

1 **Characterizing social contact patterns in rural and urban Mozambique: the GlobalMix** 2 **study, 2021-2022**

3 **Authors**

4 Moses C. Kiti^{1,a}, Charfudin Saco^{2,a}, Obianuju G. Aguolu³, Alana Zelaya¹, Holin Chen¹, Sara S.
5 Kim¹, Nilzio Cavele², Edgar Jamisse², Corssino Tchavana², Americo Jose⁴, Ivalda Macicame⁵,
6 Orvalho Joaquim², Noureen Ahmed⁶, Carol Y. Liu¹, Inci Yildirim³, Kristin Nelson¹, Samuel M.
7 Jenness¹, Herberth Maldonado⁷, Momin Kazi⁸, Rajan Srinivasan⁹, Venkata R. Mohan⁹, Alessia
8 Melegaro¹⁰, Fauzia Malik⁶, Azucena Bardaji¹¹, Saad B. Omer^{6*}, Ben Lopman^{1*}

9 **Institutional affiliation**

10 ¹ Emory University, USA

11 ² Manhica Health and Research Centre (CISM), Mozambique

12 ³ Yale University, USA

13 ⁴ Polana-Caniço Health and Research Centre (CISPOC), Mozambique

14 ⁵ National Institute of Health, Mozambique

15 ⁶ University of Texas SouthWestern Medical College, USA

16 ⁷ Universidad del Valle de Guatemala, Guatemala

17 ⁸ Aga Khan University, Pakistan

18 ⁹ Christian Medical College, India

19 ¹⁰ Bocconi University, Italy

20 ¹¹ ISGlobal, Spain

21 ^a Joint first authors

22 ^{*} Contributed equally

23 **1. Abstract**

24 **Introduction:**

25 There are few sources of empirical social contact data from resource-poor settings thus limiting
26 the development of contextual mathematical models of disease transmission and control.

27 **Methods:**

28 We collected and analyzed cross-sectional survey data from rural and urban sites in
29 Mozambique. Participants, including infants, were recruited. They reported retrospectively, in a
30 paper diary, individuals with whom they had a co-located physical or conversation contact, as
31 well as their age, sex, relationship, frequency, and duration of the contact. We compared vaccine
32 effects (VE) by parameterizing transmission models using empirical and synthetic contact rates.

33 **Results:**

34 1363 participants recruited between April 2021 and April 2022 reported a mean of 8.3 (95% CI
35 8.0–8.6) contacts per person on day 1. Mean contact rates were higher in the rural compared to
36 urban site (9.8 [9.4–10.2 vs 6.8 [6.5–7.1], $p < 0.01$), respectively. Participants aged ≤ 18 years

37 were the main drivers of higher physical contacts. In the model, we report higher VE in the rural
38 site when comparing empirical to synthetic contact matrices (32% vs 29%, respectively), and
39 lower corresponding VE in urban site (32% vs 35%). These effects were prominent in the
40 younger (0-9 years old) and older (60+ years) individuals.

41 Conclusion:

42 Our work suggests differences in contact rates and patterns between rural and urban sites in
43 Mozambique, with corresponding differences in vaccine effects on an infectious pathogen. We
44 also demonstrate the utility of empirical data in infectious disease modelling for high-burden,
45 low-income settings.

46 **2. Background**

47 Human social contact patterns drive the transmission of pathogens that spread through proximity
48 to infectious individuals. Data on social contact patterns are critical to understand “who-contacts-
49 whom” and infer who-acquires-infection-from-whom, providing insight on potential control
50 measures such as physical distancing and vaccination. Underlying the patterns of contact are
51 demographic, sociocultural, and economic determinants which vary within and across regions
52 resulting to corresponding variation in contact patterns. Unfortunately, such critical data are not
53 as widely available in low-and-middle-income countries (LMICs), including Mozambique [1], as
54 they are in high-income countries (HICs) [2]. Where data are available, they were collected
55 across e.g., rural-urban divide [3–6] or informal settlement [7,8] limiting the representativeness
56 of the data. More recent studies in LMICs have incorporated innovative methods to obtain data
57 from hard-to-reach populations especially children and illiterate adults by either using shadows
58 [3,9] interviewer-led questionnaires [4,5], or wireless proximity sensors [10,11]. In the absence
59 of locally collected data, simulated contact rates for LMIC populations may also be derived by
60 projecting empirical data collected from HICs (e.g., POLYMOD data) and scaled using local
61 demographic patterns [12]. However, these extrapolations likely mischaracterize contact
62 patterns in important ways when they differ for reasons aside from demographics.

63 During the early phase of the COVID-19 pandemic, it became clear the extent to which human
64 contact patterns could be modified, and at essentially a global scale [13–17] In the absence of
65 vaccines or pharmaceutical interventions, populations responded by “physical distancing”, i.e.,
66 drastically reducing the number and riskiness of contacts. Work contacts and those out of the
67 home decreased and masking increased. Again, however, relatively little data were collected
68 from LMICs, limiting our ability to quantify these changes and use such data to develop models
69 of interventions.

70 Starting 2021, during the COVID-19 pandemic, we launched the “GlobalMix Study” to collect
71 social contact data from four LMICs: Mozambique, Guatemala, India, and Pakistan. Across all
72 four countries, we are collecting data from selected rural and urban areas using standardized
73 questionnaires and methods that were customized for each context considering social and
74 cultural norms [18]. In this paper, we present the methods and results from Mozambique for
75 which we have complete datasets.

76 **3. Methods**

77 3.1 Study objectives

78 The main aim of this study was to characterize the age and location patterns of social contact and
79 mixing with respect to directly transmitted infections in rural and urban sites of Mozambique.
80 Then, to examine the impact of using these contextual data, we simulated the transmission of a
81 hypothetical respiratory virus and effects of vaccination in a model using contact data generated
82 from this study (henceforth called empirical data) and compared to empirically constructed
83 contact data (henceforth called synthetic data).

84 3.2 Study design

85 This was a cross-sectional study conducted from March 2021 through to April 2022. Our data
86 collection period coincided with active SARS-CoV-2 transmission in Mozambique [19]. The
87 rural site was in Manhiça District where the Manhiça Health and Research Centre (CISM) runs
88 the Manhiça Health and Demographic Surveillance System (Manhiça HDSS) [20]. The urban
89 site was in Maputo City where the Polana-Caniço Health and Research Centre (CISPOC) of the
90 National Institute of Health of Mozambique runs the Polana-Caniço HDSS [21]. Prior to
91 collecting the social contact data, we held 25 focus group discussions and 40 cognitive
92 interviews with community members drawn from the two sites. We aimed to understand the
93 determinants of human interaction at the study sites, and explore the perceptions, acceptability,
94 and utility of paper diaries for data collection. We also wanted to get community buy-in and get
95 good practice recommendations on our research implementation process.

96 Complete details of the sample size, data collection tools and procedures, and data analysis
97 methods have been described in our protocol [18]. Briefly, in Mozambique, we aimed to collect
98 data from 630 individuals per site equally stratified into 10 age groups: 0-5 months, 6-11 months,
99 1-4, 5-9, 10-14, 15-19, 20-29, 30-39, 40-59 and 60+ years, resulting in 63 participants per age
100 group. The data collection was divided into three phases: enrollment, diary-keeping, and exit
101 interview. Individuals were randomly selected by age and sex from the HDSS registers. Each
102 participant filled in their social contacts over two days. We defined a social contact (henceforth
103 called contact) as a two-way face-to-face encounter that involved either i) physical touch
104 involving a skin-to-skin touch or over clothes, or 2) non-physical, whereby two or more
105 individuals had a conversation interaction only while standing within arm's length of each other
106 and with no physical barrier between them. Additional qualitative questions included difficulty
107 encountered during filling the diary and perceived social behavior and mobility change during
108 the study period compared to periods before the COVID-19 outbreak. Contact data, the
109 enrollment and exit interview were captured electronically by fieldworkers via REDCap [22]
110 forms coded in portable electronic tablets. The diary codebook in English is available in our
111 GitHub repository (See **10. Availability of data and materials** section). All children <10 years
112 old and illiterate individuals ≥ 10 years old selected or were assigned a shadow to record contacts
113 on behalf of the participant. In addition, the shadows were instructed to be discrete and did not
114 need to follow the participant all day.

115 3.3 Data analysis

116 3.3.1 Characteristics of contact patterns

117 We estimated the mean contact rates per person with bootstrapped 95% confidence intervals (CI)
118 and median contact rates per person with inter-quartile range (IQR) over two days and for day 1
119 only. Assuming x_{ij} represents the total number of contacts between participants in age group i
120 and contacts in age group j , then the mean number of reported contacts (m_{ij}) was calculated as
121 x_{ij}/n_i , where n_i is the study population in group i .

122 The mean contact rates were stratified by site, then further by age, sex, day of the week
123 (weekday vs weekend), type of contact (physical or conversation only), household membership

124 (household member vs non-household member), occupation/ daily activity, and whether the
125 participant reported symptoms of acute respiratory infections (ARI) or acute gastroenteritis
126 (AGE) within the 14 days prior to survey. We used the Wilcoxon rank sum test to assess the
127 difference between median contact rates within the sites for each covariate, and between the rural
128 and urban sites.

129 Lastly, we computed age-stratified contact matrices to quantify the interactions between age
130 groups. We adjusted the contact matrices to account for reciprocity, assuming that the total
131 number of contacts from age group i to j were equal to the number of contacts from age group j
132 to i i.e., $m_{i,j} = m_{j,i}$ [6]. We present the age-specific contact matrix using data from day 1 only
133 using the revised formula $((m_{i,j} + m_{j,i}) / (n_i + n_j))$ which does not adjust for Mozambique age-
134 specific population size.

135 3.3.2 Characterizing location-specific proximity contact exposures.

136 We compared close, individually recorded diary contacts (close contacts) to co-location with
137 others but without direct interaction (proximity contacts) collected using the place-use surveys.
138 We describe the number of unique visits locations including other person's home, street,
139 market/shop, transport/hub, agricultural field, school, work, place of worship, well, playground
140 or other and the distribution of time spent at each place. We then compared the number of
141 proximity contacts per participant at each location type to the total close contacts per participant
142 at the same location type we examined for differences in patterns between the urban and rural
143 site.

144 3.3.3 Transmission and vaccine effects modeling.

145 To explore the utility of the empirical contact patterns, we compared rural and urban respiratory
146 virus transmission models parameterized with empirical data to those parameterized using
147 synthetic contact data modelled using Mozambique demographic data and POLYMOD data [12].
148 We built a deterministic susceptible-infectious-recovered (SIR) model with a vaccine conferring
149 protection against infection. We computed the mean number of contacts by re-classifying the
150 participants into six 10-year age classes from 0–59 years and one age group for individuals aged
151 60 years or older for compatibility with the synthetic data. We weighted the empirical contact
152 rates using 2021 rural and urban population distribution data and adjusted for reciprocity using
153 the *socialmixr* package. We modeled vaccination as 'leaky,' providing partial protection for
154 those vaccinated (50% vaccine coverage, 50% effectiveness), duration of illness as 7 days and
155 fixed the basic reproduction number at 2.5. We calculated the attack rate for no vaccine (AR_0)
156 and vaccine (AR_v) scenarios separately for rural and urban site and present the overall vaccine
157 effect (VE) calculated as the percent reduction of cases in the presence versus absence of
158 vaccine, i.e.,

159 $VE = \left(1 - \left(\frac{AR_v}{AR_0}\right)\right) * 100$. We used the *EpiModel* package to run all transmission models. All
160 analysis was done in R v4.3.2.

161 3.6 Ethical considerations

162 This study was approved by CISM Internal Scientific Committee (CCI/03/2020), the Internal
163 Ethical Review Board (initial approval CIBS-CISM/011/2020), the Emory University
164 Institutional Review Board (approval number 00105630), and Yale University (reliance
165 agreement approval number 2000026911). Field staff were responsible for obtaining written
166 informed consent from each participant prior to any data collection. Adults (≥ 18 years) provided
167 written consent, children (13–17 years) required personal written assent and parental consent,
168 while minors (< 13 years) only required parental consent.

169 4. Results

170 4.1 Baseline characteristics of participants

171 We approached 1693 residents and retained a total of 1363 (81% overall participation rate)
172 participants across both sites. We report equal participation rate in the rural (676/800, 85%) and
173 urban (687/893, 86%) site. We exceeded our target sample size by 103 participants particularly
174 in those aged 40–59 years. Of the 1363 participants, 666 (49%) were female equally distributed
175 by site. By site, there was no major difference in number of participants recruited by age, sex,
176 and school enrollment status (**Table 1**).

177 The mean household size was 5.5 (range 1–18) and 5.7 (range 1–20) in the rural and urban sites,
178 respectively. Overall, 379 (45%) households had 4–6 members and 4 (5%) had one resident.
179 When we omit children, students and the unemployed, the most common reported occupations in
180 the urban site were small business owners and vendors classified as businesspeople (16%,
181 59/366), office workers (15%, 56/366), and casual laborers (15%, $n=56/366$) while farmers
182 comprised 16% (64/394) in the rural site. 942 (69%) of participants reported wearing a mask
183 inside or outside the house, with no difference by site. About one-fifth (233) of participants
184 reported having > 1 ARI symptom and 26 (2%) reported at least one AGE symptom.

185 About half (51%, $n=701$) of the participants were able to read and write. The majority (88%,
186 $n=1200$) of the participants said that they reported all contacts. However, 51% (698/1363)
187 required assistance from a fieldworker to fill in the diary at the end of the two days (rural 43% vs
188 urban 56%). Generally, all children < 5 years ($n=409$) had a family member as a shadow, and out
189 of these 243 (50%) required additional assistance from the fieldworker. Among other ages, there
190 was no difference in proportion of those requiring a shadow/ help from fieldworker compared to
191 no help, apart from 15–19 (33% [41/124]) and 60+ year-olds (60%, [75/124]). Eight participants
192 reported testing positive for SARS-CoV-2, all of whom reported either going to quarantine
193 (government facility, $n=5$) or self-isolation at home ($n=3$).

194 4.2 Contact patterns

195 Participants reported a total of 17674 contacts over two days, with 41% unique contacts ($n=3904$
196 day 1 only and $n=3250$ day 2 only) and 59% ($n=10304$) reported on both days (repeat contacts).
197 Participants reported an overall mean of 13.1 (95% CI 12.6–13.5) contacts over two days (**Table**
198 **2**). We observe a significant difference in the mean number of contacts reported on day 1
199 compared to day 2 (8.3 [8.0–8.6] vs 5.5 [5.3–5.7]), respectively, paired t-test $p < 0.01$). Since diary
200 completion dates were randomly assigned, the actual mean contacts should not vary between the

201 first and second date of diary completion. Therefore, we believe that the observed difference is a
202 result of reporting bias resulting from participant fatigue; from here henceforth we report the
203 mean and median number of contacts on day 1 only. The rural mean contact rate was
204 significantly higher than the urban rate (9.8 [9.4–10.2] vs 6.8 [6.5–7.1], $p<0.01$) (**Figure 1 A &**
205 **B**). Contact rates were higher in rural areas for each age group (Figure 1 C). While rural mean
206 number of contacts with non-household members was significantly higher than contacts with
207 household members (6.6 vs 3.9, $p<0.01$), there was marginal difference in the urban site (4.2 vs
208 3.6, $p=0.46$). Corresponding median values for day 2 by site are available in **Supplementary**
209 **table 1**.

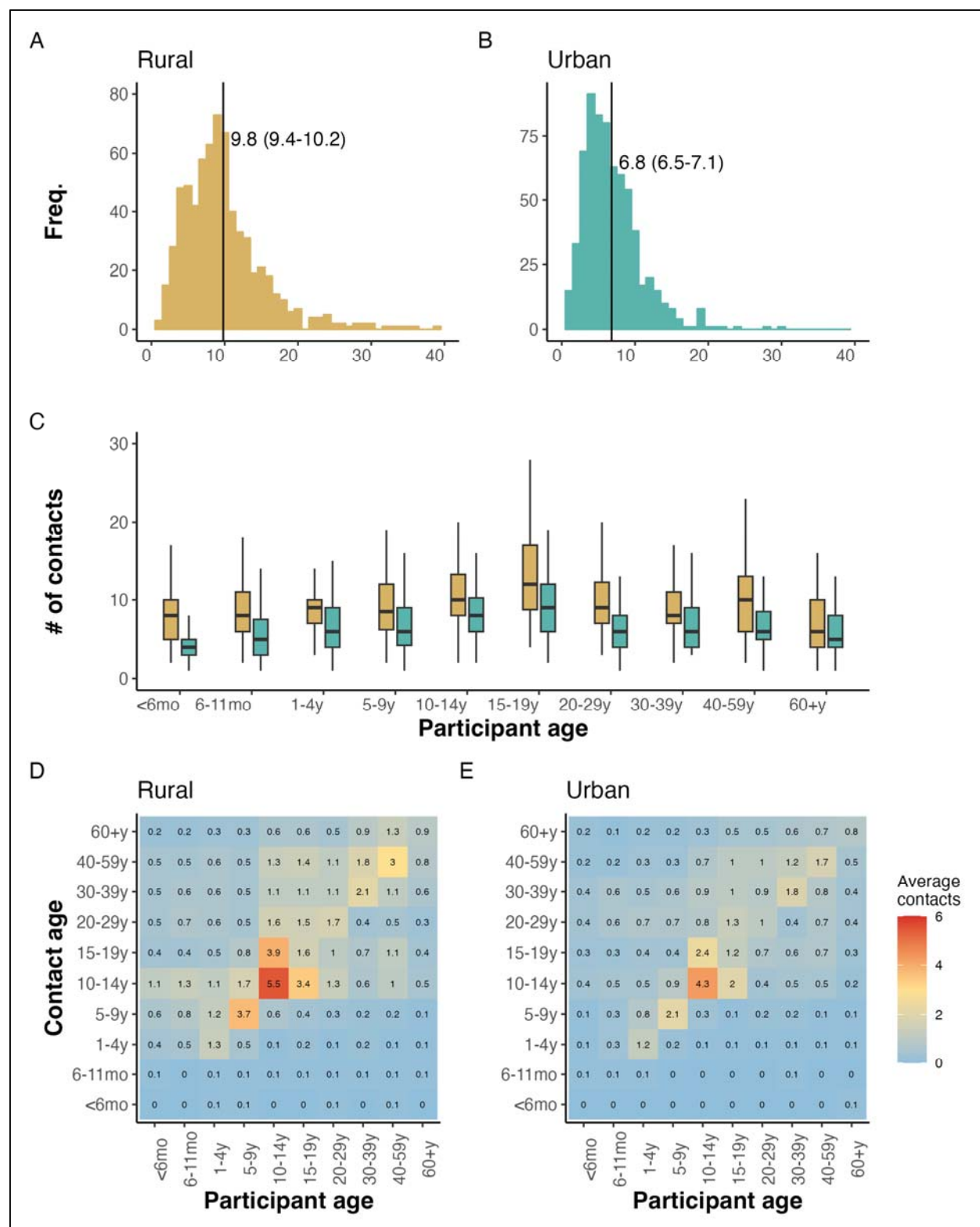
210 Physical contacts were, on average, more numerous than conversation only contacts in both rural
211 (6.7 vs 4.9, $p<0.01$) and urban areas (5.3 vs 3.3, $p<0.01$). Participants aged ≤ 18 years were the
212 main drivers of the higher physical contacts. Out of all participants, 803 (59%) reported having
213 the same number of social contacts compared to periods before the COVID-19 pandemic. From
214 these, urban participants reported either significantly fewer ($n=238$, mean = 5.9 [95% CI 5.5–
215 6.3]) or more ($n=28$, 11.7 [9.3–14.1]) mean number of contacts compared to those who reported
216 no change ($n=410$, mean = 7.0 (6.7–7.4)). In the rural site, 74 (11%) of the participants reported
217 more mean contacts than usual (12.6, 95% CI 10.9–14.3) and this was significantly different
218 from those who reported either no change ($n=384$, mean = 9.6 [9.1–10.1]) or fewer contacts
219 ($n=215$, mean = 9.2 [8.5–9.9]).

220 **Table 1. Characteristics of study participants in rural and urban sites in Mozambique.**

	Overall N = 1363 (%)	Rural N = 676 (%)	Urban N = 687 (%)
Sex¹			
Female	666 (49)	332 (49)	334 (49)
Male	696 (51)	343 (51)	353 (51)
Participant age			
<6mo	128 (9)	62 (9)	66 (10)
6-11mo	146 (11)	82 (12)	64 (9)
1-4y	135 (10)	63 (9)	72 (10)
5-9y	122 (9)	64 (9)	58 (8)
10-14y	125 (9)	61 (8)	64 (9)
15-19y	124 (9)	64 (9)	60 (9)
20-29y	125 (9)	64 (9)	61 (9)
30-39y	125 (9)	64 (9)	61 (9)
40-59y	209 (15)	89 (13)	120 (17)
60+y	124 (9)	63 (9)	61 (9)
Able to read and write			
Yes	701 (51)	293 (43)	408 (59)
Currently enrolled in school			
Yes	368 (28)	173 (26)	195 (29)
Occupation/ daily activity²			
Child	274 (23)	144 (24)	130 (22)
Unemployed	162 (14)	97 (16)	65 (11)
Student	324 (27)	153 (26)	171 (29)
Homemaker	33 (3)	9 (2)	24 (4)
Casual laborer	78 (7)	22 (4)	56 (9)
Farmer	70 (6)	64 (11)	6 (1)
Businessperson	66 (6)	7 (1)	59 (10)
Office worker	83 (7)	27 (5)	56 (9)
Retired	20 (2)	5 (1)	15 (3)
Other	74 (6)	58 (10)	14 (2)
Regular mask use			
Yes	942 (69)	435 (64)	507 (74)
Acute gastroenteritis			
Diarrhea/ vomiting	27 (2)	15 (2)	12 (2)
Acute respiratory infection			
>1 symptom	233 (17)	122 (18)	111 (16)
Who filled the diary?			
Self	665 (49)	361 (54)	304 (44)
By fieldworker	698 (51)	315 (46)	383 (56)

¹ 2 participants did not report their sex.

² 179 participants did not report their Occupation/ daily activity



221

222

223

224

Figure 1. Distribution patterns of number of contacts in rural and urban Mozambique.

Panels A and B show the density distribution of the number of contacts per person in the rural and urban site, respectively, with the mean and 95% CI shown by the black vertical line. Panel C

225 shows boxplots of the distribution of number of contacts by site. The middle line shows the
 226 median number of contacts and the lower and upper line in the boxplot shows the lower and
 227 upper quartile of the distribution, respectively. Panels D and E show the age specific contact
 228 matrices depicting the average mean number of individuals in age-group j (y-axis) with whom a
 229 participant in age-group i (x-axis) came into contact with.

230

231 **Table 2. Median (IQR) and mean (bootstrapped 95% CI) number of contacts on day 1**
 232 **only.**

	Rural		Urban	
	Median	Mean	Median	Mean
Sex				
Female	9 (6–11)	9.6 (8.9–10.2)	6 (4–8)	6.7 (6.3–7.1)
Male	9 (7–13)	10.1 (9.5–10.6)	6 (4–9)	7 (6.6–7.4)
Age				
<6mo	8 (5–10)	7.8 (6.8–8.9)	4 (3–5)	4.4 (3.8–4.9)
6–11mo	8 (6–11)	8.6 (7.7–9.4)	5 (3–7.5)	5.8 (5–6.6)
1–4y	9 (7–10)	8.9 (8–9.9)	6 (4–9)	7.1 (6.1–8.1)
5–9y	9 (6.8–12)	10.1 (8.6–11.6)	6 (4.25–9)	7.2 (6.1–8.2)
10–14y	10 (8–14)	11 (9.5–12.4)	8 (6–10.3)	8.7 (7.7–9.7)
15–19y	12 (8.8–17)	12.8 (11.4–14.2)	9 (6–12)	9.6 (8.3–10.9)
20–29y	9 (7–12.25)	10.8 (9.3–12.3)	6 (4–8)	6.1 (5.5–6.8)
30–39y	8.5 (7–11)	9.4 (8.1–10.7)	6 (4–9)	7 (6.1–7.9)
40–59y	10 (6–13)	11 (9.5–12.5)	6 (5–8.5)	6.8 (6.1–7.5)
60+y	6 (4–10)	7.4 (6.1–8.7)	5 (4–8)	5.6 (4.8–6.4)
Occupation/daily activity				
Child	8 (5–10)	8.2 (7.6–8.9)	4 (3–6)	5.1 (4.6–5.6)
Unemployed	7.5 (5–12)	8.9 (7.7–10)	6 (4–9)	7.1 (5.9–8.2)
Student	10 (8–14)	11.4 (10.5–12.3)	8 (5–10)	8.4 (7.7–9)
Homemaker	9 (7–10)	8 (5.2–10.8)	6 (4.75–8)	6.8 (5.5–8)
Casual laborer	8 (6–11.5)	9.5 (6.9–12.1)	5.5 (4.8–7)	6 (5.4–6.7)
Farmer	9 (6–13.25)	10.6 (8.9–12.4)	3.5 (3–4)	3.8 (2–5.6)
Businessperson	10 (7.5–11.5)	10.1 (5.1–15.2)	6 (4–8)	7.1 (6–8.3)
Office worker	8 (7–11)	9.9 (7.3–12.4)	5 (3.5–5.5)	5.1 (3.4–6.9)
Retired	5 (4–6)	5.6 (2.2–9)	10 (8.3–10)	9.1 (7.2–11.1)
Other	10 (7–13)	11.1 (9.5–12.7)	NA	NA
Household size				
1	6 (4–9.3)	7.2 (5.4–8.9)	4 (2.5–7)	5.2 (2.4–8)
2–3	7 (4–10)	8.1 (6.9–9.2)	5 (3–7.5)	5.6 (4.9–6.3)
4–6	9 (6–12)	9.6 (8.8–10.3)	6 (4–8)	6.5 (6–7)
7–10	10 (8–13)	10.9 (9.9–11.9)	7 (5–9)	7.8 (6.8–8.7)
10+	14 (9.5–16.5)	13.8 (12–15.7)	9 (6–12)	9.6 (7.9–11.4)
Household membership				
Member	9 (6–12)	10 (9.6–10.4)	6 (4–9)	6.9 (6.6–7.2)
Non-member	7 (4–9.5)	7.6 (6–9.2)	5 (3–8)	5.7 (4.7–6.8)
Enrolled in school				
Yes	9 (7–13)	10.8 (9.9–11.7)	8 (5–10)	8 (7.4–8.6)
No	9 (6–11)	9.4 (8.9–9.9)	6 (4–8)	6.3 (6–6.7)
Weekday/Weekend				
Weekday	9 (6–12)	9.8 (9.3–10.3)	6 (4–9)	7 (6.6–7.3)
Weekend	9 (6–12)	9.8 (9.1–10.5)	6 (4–8.5)	6.4 (5.9–7)
ARI symptoms				
>1 symptom	10 (7–13)	10.7 (9.7–11.6)	6 (4–8)	7.1 (6.3–7.9)
AGE symptoms				
Yes	9 (8–13)	9.7 (7.2–12.1)	6 (4–8.5)	6.6 (4.5–8.8)

233

234 *4.2.1 Contact matrices*

235 The patterns of age-specific social contacts are generally similar when comparing the urban and
 236 rural sites (**Figure 1** panel D and E). The urban matrix, however, suggests lower mean number of
 237 contacts across all ages compared to the rural site. Participants aged 5-14 years (school-going
 238 children) and older working adults aged 30-59 years in both rural and urban areas reported
 239 higher assortative (with same age) mean contacts. Older school goers aged 15-19 years also
 240 reported on average, high number of contacts with 10-14-year-olds in both sites. Another peak in
 241 mean number of contacts is observed between 30-39- and 40-59-year-olds, driven mostly by
 242 conversation only contacts (**Supplementary figure 2** panels (C) and (D)). In both sites, few
 243 interactions are reported between the other ages and little to no contacts are reported with infants.
 244 Despite low mean contacts reported by infants, the mean number of contacts between them and
 245 other ages generally increased with age to peak at 10-14 years in rural and 30-39 years among
 246 urban dwellers.

247 *4.2.2 Patterns of contact by location*

248 In both urban and rural areas, the estimated number of co-located individuals by far exceeded the
 249 number of contacts reported by participants. Overall, participants reported a mean of 26.9 (95%
 250 CI 23.5-30.3) proximity contacts compared to 6.8 (6.5-7.1) close contacts in the urban site, and
 251 23.1 (20.3-25.9) proximity contacts and 9.8 (9.4-10.2) close contacts in the rural site. The three
 252 locations with highest mean number of contacts were places of worship, schools, and transport
 253 hubs (**Table 3**). In addition, rural participants were more likely (n=752 visits, [48%]) to visit
 254 other homes compared to urban participants (n=288 visits [29%]). Despite overall numbers being
 255 similar, the locations where contact occurred was meaningfully different between urban and rural
 256 sites.

257 **Table 3. Number of times that participants reported visiting a location (N) and the median**
 258 **(IQR) number of people reported per location on day 1.**

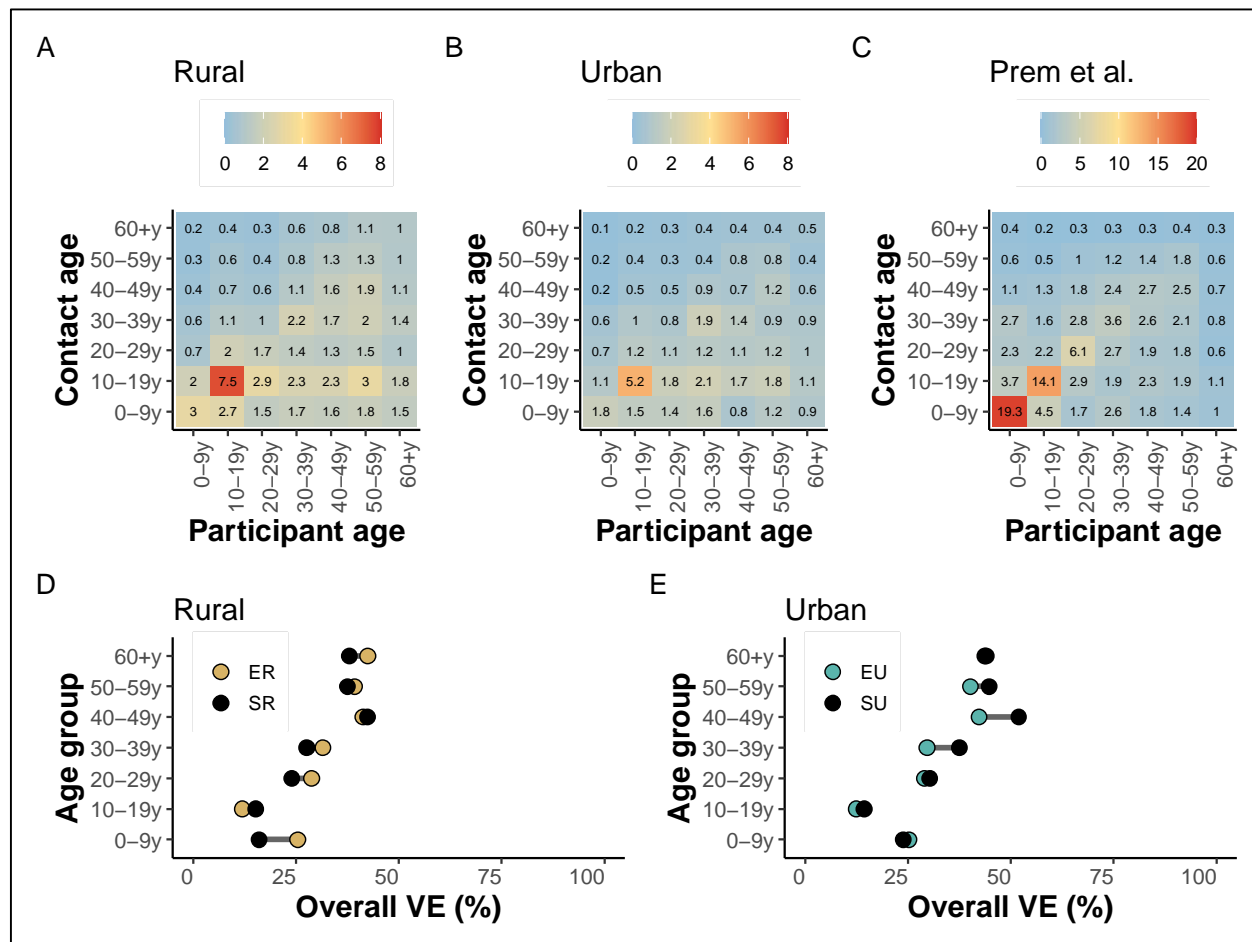
Place visited	Rural		Urban	
	N (%)	Median (IQR)	N (%)	Median (IQR)
Other home	752 (48)	4 (3-6)	261 (29)	4 (3-7)
Street	369 (23)	2 (1-5)	342 (37)	4 (3-9)
Market/Shop	92 (6)	6 (3-12)	77 (8)	10 (3-30)
Transport/Hub	90 (5)	16 (11-18)	60 (6)	18 (12-20)
Agricultural field	104 (7)	3 (1.8-6.3)	8 (1)	5.5 (1-6.8)
School	62 (3)	25 (15-25.8)	63 (6)	30 (24-33.5)
Work	45 (2)	6 (4-20)	54 (6)	6 (3-10)
Place of worship	26 (2)	20 (10.3-30)	23 (2)	30 (17.5-45)
Well	17 (1)	4 (2-6)	1	7
Playground	1	15	1	13
Other	45 (3)	10 (5-23)	47	7 (3.5-18)

259

260 *4.2.3 Sensitivity of transmission model disease dynamics to our empirical contact matrices*

261 To recap, we restructured the age-specific matrices presented in **Figure 1** D and E into 7 age
 262 groups (see Methods) depicted in **Figure 2** panels A and B. Assortative contacts in empirical
 263 data were highest among 10-19-year-olds (higher in rural compared to urban) compared to
 264 synthetic values for 0-9-year-olds (panel C, data available from [12]).

265 In the rural and urban site, the values of the empirical overall vaccine effects (VE) overlapped in
 266 most ages. However, synthetic VE were marginally higher in 10–19 and 40–49-year-olds in the
 267 rural site (**Figure 2 D**), and higher in all ages (particularly adults aged 30–59-year-olds) apart
 268 from children aged 0–9-year old (**Figure 2 E**) in the urban site. Specifically, we report higher
 269 attack rates in unvaccinated (AR) compared to vaccinated (ARv) individuals (AR: 94%, ARv:
 270 77%) and lower overall vaccine effects (18%) for 0–9-year-olds compared to attack rates using
 271 empirical data (rural AR: 84%, rural ARv, 62%, rural vaccine effects: 26%; urban AR: 84%,
 272 urban ARv: 62%, urban vaccine effects: 26%). Additionally, synthetically contacts among 60+
 273 years old were underestimated compared to empirical values, producing notably lower attack
 274 rates among this age group (Synthetic AR: 49%, Synthetic ARv: 31%; Rural AR: 40%, Rural
 275 ARv: 24%; Urban AR: 36%, Urban ARv: 21%).



276
 277 **Figure 2. Contact matrices and modeled age-specific vaccine effects (VE).** Panels A and B
 278 show rural and urban contact matrices from our empirical data, respectively, while panel C
 279 shows the synthetic contact matrix derived from Mozambique-specific demographic data by
 280 Prem et al [12]. Panels D and E shows the overall vaccine effects (VE) of a respiratory infection
 281 comparing synthetic and empirical contact rates in the rural and urban site, respectively. Legend
 282 key: ER=Empirical Rural, SR=Synthetic Rural, EU=Empirical Urban, SU=Synthetic Urban.

283 5. Discussion

284 We present results from a two-day cross-sectional study aiming to quantify social contact rates
285 among residents of a rural and urban site in Mozambique during the COVID-19 pandemic. We
286 first engaged with the local community to get their views on the suitability and acceptability of
287 our tools and study procedures. We made several key observations. First, we used the qualitative
288 outcomes to modify the format and content of the paper diaries to make them more user friendly
289 (see additional details in **Supplementary text 1**). Second, in our sample population, participants
290 from the rural site had significantly higher average number of contacts compared to the urban
291 site. Third, in each site, reported mean contacts increased with age to peak at school-going
292 children/ teenagers aged 15-19 years, and mean contacts were higher among adults (>18 years)
293 compared to children <5 years. Fourth, mixing was assortative (increased frequency of contacts
294 within the same ages) among school-going children, and less pronounced inter-generational
295 mixing particularly in the urban site. Finally, we quantify the importance of using empirical
296 contact data to model disease dynamics and interventions. In model simulations of an emerging
297 respiratory pathogen, we find meaningfully different attack rates and vaccine impact among both
298 child and elderly groups when using our local data compared with widely used contact matrices
299 modelled from other settings.

300 In the earlier phases of the COVID-19 pandemic, Mozambique adopted physical distancing
301 policies including school closures and restrictions on social gatherings, but with no express
302 requirement to stay at home [23]. This was similar to measures implemented globally to reduce
303 transmission of SARS-CoV-2. Few countries in sub-Saharan Africa e.g., South Africa [24],
304 Kenya [3,5], Zimbabwe [4], Uganda [9], and Somaliland [8] had existing empirical social
305 contact data collected in various settings prior to 2020. There are sparse data on contact patterns
306 during the pandemic from LMICs (examples are from Kenya [7] and Malawi [6]) compared to
307 high income countries such as the United Kingdom and Europe [17] and USA [16]. Longitudinal
308 data during the pandemic would have been critical for a better understanding of the transmission
309 pathways of the novel virus in LMICs taking account of demographic, social and economic
310 contexts, and provide insights on how to enhance NPI use in place and identify priority groups
311 for immunization once vaccines became available albeit in limited supply and use in LMICs.

312 Compared to pandemic periods in Kenya [7] and Malawi [6], we observed lower mean number
313 of contacts in Mozambique but a higher number of contacts reported by participants in the rural
314 compared to the urban site. However, we interpret this with care since data from Kenya and
315 Malawi were collected from high-density settlements where individuals may have been unable to
316 fully adhere to physical distancing mainly due to economic reasons. The Government of
317 Mozambique periodically revised physical distancing policies to curb the spread of SARS-CoV-
318 2 [25], and we also think that the measures were not strictly adhered to particularly among
319 school-going children and working adults.

320 We implemented several innovations in the GlobalMix study. First, we collected contact data
321 from participants over two consecutive days. The average number of contacts stratified across
322 different covariates of interest remained constant over two days, suggesting the stability of
323 participant's recall as well as the number and nature of contacts made over multiple days. The
324 stability of contact networks across days has been previously demonstrated in rural households in
325 Kenya and Malawi using autonomous methods that minimize recall bias [10,11]. A second
326 innovation of our study was an estimation of group proximity contacts at locations frequently

327 visited by participants. It is important to note that participants report almost four times the
328 number of proximity contacts compared to detailed individually reported contacts. This suggests
329 the potential to substantially underestimate the number of interactions that could lead to
330 transmission events particularly in highly mobile age groups congregating in social venues.

331 Finally, our transmission model simulation demonstrates the importance of contextual empirical
332 social contact data. Although advanced methods for projecting social contact patterns onto
333 regions without data exist (e.g., [12]), we found that age-specific infection attack rates from a
334 model using empirical contact data differed meaningfully compared to a model parameterized
335 with synthetic contact rates. Importantly, we found the largest differences in attack rates
336 (comparing vaccinated versus unvaccinated individuals) resulted in increased vaccine impact in
337 the youngest (0-9 years) age group, who often represent the most vulnerable group. These
338 findings were consistent with a Uganda model where use of local contact pattern data resulted in
339 larger epidemics in young children and smaller epidemics in adults aged >35 years compared to
340 using UK-based contact data [9]. It is also notable that we observed distinct contact patterns
341 resulting in divergent model results for the Mozambique rural and urban sites. This highlights the
342 impact of sub-national differences in contact patterns resulting in variation in disease dynamics.
343 Such insights are not possible with widely used contact data which are generally available at the
344 national level.

345 **6. Conclusion**

346 We present empirical results of a cross-sectional study quantifying rates and patterns of human
347 social contacts relevant for the spread of directly transmitted infections in rural and urban areas
348 of Mozambique. We demonstrate the possibility of collecting high quality social contact data
349 from resource-poor settings thus reducing reliance on synthetic data modelled from high-income
350 countries. Finally, we demonstrate the potential advantages of empirical compared to synthetic
351 data in a transmission and vaccine control model and advocate for the use of contextual data in
352 similar studies. Nevertheless, some questions remain including if social contact patterns changed
353 in this setting as non-pharmaceutical interventions were relaxed. Finally, we are completing data
354 collection from three other LMICs and we will make all our data collection tools, data, and
355 analysis scripts findable, accessible, interoperable, and reusable (FAIR).

356 **7. References**

- 357 1. Sidat M, Capatine I. Infecção por SARS-CoV-2 em Moçambique: a epidemiologia e os
358 avanços alcançados com a vacinação contra a COVID-19. *Anais do Instituto de Higiene e*
359 *Medicina Tropical*. 2022; 90-98 Páginas. doi:10.25761/ANAISIHMT.432
- 360 2. Hoang T, Coletti P, Melegaro A, Wallinga J, Grijalva CG, Edmunds JW, et al. A Systematic
361 Review of Social Contact Surveys to Inform Transmission Models of Close-contact
362 Infections. *Epidemiology*. 2019;30: 723–736. doi:10.1097/EDE.0000000000001047
- 363 3. Kiti MC, Kinyanjui TM, Koech DC, Munywoki PK, Medley GF, Nokes DJ. Quantifying
364 age-related rates of social contact using diaries in a rural coastal population of Kenya. *PLoS*
365 *ONE*. 2014;9. doi:10.1371/journal.pone.0104786

- 366 4. Melegaro A, Fava ED, Poletti P, Merler S, Nyamukapa C, Williams J, et al. Social contact
367 structures and time use patterns in the manicaland province of Zimbabwe. *PLoS ONE*.
368 2017;12: 1–17. doi:10.1371/journal.pone.0170459
- 369 5. Del Fava E, Adema I, Kiti MC, Poletti P, Merler S, Nokes DJ, et al. Individual’s daily
370 behaviour and intergenerational mixing in different social contexts of Kenya. *Scientific*
371 *Reports*. 2021;11. doi:10.1038/s41598-021-00799-1
- 372 6. Thindwa D, Jambo KC, Ojal J, MacPherson P, Phiri MD, Khundi M, et al. Social mixing
373 patterns relevant to infectious diseases spread by close contact in urban Blantyre, Malawi.
374 *Epidemiology*; 2021 Dec. doi:10.1101/2021.12.16.21267959
- 375 7. Quaife M, Van Zandvoort K, Gimma A, Shah K, McCreesh N, Prem K, et al. The impact of
376 COVID-19 control measures on social contacts and transmission in Kenyan informal
377 settlements. *BMC Medicine*. 2020;18. doi:10.1186/s12916-020-01779-4
- 378 8. van Zandvoort K, Bobe MO, Hassan AI, Abdi MI, Ahmed MS, Soleman SM, et al. Social
379 contacts and other risk factors for respiratory infections among internally displaced people in
380 Somaliland. *Epidemics*. 2022;41: 100625. doi:10.1016/j.epidem.2022.100625
- 381 9. le Polain de Waroux O, Cohuet S, Ndazima D, Kucharski AJ, Juan-Giner A, Flasche S, et al.
382 Characteristics of human encounters and social mixing patterns relevant to infectious
383 diseases spread by close contact: A survey in Southwest Uganda. *BMC Infectious Diseases*.
384 2018;18: 172. doi:10.1186/s12879-018-3073-1
- 385 10. Kiti MC, Tizzoni M, Kinyanjui TM, Koech DC, Munywoki PK, Meriac M, et al.
386 Quantifying social contacts in a household setting of rural Kenya using wearable proximity
387 sensors. *EPJ Data Science*. 2016;5: 21. doi:10.1140/epjds/s13688-016-0084-2
- 388 11. Ozella L, Paolotti D, Lichand G, Rodríguez JP, Haenni S, Phuka J, et al. Using wearable
389 proximity sensors to characterize social contact patterns in a village of rural Malawi. *EPJ*
390 *Data Science*. 2021;10. doi:10.1140/epjds/s13688-021-00302-w
- 391 12. Prem K, van Zandvoort K, Klepac P, Eggo RM, Davies NG, Cook AR, et al. Projecting
392 contact matrices in 177 geographical regions: An update and comparison with empirical data
393 for the COVID-19 era. *PLoS Computational Biology*. 2021;17: 1DUMMUY.
394 doi:10.1371/journal.pcbi.1009098
- 395 13. Liu CY, Berlin J, Kiti MC, Del Fava E, Grow A, Zagheni E, et al. Rapid Review of Social
396 Contact Patterns During the COVID-19 Pandemic. *Epidemiology*. 2021;32: 781–791.
397 doi:10.1097/ede.0000000000001412
- 398 14. Kiti MC, Aguolu OG, Liu CY, Mesa AR, Regina R, Woody M, et al. Social contact patterns
399 among employees in 3 U.S. companies during early phases of the COVID-19 pandemic,
400 April to June 2020. *Epidemics*. 2021;36: 100481. doi:10.1016/j.epidem.2021.100481
- 401 15. Gimma A, Munday JD, Wong KLM, Coletti P, van Zandvoort K, Prem K, et al. CoMix:
402 Changes in social contacts as measured by the contact survey during the COVID-19

- 403 pandemic in England between March 2020 and March 2021. medRxiv. 2021;08:
404 2021.05.28.21257973. Available:
405 <http://medrxiv.org/content/early/2021/05/30/2021.05.28.21257973.abstract>
- 406 16. Nelson KN, Siegler AJ, Sullivan PS, Bradley H, Hall E, Luisi N, et al. Nationally
407 representative social contact patterns among U.S. adults, August 2020-April 2021.
408 *Epidemics*. 2022;40: 100605. doi:10.1016/j.epidem.2022.100605
- 409 17. Wong KL, Gimma A, Coletti P, CoMix Europe Working Group, Faes C, Beutels P, et al.
410 Social contact patterns during the COVID-19 pandemic in 21 European countries – evidence
411 from a two-year study. *Epidemiology*; 2022 Jul. doi:10.1101/2022.07.25.22277998
- 412 18. Aguolu OG, Kiti MC, Nelson K, Liu CY, Sundaram M, Gramacho S, et al. Comprehensive
413 profiling of social mixing patterns in resource poor countries: a mixed methods research
414 protocol. medRxiv; 2023. p. 2023.12.05.23299472. doi:10.1101/2023.12.05.23299472
- 415 19. Martínez-Martínez FJ, Massinga AJ, De Jesus Á, Ernesto RM, Cano-Jiménez P, Chiner-Oms
416 Á, et al. Tracking SARS-CoV-2 introductions in Mozambique using pandemic-scale
417 phylogenies: a retrospective observational study. *The Lancet Global Health*. 2023;11: e933–
418 e941. doi:10.1016/S2214-109X(23)00169-9
- 419 20. Nhacolo A, Jamisse E, Augusto O, Matsena T, Hunguana A, Mandomando I, et al. Cohort
420 Profile Update: Manhiça Health and Demographic Surveillance System (HDSS) of the
421 Manhiça Health Research Centre (CISM). *International Journal of Epidemiology*. 2021;50:
422 395–395. doi:10.1093/ije/dyaa218
- 423 21. Sistema de vigilância demográfica e de saúde (HDSS)-Polana Caniço. 2019. Available:
424 [https://ins.gov.mz/wp-content/uploads/2020/10/HDSS-Polana-Caniço_Relatório-Avaliação-](https://ins.gov.mz/wp-content/uploads/2020/10/HDSS-Polana-Caniço_Relatório-Avaliação-de-base-19102020.pdf)
425 [de-base-19102020.pdf](https://ins.gov.mz/wp-content/uploads/2020/10/HDSS-Polana-Caniço_Relatório-Avaliação-de-base-19102020.pdf)
- 426 22. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O’Neal L, et al. The REDCap
427 consortium: Building an international community of software platform partners. *J Biomed*
428 *Inform*. 2019/05/09 ed. 2019;95: 103208–103208. doi:10.1016/j.jbi.2019.103208
- 429 23. Lane J, Means AR, Bardosh K, Shapoval A, Vio F, Anderson C, et al. Comparing COVID-
430 19 physical distancing policies: results from a physical distancing intensity coding
431 framework for Botswana, India, Jamaica, Mozambique, Namibia, Ukraine, and the United
432 States. *Global Health*. 2021;17: 124. doi:10.1186/s12992-021-00770-9
- 433 24. Johnstone-Robertson SP, Mark D, Morrow C, Middelkoop K, Chiswell M, Aquino LDH, et
434 al. Social mixing patterns within a South African township community: implications for
435 respiratory disease transmission and control. *American Journal of Epidemiology*. 2011;174:
436 1246–55. doi:10.1093/aje/kwr251
- 437 25. Júnior A, Dula J, Mahumane S, Koole O, Enosse S, Fodjo JNS, et al. Adherence to COVID-
438 19 Preventive Measures in Mozambique: Two Consecutive Online Surveys. *IJERPH*.
439 2021;18: 1091. doi:10.3390/ijerph18031091

440 **8. Acknowledgements**

441 We acknowledge all the community members and local authorities who have supported and/or
442 participated in data collection. We also thank all the researchers and field staff involved in this
443 work, both from Manhiça and Polana Caniço HDSS. Our special thanks are extended to the
444 National Institute of Health of Mozambique (INS) and the CISPOC Directorate, who willingly
445 participated in the research by including the urban site of the Polana-Caniço HDSS.

446 **9. Funding**

447 This work was supported by the National Institutes of Health (NIH) grant number: R01
448 HD097175-01. The views expressed in this publication are those of the author(s) and not
449 necessarily those of NIH or the US government. The funders had no role in study design, data
450 collection and analysis, decision to publish, or preparation of the manuscript. CISM is supported
451 by the Government of Mozambique and the Spanish Agency for International Development
452 (AECID).

453 **10. Availability of data and materials**

454 All data produced in the present study are available on our GitHub repository:
455 <https://github.com/lopmanlab/globalmix-mozambique-aim1>. We also include the survey
456 codebooks in English and Portuguese, as well as the Portuguese version of the diary to portray
457 the age group images used in the study.

458 **11. Authors contribution**

Contributor Role	Role Definition	Author
Conceptualization	Ideas; formulation or evolution of overarching research goals and aims.	Charfudin Sacoor, Azucena Bardají, Moses C. Kiti, Obianuju G. Aguolu, Fauzia Malik, Saad B. Omer, Benjamin Lopman,
Funding Acquisition	Acquisition of the financial support for the project leading to this publication.	Saad B. Omer, Benjamin Lopman
Data Curation	Management activities to annotate (produce metadata), scrub data and maintain research data (including software code, where it is necessary for interpreting the data itself) for initial use and later reuse.	Moses C. Kiti, Holin Chen, Sara Kim, Carol Liu
Formal Analysis	Application of statistical, mathematical, computational, or other formal techniques to analyze or synthesize study	Moses C. Kiti, Holin Chen, Sara Kim, Carol Liu, Charfudin Sacoor

	data.	
Investigation	Conducting a research and investigation process, specifically performing the experiments, or data/evidence collection.	Moses C. Kiti, Alana Zelaya, Obianuju G. Aguolu, Saad B. Omer, Benjamin Lopman, Edgar Jamise, Americo Jose, Charfudin Sacoor, Ivalda Macicame, Felizarda Amosse, Nilzio Cavele, Corssino Tchavana, Migdalia Uamba
Methodology	Development or design of methodology; creation of models.	Moses C. Kiti, Charfudin Sacoor, Azucena Bardají, Alana Zelaya, Obianuju G. Aguolu, Inci Yildirim, Ivalda Macicame, Felizarda Amosse, Nilzio Cavele, Corssino Tchavana, Migdalia Uamba, Edgar Jamise, Americo Jose, Saad B. Omer, Benjamin Lopman, Fauzia Malik
Project Administration	Management and coordination responsibility for the research activity planning and execution.	Moses C. Kiti, Charfudin Sacoor, Noureen Ahmed, Alana Zelaya, Obianuju G. Aguolu, Saad B. Omer, Benjamin Lopman, Fauzia Malik
Resources	Provision of study materials, reagents, materials, patients, laboratory samples, animals, instrumentation, computing resources, or other analysis tools.	Moses C. Kiti, Charfudin Sacoor, Nilzio Cavele, Corssino Tchavana, Edgar Jamise, Americo Jose, Sara Kim, Carol Liu
Software	Programming, software development; designing computer programs; implementation of the computer code and supporting algorithms; testing of existing code components.	Moses C. Kiti, Sara Kim, Carol Liu, Corssino Tchavana, Nilzio Cavele, Obianuju G. Aguolu, Fauzia Malik, Saad B. Omer, Benjamin Lopman
Supervision	Oversight and leadership responsibility for the research activity planning and execution, including mentorship external to the core team.	Moses C. Kiti, Charfudin Sacoor, Azucena Bardají, Nilzio Cavele, Corssino Tchavana, Edgar Jamise, Americo Jose, Obianuju G. Aguolu, Saad B. Omer, Benjamin Lopman, Dr Fauzia Malik
Validation	Verification, whether as a part of the activity or separate, of the overall replication /reproducibility of results/experiments and other research outputs.	Moses C. Kiti, Obianuju G. Aguolu, Azucena Bardají, Fauzia Malik, Noureen Ahmed, Saad B. Omer, Benjamin Lopman,

Visualization	Preparation, creation and/or presentation of the published work, specifically visualization/data presentation.	Moses C. Kiti, Charfudin Sacoor, Corssino Tchavana, Orvalho Augusto, Sara Kim, Carol Liu
Writing – Original Draft Preparation	Creation and/or presentation of the published work, specifically writing the initial draft (including substantive translation).	Moses C. Kiti, Charfudin Sacoor, Carol Liu, Sara Kim
Writing – Review & Editing	Preparation, creation and/or presentation of the published work by those from the original research group, specifically critical review, commentary, or revision – including pre- or post-publication stages.	Moses C. Kiti, Charfudin Sacoor, Orvalho Augusto, Azucena Bardají, Obianuju G. Aguolu, Inci Yildirim, Noureen Ahmed, Carol Liu, Sara Kim, Fauzia Malik, Saad B. Omer, Benjamin Lopman

459

460 **12. Competing interests**

461 The authors declare they have no competing interest.

462

463 **13. Supplementary Information**

464 **Supplementary text 1:** Details of fieldwork, challenges experienced, and practical solutions
465 applied.

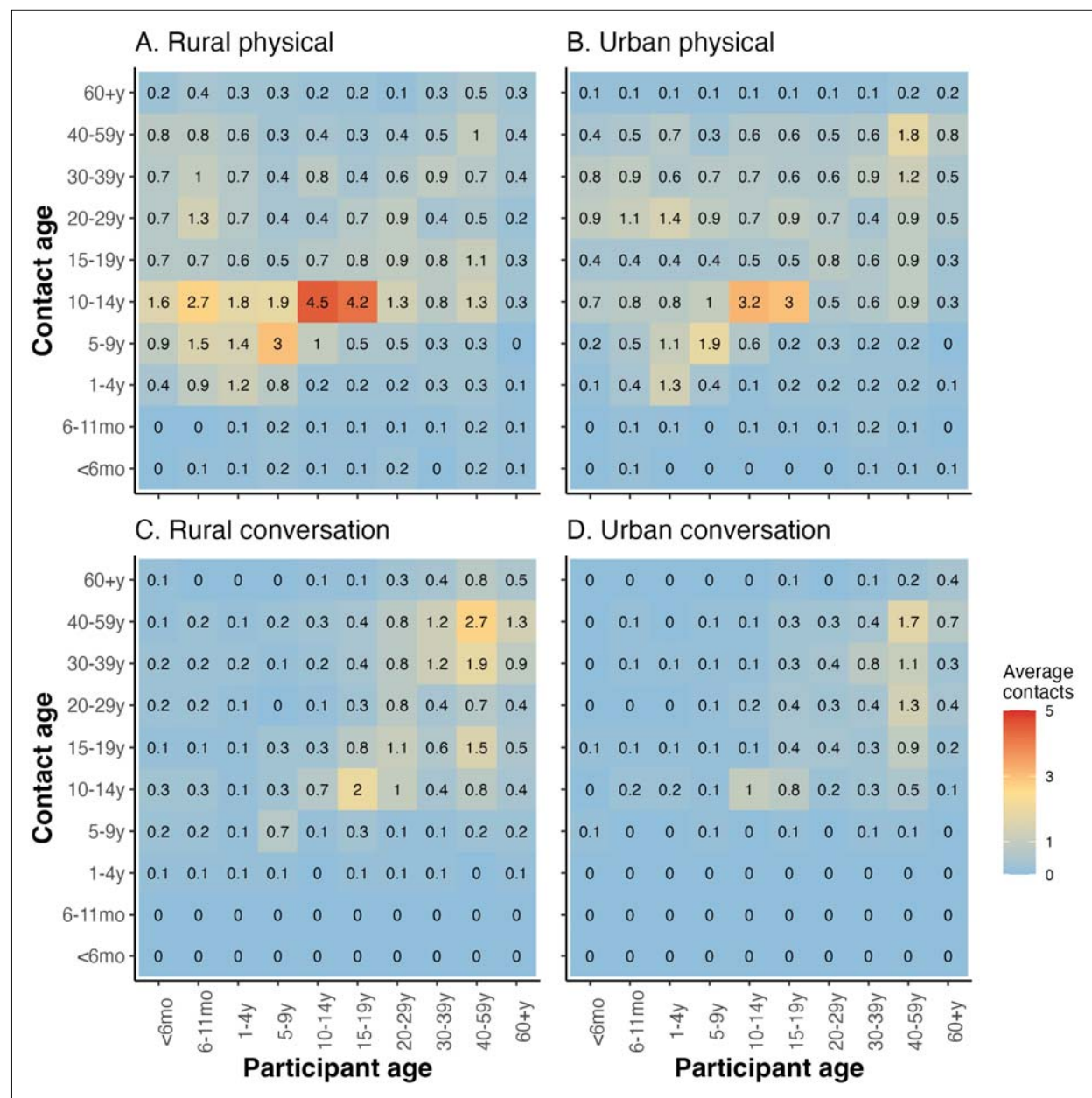
466 This was the first time that self-report paper diaries were being used in the Manhiça and Polana
467 Canhiço HDSS. It is important to note the logistic and technical challenges we experienced
468 during the conduct of our study, and our mitigation measures, to better inform future studies in
469 similar settings. Following extensive consultations with community members, we used paper
470 diaries with contextualized questions and customized visual aids to guide age selection of
471 contacts, including those met for the first time. During implementation, adult males were hard to
472 find since most left their houses early for work, market sales and to tend to their farms.
473 Absenteeism was also reported in a similar study in Kenya [3] and the authors replaced the loss-
474 to-follow-up individuals with others of similar demographic characteristics. Due to the
475 pandemic, both HDSS databases had not been updated and we resorted to door-to-door
476 recruitment for under-5-year-olds and later expanding the study catchment area to additional
477 HDSS blocks in the urban site. During the rainy season, remote areas in the rural site were
478 inaccessible thus hampering recruitment and data collection exercises. Once recruited, some
479 participants were unavailable to present their diary data during the exit visits, and almost half of
480 diaries in both sites were partly filled in, thus requiring assistance from our field staff to
481 complete the diary. Eventually, we resorted to recruiting children from mothers attending health
482 clinics, visits to workplaces for adult males, and following up participants through phone calls.
483 The lapse in time between diary data collection period and exit survey could have contributed to
484 recall bias and it was not possible to ascertain the magnitude and direction of the recall bias.
485 Despite this, we had a good buy in from the community due to our extensive community
486 engagement conducted before and during the study, which also resulted in higher participation
487 rates in general than previously observed in other sub-Saharan African settings. Furthermore, the
488 data was collected during periods of active transmission of SARS-CoV-2 in the community,
489 including certain times when certain restrictive measures were in place, such as the closure of
490 schools and parks, limitations on the number of employees at the workplace and passengers on
491 public transportation, social distancing, and other measures [23]. This context likely caused some
492 disruption, given the contingency scenario at the time, and may have influenced the
493 characteristics of social contact patterns due to the ongoing pandemic.

494

495 **Supplementary table 1: Median (IQR) rural contacts on day 2**

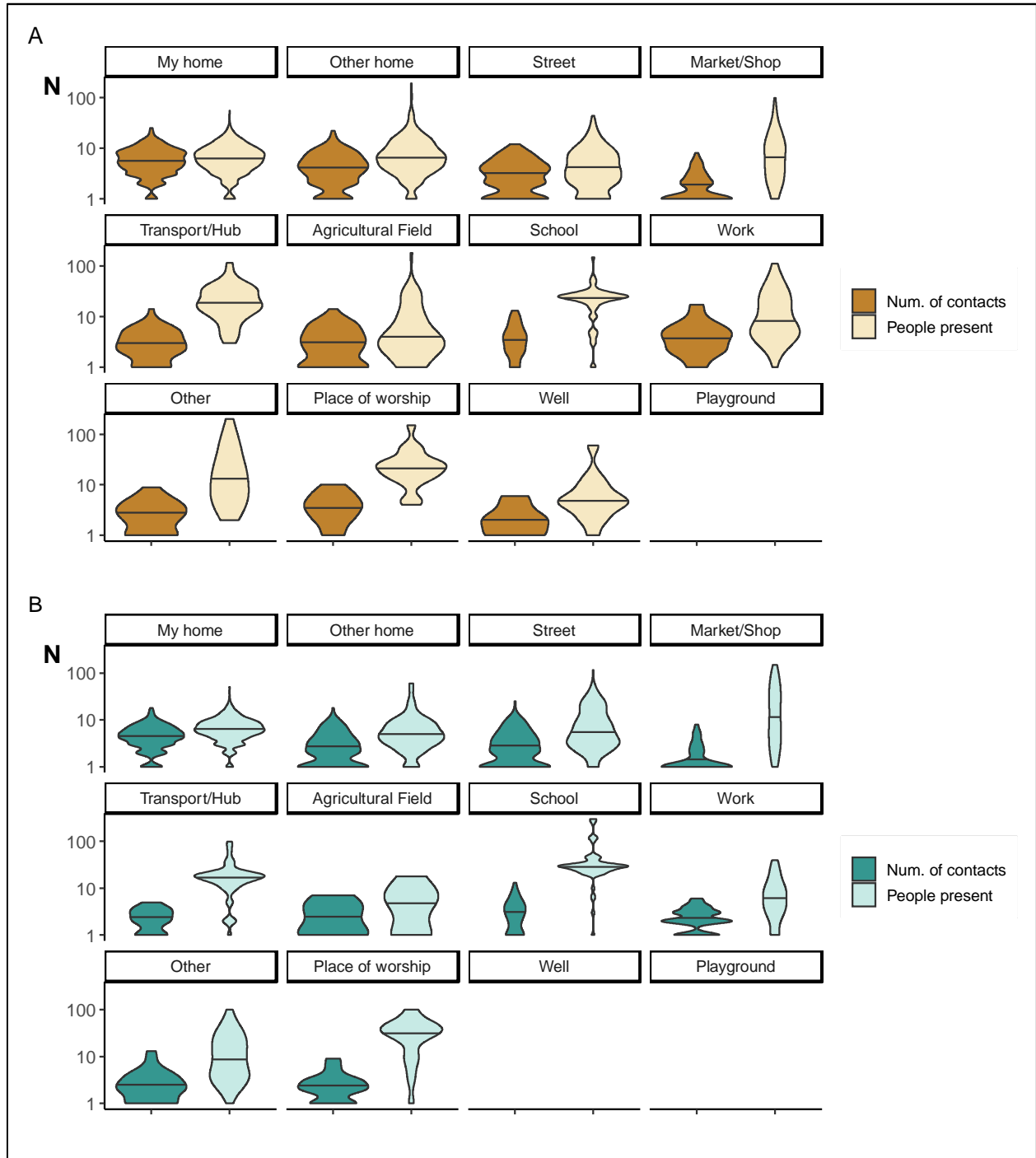
	Rural	Urban
Sex		
Female	5 (3–9)	5 (3–8)
Male	8 (5–11)	6 (4–8)
Age		
<6mo	5 (3–7)	2 (1–4)
6–11mo	5 (3–9)	3.5 (2–5.8)
1–4y	5 (3–8)	5 (3–7)
5–9y	6 (4–9)	4 (2–5)
10–14y	6 (3–9)	4 (3–7)
15–19y	7 (4–12)	6 (3–7)
20–29y	6 (4–9)	4 (3–5)
30–39y	5 (3–10)	4 (2.8–6)
40–59y	5 (3–9)	3 (2–4)
60+y	4 (2–7)	3 (1.3–4.8)
Occupation/ daily activity		
Child	5 (3–8)	3 (2–5)
Unemployed	4.5 (2–8)	3.5 (2–5.3)
Student	6 (4–10)	4 (3–7)
Homemaker	3 (2–6.5)	2 (2–3)
Casual laborer	6.5 (2.8–10.3)	3.5 (2–4.8)
Farmer	5 (2–7)	9 (5–9)
Businessperson	9 (4–9.5)	3 (2–4)
Office worker	5 (3.5–11)	4 (2–5)
Retired	7 (5.8–7)	1.5 (1–2.8)
Other	5.5 (4–10)	5 (4–6)
Household size		
1	6 (2–10)	5.5 (4–8)
2–3	5.5 (3–9)	4 (2–6)
4–6	5 (3–8.8)	4 (2–5)
7–10	5 (3–7.8)	3 (2–5)
10+	5 (3–6.8)	2 (2–5)
Enrolled in school		
Yes	6 (3–10)	4 (2–7)
No	5 (3–9)	3 (2–5)
Weekday/Weekend		
Weekday	5 (3–9)	4 (2–6)
Weekend	5 (3–9)	4 (2–5)
ARI symptoms		
No symptom	5 (3–9)	4 (2–6)
>1 symptom	6 (4–9)	4 (3–6)
AGE symptoms		
Yes	7 (3–12)	3 (2–5)
No	5 (3–9)	4 (2–6)

497 **Supplementary figure 1.** Matrices of age-specific average physical only and conversation only
 498 contacts by site. The panels are arranged by rural (A) and urban (B) physical contacts and rural
 499 (C) and urban (D) conversation contacts. Average age-specific values are generated by dividing
 500 the total number of contacts reported by age group pair with the number of participants in that
 501 age group.



502
503

504 **Supplementary figure 2.** Violin plots comparing the distribution of number of contacts versus
505 number of co-located individuals reported by participants at each location in rural (A) versus
506 urban (B) site.



507