

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

Emanuele Del Fava<sup>1,2†</sup>, Irene Adema<sup>3†</sup>, Moses C. Kiti<sup>3</sup>, Piero Poletti<sup>4</sup>, Stefano Merler<sup>4</sup>, D. James Nokes<sup>3,5</sup>, Piero Manfredi<sup>6</sup>, Alessia Melegaro<sup>1,7\*</sup>

<sup>1</sup>Carlo F. Dondena Centre for Research on Social Dynamics and Public Policy, Bocconi University, Milan, Italy

<sup>2</sup>Max Planck Institute for Demographic Research, Rostock, Germany

<sup>3</sup>KEMRI-Wellcome Trust Research Programme, Kilifi, Kenya

<sup>4</sup>Fondazione Bruno Kessler, Trento, Italy

<sup>5</sup>The Zeeman Institute for Systems Biology and Infectious Disease Epidemiology Research (SBIDER), and School of Life Sciences, University of Warwick, United Kingdom

<sup>5</sup>Department of Economics, Pisa University, Pisa, Italy

<sup>7</sup>Department of Social and Political Sciences, Bocconi University, Milan, Italy

\*Corresponding author.

†These authors contributed equally to the work.

## Abstract

Which are the characteristics of contact patterns in diverse social contexts in sub-Saharan Africa, and which types of individuals and daily behaviours may play a pivotal role in infection transmission to the most vulnerable, such as older adults? We address these questions using novel survey data on social contacts and time use from a sample of 1407 individuals from rural, urban, and slum settings in Kenya. In the rural setting, we observed the highest number of daily social contacts (11.56, SD = 0.23) and the highest share of intergenerational mixing with older adults (7.5% vs. around 4% in the urban settings). Intergenerational mixing with older adults was mainly reported by individuals spending their day mostly in the general community (around 8%) or at home (5.1%), rather than at work (1.5%) or at school (3.6%). These results are essential to define effective interventions to control infection transmission in the African context.

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 analysis of age-specific mixing patterns is key to identifying the main transmission routes bringing the disease into the most vulnerable population segments. This 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 characterized 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<sup>1-11</sup>.

The heterogeneity in the observed contact patterns appears to be strongly related to the social context in which these surveys were conducted. These include the geographical area – with studies being carried out in sub-Saharan Africa (Kenya, South Africa, Uganda, Zambia, and Zimbabwe), East Asia and Pacific (China, Fiji, Thailand, and Vietnam), and Latin America (Peru); the urbanization level – with conflicting evidence for the difference in contact numbers between rural and urban settings<sup>5,7,8</sup>; the spatial distribution – with age-assortativeness and distance from home being positively correlated<sup>8</sup>; 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<sup>7</sup>; the ethnicity – with minimal mixing found between people of different ethnic background<sup>10</sup>.

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 characterized 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<sup>12-15</sup>.

The quantification of social contact patterns 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<sup>16-19</sup>.

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 five years of age<sup>20</sup>, with school-age children as the most likely source of infection for infants, therefore both representing the best targets of vaccination campaigns<sup>21,22</sup>. 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<sup>23-26</sup>, 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<sup>27,28</sup>. To add to the clinical challenges, also the contextual complexities, especially in LMICs, have to be considered where 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<sup>7,15,19</sup>.

For this purpose, we conducted a population-based survey in three diverse settings on the southern coast of Kenya 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 urban and rural settings and how these differences affected age-specific social mixing patterns. To this aim, study sites were selected to emphasize differences in terms of urbanization and living conditions. In particular, our study population consisted of (i) one rural village within the Kilifi Health and Demographic Surveillance System (KHDSS) in Kilifi county<sup>29</sup>, and (ii) two urban areas in Mombasa county, the second largest Kenyan city, namely, an informal settlement ("slum") and an urban mixed area, characterized by a mix of slum and non-slum elements.

As part of the study, participants reported the number of different persons encountered in two randomly assigned consecutive days, the type and the 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 filled in a contact diary during two different days. 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 summarized 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 characterized 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<sup>30</sup>.

After adjusting for survey design with sampling weights<sup>7</sup>, relevant differences between settings were found in terms of sex, age, household (HH) size, SES, and type of survey day (Fig. S2 in the supplementary material). 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.

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

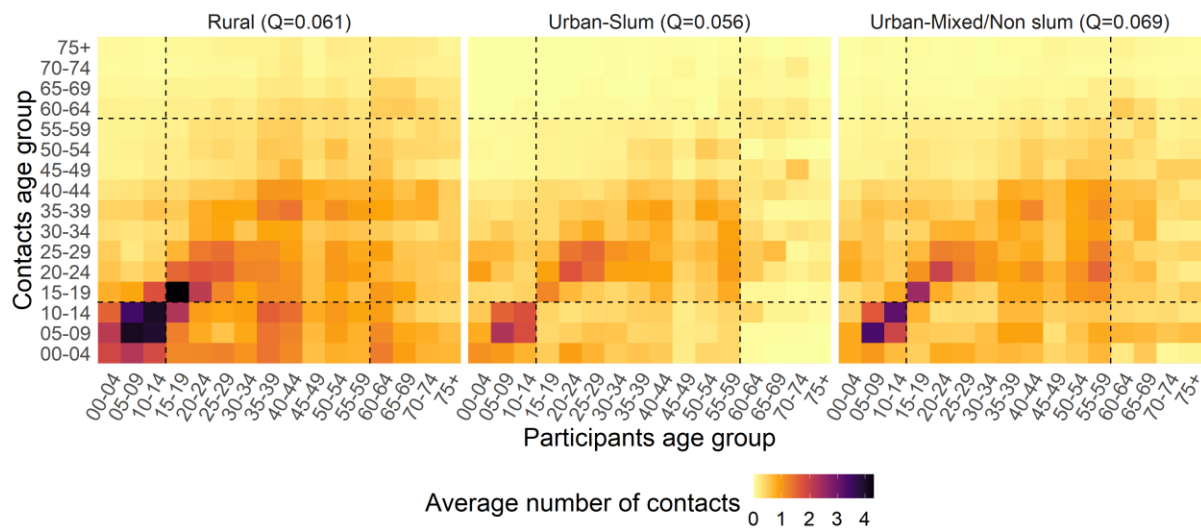
Variable	Category	Participants		Number of contacts		Relative number of contacts	
		N	%	Mean	SD	IRR	95% CI
Setting	Rural	512	36.4	11.56	0.23	1.00	[1.00,1.00]
	Urban slum	345	24.5	6.73	0.19	0.66	[0.61,0.71]
	Urban mixed/non-slum	550	39.1	7.72	0.22	0.76	[0.68,0.84]
Sex	Female	858	61.0	8.80	0.17	1.00	[1.00,1.00]
	Male	549	39.0	9.10	0.25	0.97	[0.92,1.03]
Age group	00-01y	131	9.3	5.53	0.25	0.67	[0.59,0.76]
	01-05y	192	13.6	8.70	0.31	0.98	[0.87,1.09]
	06-14y	232	16.5	11.28	0.33	1.21	[1.10,1.32]
	15-18y	139	9.9	10.49	0.52	1.14	[1.04,1.27]
	19-34y	393	27.9	7.89	0.25	1.00	[1.00,1.00]
	35-59y	248	17.6	8.60	0.34	1.04	[0.95,1.14]
	60y+	72	5.1	8.13	0.41	0.99	[0.87,1.13]
HH size	1-2p	203	14.4	6.78	0.36	1.00	[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]
	7p+	316	22.5	11.83	0.29	1.41	[1.22,1.62]
	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	1.00	[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	[0.61,1.81]
SES	Low	446	31.7	10.67	0.25	1.00	[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	1.00	[0.89,1.11]
Education level	Primary	498	35.4	10.11	0.25	1.00	[1.00,1.00]
	Secondary	274	19.5	8.15	0.29	0.98	[0.90,1.07]
	College/university	127	9.0	6.93	0.44	0.87	[0.77,0.99]
	None	483	34.3	8.60	0.22	0.99	[0.91,1.08]
	Missing	25	1.8	5.23	0.44	0.70	[0.59,0.83]
TU profile	Homestayors	995	70.7	8.45	0.15	1.00	[1.00,1.00]
	Workers	127	9.0	8.96	0.41	1.38	[1.25,1.52]
	Schoolers	104	7.4	11.56	0.58	1.24	[1.14,1.34]
	Walkers	108	7.7	10.36	0.52	1.10	[1.01,1.19]
	Commuters	63	4.5	9.86	0.68	1.21	[1.07,1.36]
	Travelers	8	0.6	6.01	1.51	1.08	[0.80,1.46]
	Missing	2	0.1	4.05	1.68	0.56	[0.30,1.05]
Day type	Weekday	908	64.5	9.07	0.16	1.00	[1.00,1.00]
	Weekend	499	35.5	8.61	0.21	1.02	[0.98,1.06]

## 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).

For each of the three settings, we built contact matrices by five-year age groups defining the intensity of interactions between the age groups (Fig. 1). For each matrix, we calculated the assortativity index  $Q$ , which quantified the importance of contacts occurring between individuals of the same age<sup>31-33</sup>. The lower the value of  $Q$ , the lower the assortativity and the higher the mixing between different 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. Overall, the index  $Q$  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. A comparison, on the one hand, with the estimates of  $Q$  for other LMICs countries (e.g., 0.051 in Zimbabwe<sup>7</sup>, and 0.075 in Vietnam<sup>3</sup> and Peru<sup>2</sup>, own calculations), and, on the other hand, with those from high-income countries in Europe (ranging from 0.14 in the United Kingdom to 0.19 in the Netherlands<sup>34,35</sup>) 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 (Figure 2C), possibly because of the higher rate of intergenerational residence and the larger HH size (Supplementary Fig. S3). A strong mixing between adults older than 20 years and younger than 60 years was also found across all settings.



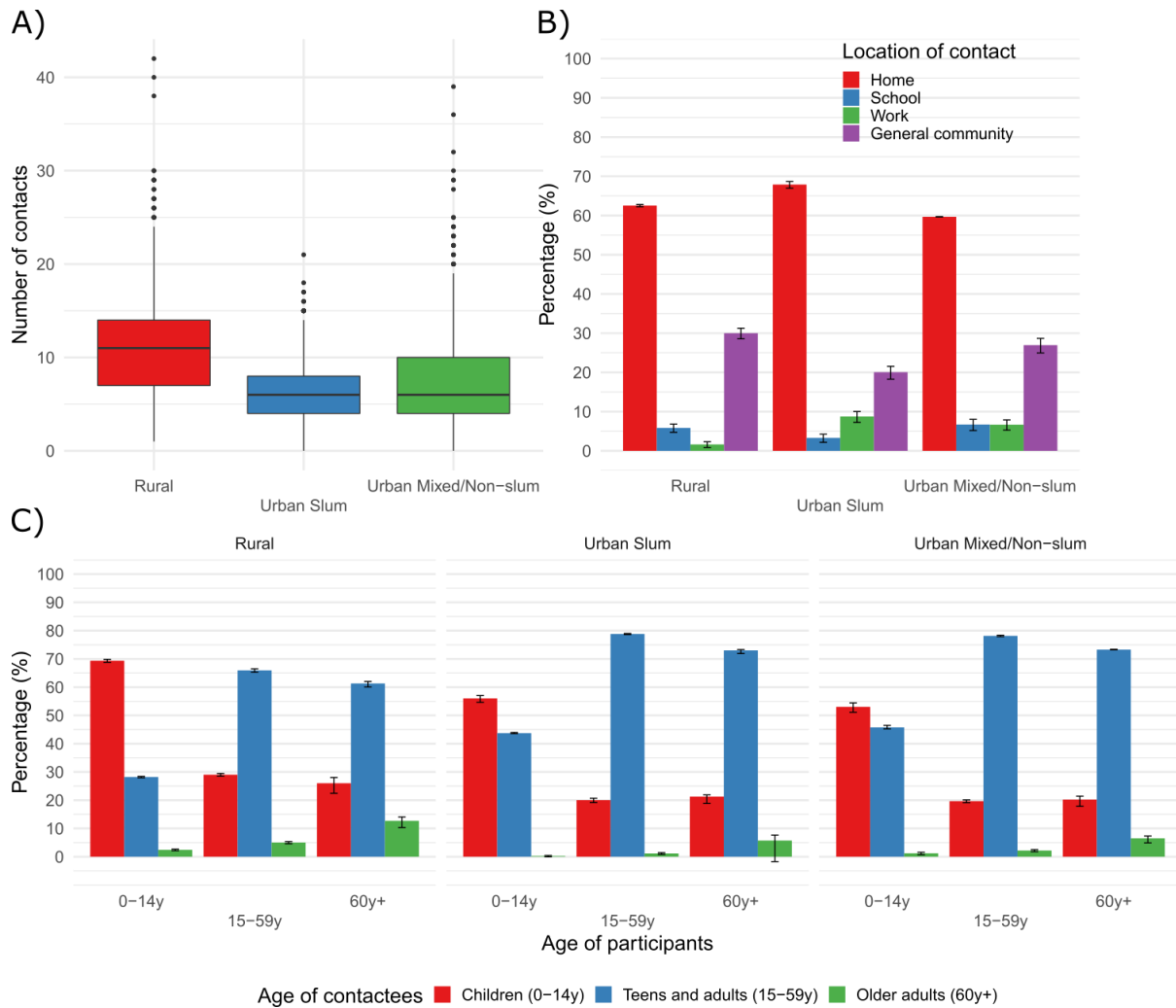


**Figure 1.** Average number of overall contacts between participants in age class  $i$  and contacted individuals in age class  $j$ , adjusting for reciprocity of contact at the population level, in each of the three settings. The dashed lines help identifying the contacts of children (0-14 years), teens and adults (15-59 years) and older adults (60+ years). At the top of each matrix, we report its assortativity index  $Q$ .

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' characteristics (Supplementary Fig. S4).

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) – accounted for a higher share of total contacts of the adults aged 60 years or more

(90.3%, of which 23.3%, CI 21.5%-24.6% with children) than for children younger than 15 years (37.1%, of which 1.7%, CI 1.5%-1.9% with older adults) or for teens and adults (26.9%, of which 3.2%, CI 2.9%-3.4% with older adults). Moreover, we found that the share of intergenerational contact reported with older adults was higher in the rural setting for both children (2.4%, CI 2.2%-2.7%) and teens and adults (5%, CI 4.6%-5.4%) (Fig. 2C).



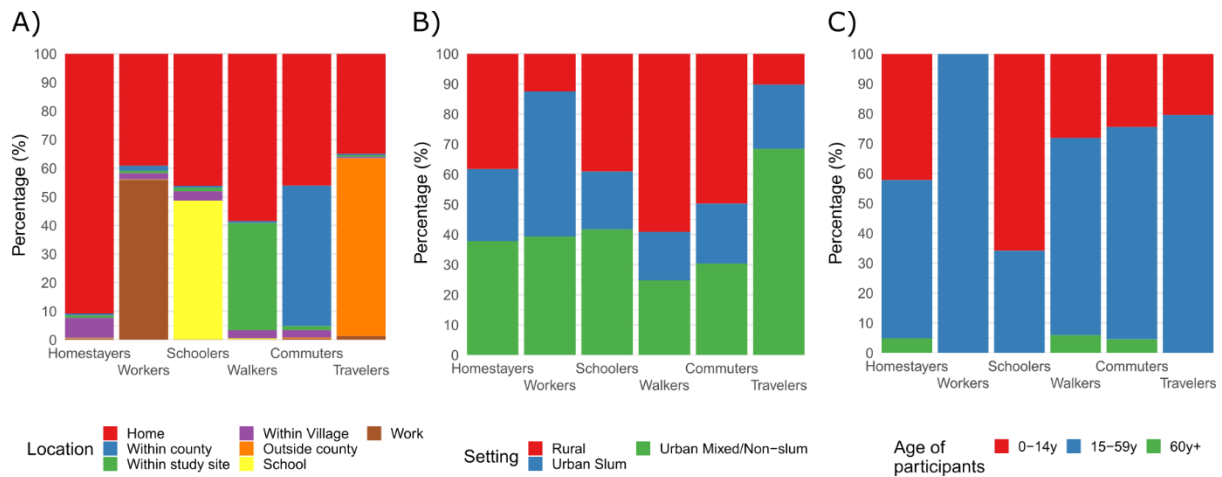
**Figure 2.** Distribution of daily social contacts by setting (A), percentage (with 95% CI) of daily social contacts by location of contact (B), and percentage (with 95% CI) of daily social contacts by setting and age of participants with children (0-14 years), teens and adults (15-59 years) and older adults (60+ years) (C).

### Differences in participants' daily routines

To understand the behavioural determinants of the number of social contacts, six profiles based on individuals' daily routine behaviour were identified by, first, applying principal component analysis (PCA) to the collected time-use (TU) data and, next, by applying a hierarchical cluster analysis to the retrieved principal components to group individuals into meaningful profiles<sup>36</sup>. Each generated TU profile group 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 (*Homestayors*) 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 characterized 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 a good share of their time *within the study site* (37.5%); the fifth group (*Commuters*: 4.6% of the total p.d.) encompassed individuals spending a relevant part of their day *within their county* (49.1%); finally, the last group (*Travelers*: 0.6% of the total p.d.) identified individuals spending a relevant part of their day *outside their county* (62.2%).

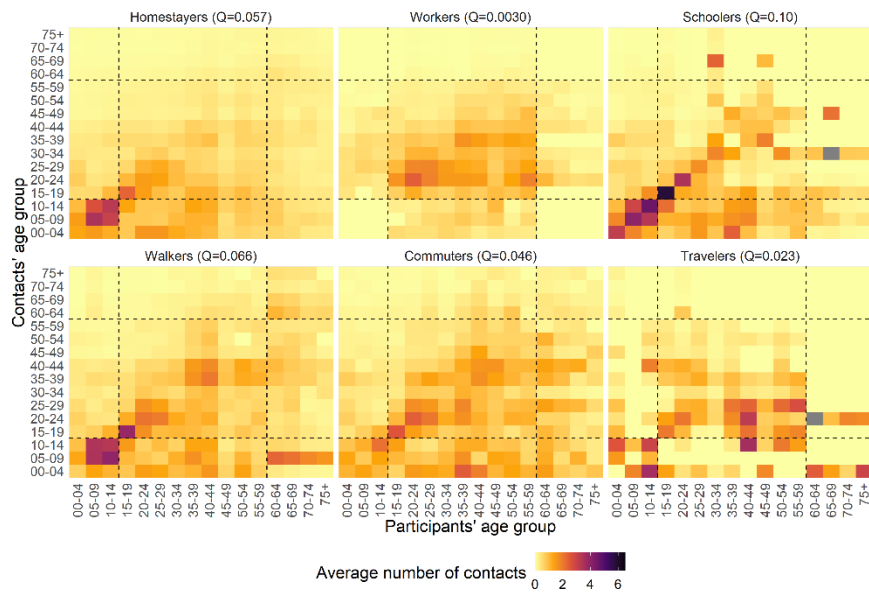


**Figure 3.** Description of the six TU profile groups by location (A), setting (B), and age group (C).

### Social contact patterns by TU profile

The TU profiles described above were used to assess differences in contact patterns between people characterized by different daily routines.

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.



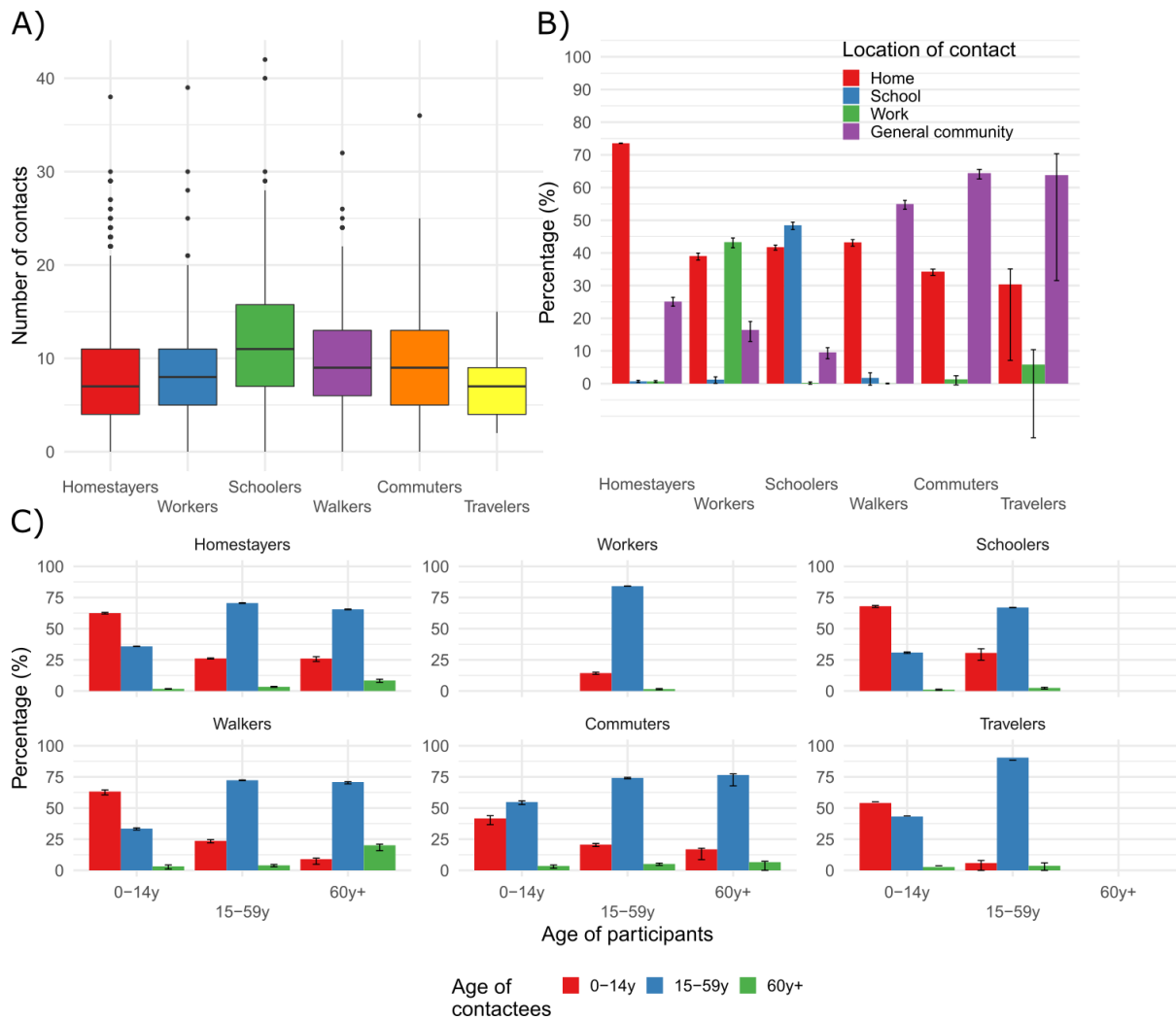
**Figure 1.** Average number of overall contacts between participants in age class  $i$  and contacted individuals in age class  $j$ , adjusting for reciprocity of contact at the population level, stratified by TU profile. The dashed lines help identifying the contacts of children (0-14 years), teens and adults (15-59 years) and older adults (60+ years). At the top of each matrix, we report its assortativity index  $Q$ .

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 (Fig. 3A, 5B) and reporting a higher number of contacts as occurring in this location.

TU profiles also varied in terms of intergenerational mixing (Fig. 5C). Contacts with children were mainly reported by Schoolers (55.4%), followed by Homestayers (43.2%) and Walkers (34%). 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 did not reach 15%. 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), Homestayors (almost 3% of contacts), and Walkers and Commuters (almost 5% of contacts). Such contacts with older adults were mainly reported by participants of the same generation; however, we also observed a substantial level of intergenerational contact reported by both children (at least 3.2%) and teens and adults (between 4% and 5%), suggesting that interactions with the older adults occurred mostly with those individuals spending a high proportion of their time in the general community.

After adjusting for participants' characteristics, we estimated that Workers and Schoolers reported overall 38% and 24% more contacts than Homestayors, respectively (Table 1). We did not find any significant difference in contacts between Workers and Homestayors at home and in the general community (Supplementary Fig. S6). Conversely, Schoolers reported less contacts than Homestayors both at home and in the general community, supporting the finding that most of their contacts occurred at school. Finally, Walkers and Commuters (who reported 10% and 21% more contacts than Homestayors, respectively) had less contacts at home and much more in the general community.



**Figure 2.** Distribution of daily social contacts (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).

## Discussion

Although most social contact studies in LMICs<sup>2,3,6</sup> have been predominantly conducted in rural settings, while those conducted in high-income countries<sup>31,34,35,37-42</sup> 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<sup>1,5,7,8</sup>. Moreover, only one study sought to investigate whether differences in social behaviour induced by daily routines entailed

differences in the average numbers of contacts<sup>7</sup>. 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<sup>5,43,44</sup>, 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<sup>7</sup>, they were consistent with what was observed by previous studies in the same area in Kenya<sup>5</sup>, in Zambia and South Africa<sup>1</sup>, 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)<sup>45</sup>. Despite the lower degree of age assortativeness in comparison to high-income countries, regardless of the type of setting, we observed 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 (Fig. 2C). 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<sup>21,22,26,28</sup>.



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<sup>5</sup>, while a small study conducted in a slum in the capital Nairobi in May 2020 found an average of 18 overall contacts per day<sup>44</sup>. 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<sup>44</sup> slums, we employed a modified version of the WHO Expanded Program on Immunization (EPI) cluster sampling technique<sup>46</sup> (see supplementary materials), 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 7am and 6pm, 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, we used face-to-face interviews conducted by a trained interviewer instead of self-reporting paper diaries<sup>5</sup> or phone interviews<sup>44</sup> to improve data quality, especially for those collected from illiterate respondents.

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 Homestayors, 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 characterized by a larger number of interactions between teens and adults and with older adults. These behaviours 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.

In conclusion, 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. More in general, 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.

## **Materials and Methods**

### **Study population**

The rural site was selected among the most rural villages 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<sup>29</sup>.

On the other hand, the urban site of Mombasa County, had a population of 939,370 in 2009 and in 2014 reported a population growth rate of 3.60% per year<sup>47</sup> and a TFR of 3.2<sup>48</sup>. Mombasa County was also characterized 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<sup>47</sup>.

### **Study design**

The study design was cross-sectional, targeting recruitment at the household level. 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<sup>5</sup>. 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.

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<sup>7,34</sup>, 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 (e.g., dawn, late morning, or dusk/early night) to a maximum of eight hours (late night). Respondents reported whether, for each time slot, they had visited or not a given location (home, school, work, and general community – within the village/estate, within the study site, within the county, or outside the county).

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.

### **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 PCA to data on household characteristics and owned assets, also collected in the study, and taking the first component as the index<sup>49</sup>. To construct the TU profiles, we applied PCA to the collected TU data to summarize all the information contained in the variables with a smaller number of dimensions, constructed to be independent of each other. We used the obtained components as input for a hierarchical cluster analysis to create groups of person-days characterized by similar time use behavior profiles<sup>36</sup>.

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<sup>50</sup>, the elements of the matrix were adjusted for reciprocity at the population level (see the supplementary materials for further details on the construction of the contact matrices).

We measured the age-assortativeness with the  $Q$  index<sup>31,33,51</sup>, 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  $Tr(\cdot)$  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<sup>32</sup>.

Generalized estimating equations (GEE) were chosen to evaluate the effect of the different participants' characteristics on the daily number of contacts<sup>52</sup>. 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)<sup>7,34,53</sup>. 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 the type of survey day (weekday or weekend). 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<sup>54</sup>); GEE models were estimated in Stata, using the procedure “xtgee”.

## References

1. Dodd, P. J. *et al.* Age- and Sex-Specific Social Contact Patterns and Incidence of *Mycobacterium tuberculosis* Infection. *Am. J. Epidemiol.* kwv160 (2015)  
doi:10.1093/aje/kwv160.
2. 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).
3. Horby, P. *et al.* Social Contact Patterns in Vietnam and Implications for the Control of Infectious Diseases. *PLoS ONE* **6**, e16965 (2011).
4. 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).
5. 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).
6. 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**, (2018).
7. Melegaro, A. *et al.* Social Contact Structures and Time Use Patterns in the Manicaland Province of Zimbabwe. *PLoS ONE* **12**, e0170459 (2017).
8. 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).

9. 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).
10. 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).
11. Huang, Y. *et al.* Spatiotemporal heterogeneity of social contact patterns related to infectious diseases in the Guangdong Province, China. *Sci. Rep.* **10**, 6119 (2020).
12. 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).
13. Menashe Oren, A. & Stecklov, G. Urban–rural disparities in adult mortality in sub-Saharan Africa. *Demogr. Res.* **39**, 136–176 (2018).
14. 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).
15. Trentini, F. *et al.* Demography, social contact patterns and the COVID-19 burden in different settings of Ethiopia: a modeling study. *medRxiv* 2020.11.24.20237560 (2020) doi:10.1101/2020.11.24.20237560.
16. 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).
17. 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).
18. Zhang, J. *et al.* Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. *Science* (2020) doi:10.1126/science.abb8001.

19. 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).
20. 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. *The Lancet* **390**, 946–958 (2017).
21. Poletti, P. *et al.* Evaluating vaccination strategies for reducing infant respiratory syncytial virus infection in low-income settings. *BMC Med.* **13**, (2015).
22. Yamin, D. *et al.* Vaccination strategies against respiratory syncytial virus. *Proc. Natl. Acad. Sci.* **113**, 13239–13244 (2016).
23. 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).
24. 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).
25. O’Driscoll, M. *et al.* Age-specific mortality and immunity patterns of SARS-CoV-2. *Nature* 1–6 (2020) doi:10.1038/s41586-020-2918-0.
26. Poletti, P. *et al.* Probability of symptoms and critical disease after SARS-CoV-2 infection. *ArXiv200608471 Q-Bio* (2020).
27. 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).
28. Davies, N. G. *et al.* Age-dependent effects in the transmission and control of COVID-19 epidemics. *Nat. Med.* 1–7 (2020) doi:10.1038/s41591-020-0962-9.
29. Scott, J. A. G. *et al.* Profile: The Kilifi Health and Demographic Surveillance System (KHDSS). *Int. J. Epidemiol.* **41**, 650–657 (2012).



30. 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**, (2018).
31. Ajelli, M. & Litvinova, M. Estimating contact patterns relevant to the spread of infectious diseases in Russia. *J. Theor. Biol.* **419**, 1–7 (2017).
32. 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).
33. 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).
34. Mossong, J. *et al.* Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases. *PLoS Med.* **5**, e74 (2008).
35. 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).
36. 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).
37. 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).
38. Ibuka, Y. *et al.* Social contacts, vaccination decisions and influenza in Japan. *J. Epidemiol. Community Health* **70**, 162–167 (2016).
39. Kwok, K. O. *et al.* Social contacts and the locations in which they occur as risk factors for influenza infection. *Proc. R. Soc. B Biol. Sci.* **281**, 20140709 (2014).

40. 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**, (2017).
41. Munasinghe, L., Asai, Y. & Nishiura, H. Quantifying heterogeneous contact patterns in Japan: a social contact survey. *Theor. Biol. Med. Model.* **16**, (2019).
42. Feehan, D. M. & Mahmud, A. S. Quantifying population contact patterns in the United States during the COVID-19 pandemic. *Nat. Commun.* **12**, 893 (2021).
43. Kiti, M. C. *et al.* Quantifying social contacts in a household setting of rural Kenya using wearable proximity sensors. *EPJ Data Sci.* **5**, (2016).
44. 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).
45. 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).
46. World Health Organization. *Immunization coverage cluster survey : reference manual*. <https://apps.who.int/iris/handle/10665/69087> (2005).
47. Kenya National Bureau of Statistics. *Mombasa Informal Settlement Survey, Kenya, 2009*. (2010).
48. 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. *Kenya Demographic and Health Survey 2014*. (2015).
49. Vyas, S. & Kumaranayake, L. Constructing socio-economic status indices: how to use principal components analysis. *Health Policy Plan.* **21**, 459–468 (2006).
50. 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).

51. 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).
52. Liang, K.-Y. & Zeger, S. L. Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22 (1986).
53. 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**, (2009).
54. Lê, S., Josse, J. & Husson, F. FactoMineR: An R Package for Multivariate Analysis. *J. Stat. Softw.* **25**, 1–18 (2008).

## Acknowledgements

We thank all the study participants for their contribution of time and data. We also thank the Tudor location community health volunteers and local administration in Mombasa for providing security throughout the data collection period. We are grateful to the field study team lead by Grace Jumbale and John Koi who dedicated their time to make sure data collection was successful. We acknowledge Edward Mundia, Lilian Mwangi, and Grieven Otieno for the digital system development, the data clerks for data entry and management, and the Kilifi HDSS registry. This article is published with the permission of the Director of KEMRI. This work has received funding from the European Research Council under the European Union's Seventh Framework Program (FP7/2007-2013)/ERC Grant Agreement 283955 and has been partially supported by the Wellcome Trust (Grant # 203077/Z/16/Z).

## Author contributions

**Emanuele Del Fava:** Conceptualization, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft, review & editing. **Irene Adema:** Data curation, Investigation, Resources, Writing – original draft, review & editing. **Moses C. Kiti:** Data curation, Investigation, Project administration, Writing – review & editing. **Piero Poletti:** Conceptualization, Investigation,

Methodology, Writing – review & editing. **Stefano Merler:** Supervision, Writing – review & editing.

**D. James Nokes:** Funding acquisition, Supervision, Writing – review & editing. **Piero Manfredi:**

Conceptualization, Supervision, Writing – review & editing. **Alessia Melegaro:** Conceptualization,

Funding acquisition, Project administration, Supervision, Writing – review & editing.

## Competing interests

The authors declare no competing interests.

## Data availability

The datasets and the source codes necessary for the reproducibility of our results are available at the following [blind for peer-review] repository:

[https://osf.io/xtw8h/?view\\_only=27b7ccf00a4247cea0a50b69617a988f](https://osf.io/xtw8h/?view_only=27b7ccf00a4247cea0a50b69617a988f).