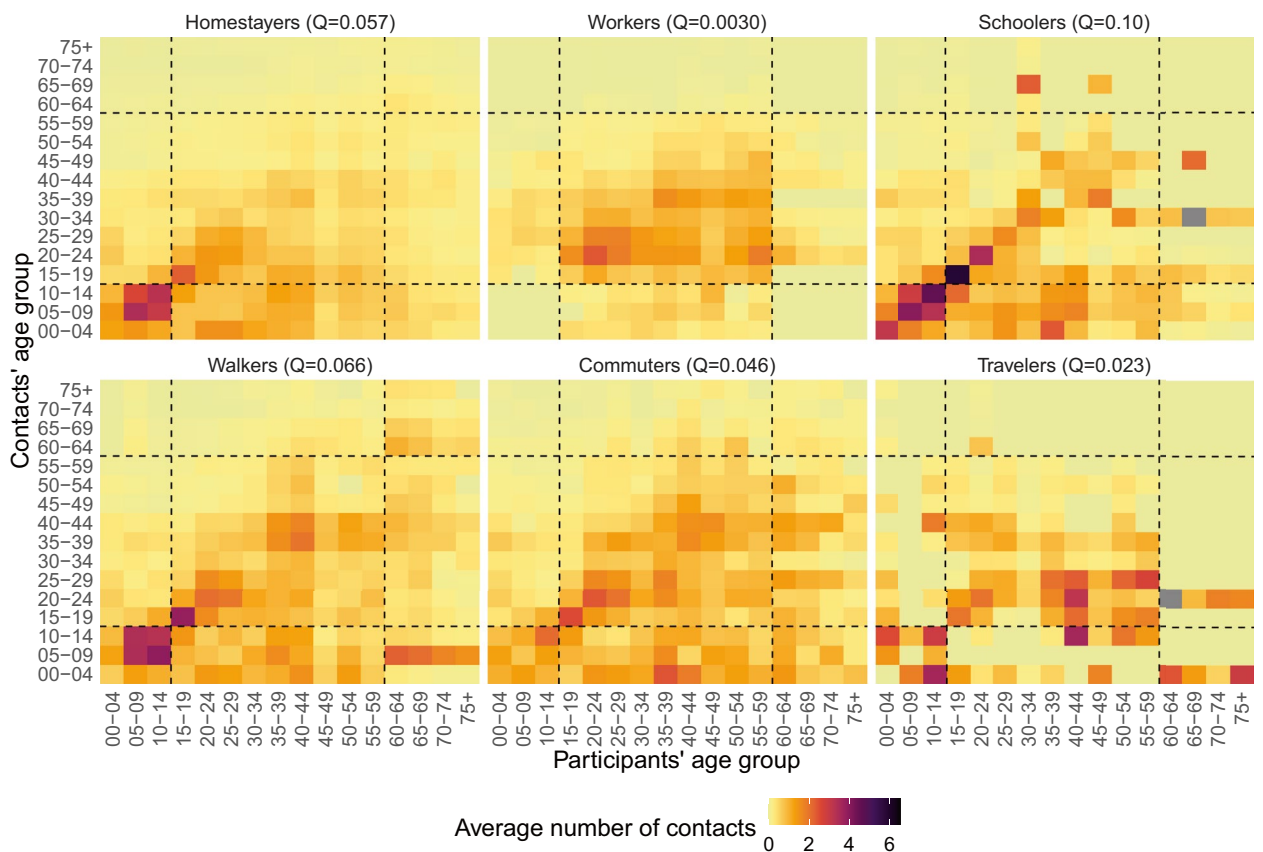
Figure 4. Average number of overall contacts between participants in age class *i* and contacted individuals in age class *j*, adjusting for reciprocity of contact at the population level, stratified by TU profile. The dashed lines help identifying the contacts of children (0–14 years), teens and adults (15–59 years) and older adults (60 + years). At the top of each matrix, we report its assortativeness index Q.

Discussion

Although most social contact studies in LMICs[7,8,11] have been predominantly conducted in rural settings, while those conducted in high-income countries[2,17,18,41,44,45] have been largely carried out in urban settings, only a few have so far presented a direct comparison of social mixing patterns between such different levels of urbanization within the same country or region[6,10,12,13]. Moreover, only one study sought to investigate whether differences in social behaviour induced by daily routines entailed differences in the average numbers of contacts[12]. Our contact study in Kenya contributes to filling this gap by (i) providing a robust comparison of social mixing patterns across diverse demographic settings along the rural–urban gradient, and (ii) identifying which individuals' profiles, based on their age mixing patterns and daily time use behaviour, can play a pivotal role in bringing different age groups in contact. The latter aspect is of great importance to identify those individuals who present a higher risk of bringing an infection to the more vulnerable segments of the population.

Even though this study was not the first one conducted in Kenya[10,46,47], it was the largest and most diverse one in terms of surveyed settings, describing three geographical areas that significantly differed in terms of socio-economic status (the rural setting being the poorer in relative terms, the urban mixed/non-slum area being the wealthier), age and household size distribution (a younger population and a higher frequency of large extended households in the rural setting, and a higher frequency of small nuclear households in the slum), and TU profiles (a higher percentage of people spending time at work in the urban settings, especially in the slum, as well as a higher percentage of people spending time in the general community in the rural setting).

The number of contacts reported by our study participants were higher in the rural setting than in the urban ones, even after adjusting for the HH size (higher in rural site) and the living arrangements (with higher intergenerational cohabitation in the rural site), as a result of the higher contact numbers reported at home and in the general community. Even though these findings differed from what was found for Zimbabwe[12], they were consistent with what was observed by previous studies in the same area in Kenya[10], in Zambia and South Africa[6], as well as with the evidence from the synthetic matrices for low-income countries constructed using setting-specific survey data (e.g., on household, school, classroom, and workplace composition, and on contact patterns)[48]. Despite a lower degree of age assortativeness in our three settings than what observed in high-income countries, we found both a higher intensity of contacts among children and a higher share of intergenerational contact with older adults in the rural setting than in the urban ones. As we found no evidence of a difference for these contact patterns regardless of being reported on weekdays versus weekend, we believe they were not entirely motivated by school attendance or the working routine during the weekdays, but rather a consequence
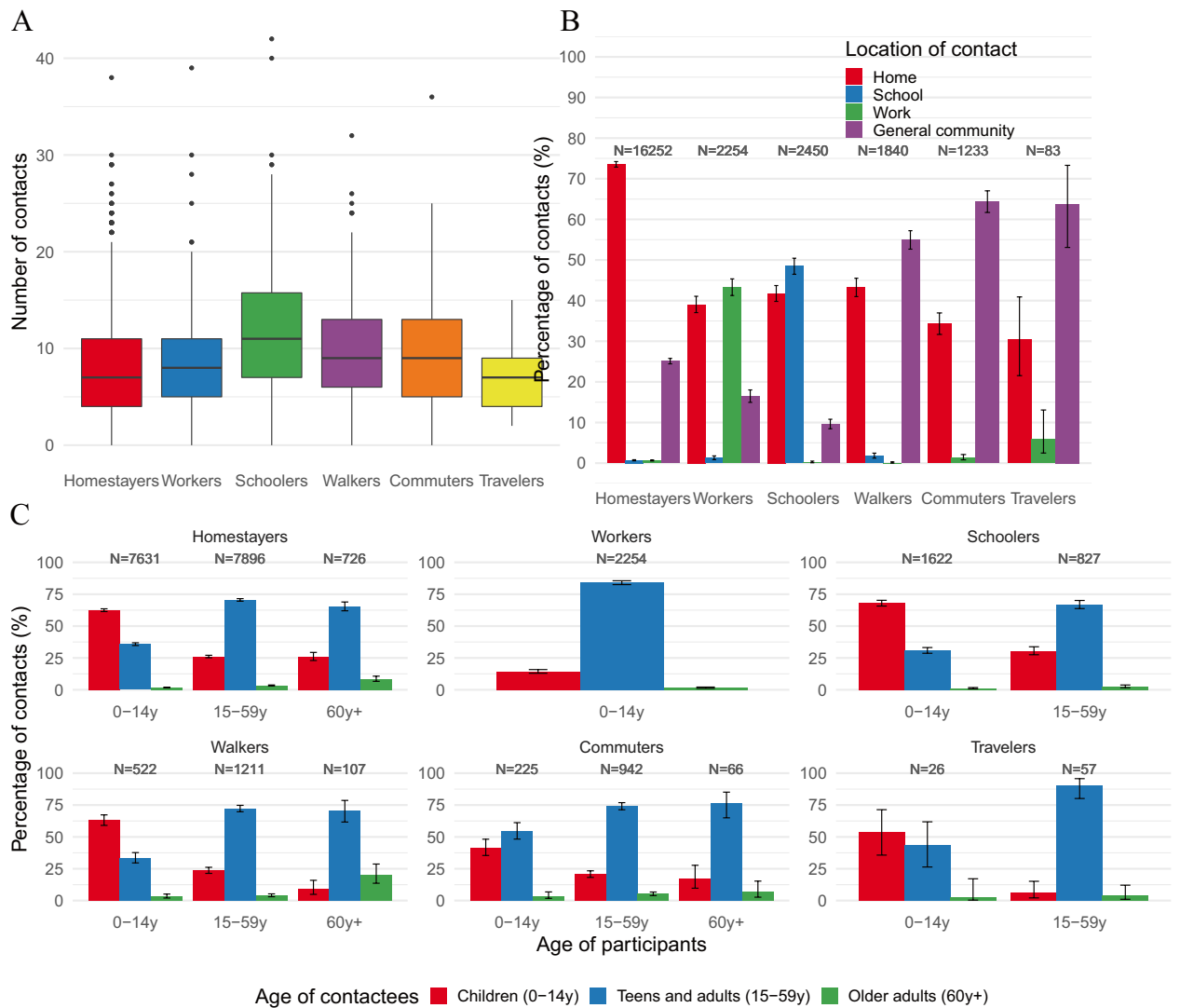
Figure 5. Distribution of daily social contacts by TU profile (**A**), percentage (with 95% CI) of daily social contacts by location of contact (**B**), and percentage (with 95% CI) of daily social contacts by TU profile group and age of participants with children (0–14 years), teens and adults (15–59 years) and older adults (60+ years) (**C**). For each bar, the respective denominator (N) is reported. All CIs are computed as Wilson CIs for a binomial proportion[42].

of the HH structures and living arrangements more common in the rural site. Such mixing behaviours should be taken into account when considering the transmission of infections with strongly age-specific severity, such as in the case of RSV or SARS-CoV-2[26,27,31,33].

Compared with previous studies conducted in Kenya, we found that our survey participants reported a lower number of contacts: a study carried out between 2011 and 2012 in Kilifi county reported 18.8 physical contacts per day in a rural setting and 16.5 in a semi urban area[10], while a small study conducted in a slum in the capital Nairobi in May 2020 found an average of 18 overall contacts per day[47]. Such discrepancies with our study might be due to differences in the sample selection, especially in the urban site. First, since we lacked a sampling list for our urban site, conversely to what happened with the study in the Nairobi[47] slums, we employed a modified version of the WHO Expanded Program on Immunization (EPI) cluster sampling technique[49] (see Supplementary Information), which may have caused some bias in the household selection. Second, the inaccessibility to some homes, especially the guarded houses in the non-slum area, may have been another possible source of bias in the household selection. Finally, the requirement that interviewers worked between 7 a.m. and 6 p.m., due to both security measures and working time restrictions, may have prevented the inclusion of working adults who might be outside the home when interviews were conducted. On the other hand, to improve data quality, especially that collected from illiterate respondents, instead of using self-reporting paper diaries[10] or phone interviews[47], we used face-to-face interviews conducted by trained interviewers, during which the data provided by respondents were thoroughly revised and entered into an electronic questionnaire.

The combined collection of time-use and contact data within the same study allowed us to investigate how people differed in terms of daily behaviour and how their behaviour was associated with a distinct age mixing

and contact numbers. At least 70% of participants spent most of their day at home or in the village, without any time at work or at school, and reported the lowest number of contacts (8.45 on average), 74% of which on the premises of their home. Intergenerational mixing patterns were found to be less relevant for Workers and Schoolers and more substantial for Homestayers, Walkers and Commuters. On the one hand, the level of intergenerational mixing increased for people spending more time farther from home, suggesting that people who spend most of their time and report most of their contacts in the general community might be relevant for the transmission of COVID-19 as they are characterised by a larger number of interactions between teens and adults and with older adults. These behaviours, as well as those of Workers and Schoolers who concentrate their contacts among people of the same generation, may be those mostly affected by social distancing and isolation, which are strategies designed to reduce the number of contacts with the most vulnerable age groups. On the other hand, the rather large mixing reported by older adults with children at home may be more difficult to control, as people who spend most of their time at home are less likely affected by public restrictions targeting individual movement, social activities, and school closures[50].

Our results highlight that the higher number of reported contacts and the higher level of intergenerational mixing, especially between children and older adults in the rural site, as opposed to the urban slum, and mainly reported by individuals who spend most of their day at home or in the general community, may have important consequences for COVID-19 disease transmission to older adults. Preliminary results on the spread of the COVID-19 pandemic in Kenya showed that the first wave (April–August 2020) was particularly concentrated within those individuals of lower SES living in the urban areas, such as individuals working in the informal sector, likely less able to restrict movement, while the successive wave (October–December 2020) also spread among the higher SES in the urban areas, as well as in the rural areas[51]. However, at the moment of writing (summer 2021), the little available information on the age distribution of the COVID-19 cases, mostly due to the poor case reporting, the demand-driven testing, and a focus on younger people[52], does not allow us to properly assess the epidemiological impact of the infection on the different generations, especially the older adults. Nonetheless, we believe that, considering the heterogeneity in contact numbers and social mixing entailed by differences in time use behaviour, rather than simply looking at the variations in contact patterns per location, may achieve a better identification of groups of individuals at higher risk of transmission, define effective measures to prevent the spread of the diseases, and design appropriate targets of future vaccination efforts.

Methods

Study population. For the rural site, we selected the administrative location of Banda-ra Salama (Kilifi County), one of the rural settlements within the Kilifi Health and Demographic Surveillance System (KHDSS), an area including most of patients admitted to Kilifi District Hospital. The HDSS had an average population of 261,919 between 2006 and 2010, a population growth rate of 2.79% per year, and a total fertility rate (TFR) of 4.7[37].

For the urban site, we chose Tudor, a subdivision of the city of Mombasa (Mombasa County), which included both an urban informal settlement and a formal urban area. Mombasa had a population of 939,370 in 2009 and in 2014 reported a population growth rate of 3.60% per year[53] and a TFR of 3.2[54]. The county was also characterised by disparities in living conditions within the city and by the presence of informal settlements (slums), which featured higher fertility and mortality compared to the rest of the country[53].

Study design. The study design was cross-sectional, targeting recruitment at the household level, and was carried out through a series of face-to-face interviews with trained fieldworkers. Individuals of all ages were grouped into seven age classes reflective of key social or behavioural groups: < 1 year (infants), 1–5 years (preschoolers), 6–14 years (primary school), 15–18 years (secondary school), 19–34 years (younger adults), 35–59 years (middle-aged adults), and > 60 years (older adults).

Estimates from a previous contact study conducted in Kenya were used to parameterize the sample size calculation[10]. To account for the larger heterogeneity in the social structure of the urban area in terms of household composition and social interactions, 1000 participants were targeted in the urban area and 500 from the rural area. Although the sample size was allocated to the seven age groups proportional to the population of the corresponding age strata, children were oversampled due to the critical role they were expected to play in infectious disease transmission and hence the need to reduce the level of uncertainty in the estimated contact rate for this age group. Recruitment was staggered over a six-month period (June 2015 to December 2015). Inclusion criteria included (i) giving informed consent, either directly or by parents if participants were minor of age, and (ii) planning to remain in the site for at least two weeks. Information sheets and consent forms were provided to all participants.

Each survey respondent was assigned two randomly picked consecutive days and requested to keep track of social contacts and visited locations during these 2 days by keeping a mental note, for those with fewer contacts, or using a tracking sheet on which they could write a few details such as name initials, presumed age, time and location where they met the contacts, colour of clothes or any other details that would to help them recall. During the interviews with the fieldworkers, background information as well as data on social contacts and time use were collected, revised, and entered into an electronic questionnaire.

Background information on the participants regarded their age and sex, their education career and occupational status, their HH size and characteristics (such as owned assets, use to derive a SES measure), and their school and work environments (e.g., distance from home, school class and workplace size).

Similarly to previous studies[2,12], contacts were defined as an interaction between two individuals, either physical (involving skin-to-skin contact), or non-physical (involving a two-way conversation with three or more words in the physical presence of another person, but no skin-to-skin contact). For each contacted individual,

participants provided information on the sex and the age of the contacted person, the type of contact, the location(s) where the contact occurred (home, school, work, general community), and the relationship with the contacted person (e.g., sibling, parent, non-relative). To distinguish short-lived contacts with long-duration contacts, and to account for the possibility that an individual may be encountered several times during the day, we collected the information on the total duration of the contact with such an individual.

Finally, participants provided information on their time-use by recording all the visited locations (the same used for contacts) during their day. For this purpose, each survey day was divided into nine time slots reflecting the position of the sun, spanning from a minimum of two hours (dawn, early morning, late morning, mid-day/lunch time, early afternoon, late afternoon, evening, and dusk) to a maximum of eight hours (late night). Respondents reported whether, for each time slot, they had visited or not each of seven given locations (home, school, work, and the general community within the village/estate, within the study site, within the county, or outside the county), which resulted in a matrix of 63 dummy variables for each combination of time slot and location.

The study was approved by the Kenya Medical Research Institute-Scientific Ethics Review Unit (KEMRI-SERU) and the University of Warwick Biomedical and Scientific Research Ethics Committee (BSREC). The parent study, DECIDE, had received ethical approval by the ERC Ethics Committee and the Ethics Committee of the Italian National Institute of Health. Individual consent was obtained from each of the study participants. All research was performed in accordance with relevant guidelines and regulations.

Statistical analysis. We analysed all collected data by adjusting for sampling weights (the inverse of the probability that an observation is included because of the sampling design), which were calculated for each site separately to compute site-specific estimates. These weights were based on both the KHDSS and the Mombasa County population age distribution, taken as reference populations.

We created two new variables for inclusion in the statistical models: a continuous SES index and a TU profile variable. The SES index was computed by applying principal component analysis to data on household characteristics and owned assets, also collected in the study, and taking the first principal component as the index[55]. To construct the TU profiles, we applied principal component analysis to the collected TU data—a set of 54 dummy variables (nine time slots for six locations, as the location "general community within the village/estate" was excluded and kept as baseline) to summarize all the information contained in the variables with a smaller number of dimensions, constructed to be independent of each other, while retaining as much information as possible. We used the obtained components as input for a hierarchical cluster analysis to create groups of person-days characterised by similar time use behaviour profiles[43].

Social contact matrices by study site and TU profile were constructed to show the average number of contacts between 5-year age classes in the respective populations (plus a single age class from 75 years or more). Each matrix element, $m_{ij}^{(rec.)}$, contained the mean number of contacts per day in age class j as reported by participants in age class i. To take into account the reciprocal nature of contacts[1], the elements of the matrix were adjusted for reciprocity at the population level (see the Supplementary Information for further details on the construction of the contact matrices).

We measured the age-assortativeness with the Q index[41,56,57], which is calculated as $Q = [Tr(P) - 1]/(n-1)$, where P is the matrix whose elements represent the fraction of total contacts of age group i with age group j, $P_{ij} = M_{ij}/\sum_j M_{ij}$, M_{ij} is the matrix with the number of total contacts between age groups, and $Tr(\bullet)$ is the trace of the matrix, i.e., the sum of its diagonal elements. The Q index takes values one as the assortativeness becomes maximal (all contacts are on the main diagonal, i.e., with individuals of the same age), while, for increasing homogenous (or proportionate) mixing, it tends to zero[39].

Generalised estimating equations (GEE) were chosen to evaluate the effect of the different participants' characteristics on the daily number of contacts[58]. The negative binomial distribution was preferred over the Poisson to allow for a higher variance in the number of contacts reported by participants than what expected under the former distribution (overdispersion)[2,12,59]. We tested, however, for the possibility of reducing the distribution to a Poisson by checking whether the 95% confidence interval (CI) of the estimated overdispersion parameter contained zero. Contact data were clustered within participants over the survey days under the assumption that responses from the same individual were correlated, while responses from different participants were assumed to be independent.

We fitted the GEE model to all contacts and then separately for each of the three settings. Except for the setting, which was included only in the overall model, the variables included in all models were the sex, the age group (0–14 years, 15–59 years, and 60 years or more), the HH size (approximately grouped in five quantile groups, i.e., 1–3, 4, 5, 6–7, and 7 or more individuals), the living arrangements (one-generation, two-generation, or three-generations HHs), the SES index, the current education level, the TU profile, and whether the contact day was a weekday or a weekend day. We interpreted the exponentiated model estimates from the negative binomial GEE as incidence rate ratios (IRR), giving the relative number of contacts per day with respect to the reference category of each variable, and we accompanied them with 95% CI.

Data cleaning and wrangling, matrix construction, and result visualization were carried out in R (using, in particular, the package "FactoMineR" for the PCA and the HCA[60]); GEE models were estimated in Stata, using the procedure "xtgee".

Data availability
The datasets and the source codes necessary for the reproducibility of our results are available at the following repository: https://osf.io/ebwrh/?view_only=19d34326583e47a19510668c6e1c17e5.

References

1. 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.* **164**, 936–944 (2006).
2. Mossong, J. *et al.* Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med.* **5**, e74 (2008).
3. Diekmann, O., Heesterbeek, J. A. P. & Metz, J. A. J. On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. *J. Math. Biol.* **28**, 365–382 (1990).
4. Zhang, J. *et al.* The impact of relaxing interventions on human contact patterns and SARS-CoV-2 transmission in China. *Sci. Adv.* **7**, eabe2584 (2021).
5. Marziano, V. *et al.* Retrospective analysis of the Italian exit strategy from COVID-19 lockdown. *Proc. Natl. Acad. Sci.* **118**, e2019617118 (2021).
6. Dodd, P. J. *et al.* Age- and sex-specific social contact patterns and incidence of *Mycobacterium tuberculosis* infection. *Am. J. Epidemiol.* https://doi.org/10.1093/aje/kwv160 (2015).
7. 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* **10**, e0118457 (2015).
8. Horby, P. *et al.* Social contact patterns in Vietnam and implications for the control of infectious diseases. *PLoS One* **6**, e16965 (2011).
9. Johnstone-Robertson, S. P. *et al.* Social mixing patterns within a South African township community: implications for respiratory disease transmission and control. *Am. J. Epidemiol.* **174**, 1246–1255 (2011).
10. Kiti, M. C. *et al.* Quantifying age-related rates of social contact using diaries in a rural coastal population of Kenya. *PLoS One* **9**, e104786 (2014).
11. le Polain de Waroux, O. *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.* **18**, 1–2 (2018).
12. Melegaro, A. *et al.* Social contact structures and time use patterns in the Manicaland Province of Zimbabwe. *PLoS One* **12**, e0170459 (2017).
13. Read, J. M. *et al.* Social mixing patterns in rural and urban areas of southern China. *Proc. R. Soc. B Biol. Sci.* **281**, 20140268 (2014).
14. Stein, M. L. *et al.* Comparison of contact patterns relevant for transmission of respiratory pathogens in Thailand and the Netherlands using respondent-driven sampling. *PLoS One* **9**, e113711 (2014).
15. Watson, C. H. *et al.* Social mixing in Fiji: Who-eats-with-whom contact patterns and the implications of age and ethnic heterogeneity for disease dynamics in the Pacific Islands. *PLoS One* **12**, e0186911 (2017).
16. Huang, Y. *et al.* Spatiotemporal heterogeneity of social contact patterns related to infectious diseases in the Guangdong Province, China. *Sci. Rep.* **10**, 6119 (2020).
17. Munasinghe, L., Asai, Y. & Nishiura, H. Quantifying heterogeneous contact patterns in Japan: A social contact survey. *Theor. Biol. Med. Model.* **16**, 1–10 (2019).
18. Fu, Y., Wang, D.-W. & Chuang, J.-H. Representative contact diaries for modeling the spread of infectious diseases in Taiwan. *PLoS One* **7**, e45113 (2012).
19. Trentini, F. *et al.* Modeling the interplay between demography, social contact patterns, and SARS-CoV-2 transmission in the South West Shewa Zone of Oromia Region, Ethiopia. *BMC Med.* **19**, 89 (2021).
20. Matz, C. J., Stieb, D. M. & Brion, O. Urban-rural differences in daily time-activity patterns, occupational activity and housing characteristics. *Environ. Health* **14**, 88 (2015).
21. Poletti, P. *et al.* The hidden burden of measles in Ethiopia: How distance to hospital shapes the disease mortality rate. *BMC Med.* **16**, 177 (2018).
22. Funk, S. *et al.* Combining serological and contact data to derive target immunity levels for achieving and maintaining measles elimination. *BMC Med.* **17**, 180 (2019).
23. Litvinova, M., Liu, Q.-H., Kulikov, E. S. & Ajelli, M. Reactive school closure weakens the network of social interactions and reduces the spread of influenza. *Proc. Natl. Acad. Sci.* **116**, 13174–13181 (2019).
24. Zhang, J. *et al.* Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. *Science* https://doi.org/10.1126/science.abb8001 (2020).
25. Shi, T. *et al.* Global, regional, and national disease burden estimates of acute lower respiratory infections due to respiratory syncytial virus in young children in 2015: A systematic review and modelling study. *Lancet* **390**, 946–958 (2017).
26. Poletti, P. *et al.* Evaluating vaccination strategies for reducing infant respiratory syncytial virus infection in low-income settings. *BMC Med.* **13**, 1–11 (2015).
27. Yamin, D. *et al.* Vaccination strategies against respiratory syncytial virus. *Proc. Natl. Acad. Sci.* **113**, 13239–13244 (2016).
28. Dowd, J. B. *et al.* Demographic science aids in understanding the spread and fatality rates of COVID-19. *Proc. Natl. Acad. Sci.* **117**, 9696–9698 (2020).
29. Nepomuceno, M. R. *et al.* Besides population age structure, health and other demographic factors can contribute to understanding the COVID-19 burden. *Proc. Natl. Acad. Sci.* **117**, 13881–13883 (2020).
30. O'Driscoll, M. *et al.* Age-specific mortality and immunity patterns of SARS-CoV-2. *Nature* https://doi.org/10.1038/s41586-020-2918-0 (2020).
31. Poletti, P. *et al.* Probability of symptoms and critical disease after SARS-CoV-2 infection. *ArXiv200608471 Q-Bio* (2020).
32. Moghadas, S. M. *et al.* The implications of silent transmission for the control of COVID-19 outbreaks. *Proc. Natl. Acad. Sci.* **117**, 17513–17515 (2020).
33. Davies, N. G. *et al.* Age-dependent effects in the transmission and control of COVID-19 epidemics. *Nat. Med.* https://doi.org/10.1038/s41591-020-0962-9 (2020).
34. Walker, P. G. T. *et al.* The impact of COVID-19 and strategies for mitigation and suppression in low- and middle-income countries. *Science* **369**, 413–422 (2020).
35. Trentini, F. *et al.* Demography, social contact patterns and the COVID-19 burden in different settings of Ethiopia: A modeling study. *medRxiv* https://doi.org/10.1101/2020.11.24.20237560 (2020).
36. UN-Habitat. *Slum Almanac 2015/2016: Tackling Improvement in the Lives of Slum Dwellers.* (UN-Habitat, 2016).
37. Scott, J. A. G. *et al.* Profile: The Kilifi health and demographic surveillance system (KHDSS). *Int. J. Epidemiol.* **41**, 650–657 (2012).
38. Williams, J. R., Manfredi, P. & Melegaro, A. The potential impact of the demographic transition in the Senegal-Gambia region of sub-Saharan Africa on the burden of infectious disease and its potential synergies with control programmes: the case of hepatitis B. *BMC Med.* **16**, 1–13 (2018).
39. Gupta, S., Anderson, R. M. & May, R. M. Networks of sexual contacts: Implications for the pattern of spread of HIV. *AIDS* **3**, 807–818 (1989).
40. Hoang, T. V. *et al.* Close contact infection dynamics over time: Insights from a second large-scale social contact survey in Flanders, Belgium, in 2010–2011. *BMC Infect. Dis.* **21**, 274 (2021).

41. Ajelli, M. & Litvinova, M. Estimating contact patterns relevant to the spread of infectious diseases in Russia. *J. Theor. Biol.* **419**, 1–7 (2017).
42. Brown, L. D., Cai, T. T. & DasGupta, A. Interval estimation for a binomial proportion. *Stat. Sci.* **16**, 101–133 (2001).
43. Husson, F., Josse, J. & Pages, J. Principal component methods-hierarchical clustering-partitional clustering: Why would we need to choose for visualizing data. *Appl. Math. Dep.* 1–17 (2010).
44. Béraud, G. *et al.* The French Connection: The first large population-based contact survey in France relevant for the spread of infectious diseases. *PLoS One* **10**, e0133203 (2015).
45. Leung, K., Jit, M., Lau, E. H. Y. & Wu, J. T. Social contact patterns relevant to the spread of respiratory infectious diseases in Hong Kong. *Sci. Rep.* **7**, 1–12 (2017).
46. Kiti, M. C. *et al.* Quantifying social contacts in a household setting of rural Kenya using wearable proximity sensors. *EPJ Data Sci.* **5**, 1–21 (2016).
47. Quaife, M. *et al.* The impact of COVID-19 control measures on social contacts and transmission in Kenyan informal settlements. *BMC Med.* **18**, 316 (2020).
48. Prem, K., Cook, A. R. & Jit, M. Projecting social contact matrices in 152 countries using contact surveys and demographic data. *PLoS Comput. Biol.* **13**, e1005697 (2017).
49. World Health Organization. *Immunization Coverage Cluster Survey: Reference Manual.* https://apps.who.int/iris/handle/10665/69087. Accessed 26 Oct 2021 (2005).
50. Wagner, M. *et al.* Using contact data to model the impact of contact tracing and physical distancing to control the SARS-CoV-2 outbreak in Kenya. *Wellcome Open Res.* **5**, 212 (2020).
51. Brand, S. P. C. *et al.* COVID-19 transmission dynamics underlying epidemic waves in Kenya. *medRxiv* https://doi.org/10.1101/2021.06.17.21259100 (2021).
52. Uyoga, S. *et al.* Seroprevalence of anti–SARS-CoV-2 IgG antibodies in Kenyan blood donors. *Science* **371**, 79–82 (2021).
53. Kenya National Bureau of Statistics. *Mombasa Informal Settlement Survey, Kenya, 2009.* Nairobi: Kenya National Bureau of Statistics (2010).
54. Kenya National Bureau of Statistics, Ministry of Health/Kenya, National AIDS Control Council/Kenya, Kenya Medical Research Institute, National Council for Population and Development/Kenya, and ICF International. *Kenya Demographic and Health Survey 2014.* Rockville, MD, USA: Kenya National Bureau of Statistics, Ministry of Health/Kenya, National AIDS Control Council/Kenya, Kenya Medical Research Institute, National Council for Population and Development/Kenya, and ICF International (2015).
55. Vyas, S. & Kumaranayake, L. Constructing socio-economic status indices: How to use principal components analysis. *Health Policy Plan.* **21**, 459–468 (2006).
56. Fumanelli, L., Ajelli, M., Manfredi, P., Vespignani, A. & Merler, S. Inferring the structure of social contacts from demographic data in the analysis of infectious diseases spread. *PLoS Comput. Biol.* **8**, e1002673 (2012).
57. Iozzi, F. *et al.* Little Italy: An agent-based approach to the estimation of contact patterns—fitting predicted matrices to serological data. *PLoS Comput. Biol.* **6**, e1001021 (2010).
58. Liang, K.-Y. & Zeger, S. L. Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22 (1986).
59. Hens, N. *et al.* Mining social mixing patterns for infectious disease models based on a two-day population survey in Belgium. *BMC Infect. Dis.* **9**, 1–18 (2009).
60. Lê, S., Josse, J. & Husson, F. FactoMineR: An R package for multivariate analysis. *J. Stat. Softw.* **25**, 1–18 (2008).

Acknowledgements

Author contributions

Conceptualization: E.D.F., P.P., P.M., A.M.; methodology: E.D.F., P.P.; software: E.D.F.; formal analysis: E.D.F.; investigation: E.D.F., I.A., M.C.K., P.P.; resources: I.A.; data curation: I.A., M.C.K.; writing—original draft: E.D.F., I.A.; writing—review & editing: E.D.F., I.A., M.C.K., P.P., S.M., D.J.N., P.M., A.M.; visualization: E.D.F.; supervision: S.M., D.J.N., P.M., A.M.; project administration: M.C.K., A.M.; funding acquisition: D.J.N., A.M.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-00799-1.

Correspondence and requests for materials should be addressed to A.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.