

1 **TITLE PAGE**

2 Title: Characteristics of human encounters and social mixing patterns relevant to infectious
3 diseases spread by close contact: A survey in southwest Uganda

4 Short title: Social mixing patterns in Uganda

5 Authors: O le Polain de Waroux¹, S Cohuet², D Ndazima³, A J Kucharski¹, A Juan-Giner², S
6 Flasche¹, E Tumwesigye⁵, R. Arinaitwe³, J Mwanga-Amumpaire^{3,4}, Y Boum Il³, F Nackers⁶, F
7 Checchi¹, R F Grais², W J Edmunds¹

8 Affiliations

- 9 1. Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical
10 Medicine, London, United Kingdom
- 11 2. Epicentre, Paris, France
- 12 3. Epicentre, Uganda Research Centre, Mbarara, Uganda
- 13 4. Mbarara University Of Science and Technology (MUST), Mbarara, Uganda
- 14 5. Kabwohe Medical Research Centre, Kabwohe, Uganda
- 15 6. Epicentre, Brussels, Belgium

16 Corresponding author: Olivier.lepolain@gmail.com

17

18 Keywords: Social mixing, contact, transmission, spatial, infection, Africa

19

20 **ABSTRACT**

21 Quantification of human interactions relevant to infectious disease transmission through
22 social contact is central to predict disease dynamics, yet data from low-resource settings
23 remain scarce. We undertook a social contact survey in rural Uganda, whereby participants
24 were asked to recall details about the frequency, type, and socio-demographic
25 characteristics of any conversational encounter that lasted for ≥ 5 minutes (henceforth
26 defined as 'contacts') during the previous day. An estimate of the number of 'casual
27 contacts' (i.e. < 5 minutes) was also obtained. A total of 568 individuals were included. On
28 average participants reported having routine contact with 7.2 individuals (range 1-25).
29 Children aged 5-14 years had the highest frequency of contacts and the elderly (≥ 65 years)
30 the fewest ($P < 0.001$). A strong age-assortative pattern was seen, particularly outside the
31 household and increasingly so for contacts occurring further away from home. Adults aged
32 25-64 years tended to travel more and further than others, and males travelled more
33 frequently than females. Our study provides detailed information on contact patterns and
34 their spatial characteristics in an African setting. It therefore fills an important knowledge
35 gap that will help more accurately predict transmission dynamics and the impact of control
36 strategies in such areas.

37

38

39

40

41 **INTRODUCTION**

42 Quantification of human interactions relevant to the spread of these infectious diseases is
43 essential to accurately predict their infection dynamics and the impact of control strategies
44 [1, 2].

45 Detailed surveys of social mixing patterns have now been undertaken in a number of
46 settings [2-13]. Studies have shown that people tend to mix with other individuals of their
47 own age (i.e. assortative mixing); however, the frequency of contact, the degree of
48 intergenerational mixing and the characteristics of mixing tend to vary between settings,
49 depending on factors such as household size, population density and local activities, among
50 others [3-11].

51 Data from low-resource settings remain scarce, with only three studies in Africa published
52 to date [10, 12, 13], and none from Uganda.

53 With the exception of a recent study from China [11], the spatial dispersal of social contacts
54 relevant for transmission has often been overlooked, and there is – to our knowledge – no
55 published information from low-income settings. Spatial mobility is particularly important
56 for epidemic risk prediction of novel and re-emergent diseases, and for the optimization of
57 routine control programmes [14].

58 To address this knowledge gap, we set up a study of social contacts relevant to the spread of
59 infections transmitted through the respiratory route or by close contact, in rural southwest
60 Uganda.

61

62 MATERIALS AND METHODS

63 The study was conducted in four sub-counties of Sheema North Sub-District (southwest
64 Uganda), an area with a total of about 80,000 inhabitants. About half (49%) of the district's
65 population is <15 years. The area is primarily rural.

66 Study design

67 We conducted a two-stage age-stratified community-based study in Sheema North Sub-
68 district between January and March 2014 on a subset of individuals included in a
69 population-based survey of nasopharyngeal carriage of *Streptococcus pneumoniae* (Nackers
70 F et al. manuscript in preparation). The sample size calculations by age group for the
71 nasopharyngeal carriage study included 538 children <2 years, 323 children aged 2 – 4 years,
72 583 aged 5 – 14 years and 327 individuals aged ≥ 15 years. Using the same age groups, but
73 different inclusion probabilities, we estimated that, based on previous results [12, 15],
74 including all 327 individuals ≥ 15 years and a subset of 90 children <2 years, 90 children aged
75 2-4 years and 180 children aged 5 – 14 years, for a total target sample size of 687, would
76 provide a precision of just over 1 contact on the mean number of contacts per day, and
77 enable detection of a 20% difference in the average number of daily contacts by age group,
78 accounting for 10% non-response.

79 Individuals were selected from 60 clusters randomly sampled from the exhaustive list of 215
80 villages and two small towns in the sub-county, with an inclusion probability proportional to
81 the size of the village or town. Within each cluster 11 or 12 households were randomly
82 selected and in each household only one individual was selected for inclusion in the study. A
83 household was defined as a group of individuals living under the same roof and sharing the

84 same kitchen on a daily basis. One individual from each household was randomly selected
85 from within a predefined age group based on a random sequence of age groups according
86 to the age group sampling quota by cluster.

87 When nobody in the household was from that age group, either someone from another age
88 group was selected providing that the quota for that age group had not been reached in the
89 cluster, or the closest neighbouring household was visited instead. In case of non-response,
90 another attempt was made later in the day or the following Saturday. Survey teams had a
91 day off on Thursdays and Sundays.

92 **Data collection**

93 Informed consent was sought for individuals aged > 13 years, and consent was sought from
94 a parent or guardian otherwise.

95 Ethical approval was obtained from the Ethical review boards of Médecins Sans Frontières
96 (MSF), the Faculty of Medicine Research & Ethics Committee of the Mbarara University of
97 Science and Technology (MUST), the Institutional Ethical Review Board of the MUST, the
98 Uganda National Council for Science and Technology (UNCST) and the London School of
99 Hygiene and Tropical Medicine (LSHTM).

100 Participants were asked to recall information on the frequency, type and duration of social
101 encounters from the time they woke up the day before the survey until when they woke up
102 on the survey day (~ 24 hours).

103 We defined contacts as individuals with whom there was at least one two-way
104 conversational encounter (three or more words) lasting for ≥ 5 minutes. Participants were

105 first asked to list all the places they had visited in the previous 24 hours, the number of
106 people they had contact with, their relationship with each individual mentioned, the age (or
107 estimated age) of each listed contact and how long the encounter lasted for. Contacts
108 involving skin-to-skin touch or sharing utensils passed directly from mouth-to-mouth were
109 defined as ‘physical’ contacts.

110 We defined as ‘casual contacts’ short conversational encounters lasting less than 5 minutes.
111 Participants were only asked to estimate the number of casual contacts they had, based on
112 pre-defined categories (<10, 10-19, 20-29, ≥30), but were not asked to provide detailed
113 information about the nature of the encounter or the socio-demographic characteristics of
114 the person met. Casual contacts are generally inaccurately reported in social contact surveys
115 [7], particularly in a retrospective design, and most contacts important for the transmission
116 of respiratory infections are believed to be close rather than casual [6].

117 The questionnaire was designed in English, translated to Ruyankole, the local language, and
118 back-translated to English for consistency (Supporting Information Text S1). For children <5
119 years, parents were asked about their child’s encounters and whereabouts. Children aged 5
120 – 14 years were interviewed directly, using an age-appropriate questionnaire.

121 Geographical coordinates from each participant’s household and of the centre point of each
122 village were taken using handheld GPS devices.

123 Questionnaires completed in the field were double entered on a preformatted data entry
124 tool by two data managers working independently. Data entry conflicts were identified
125 automatically and resolved as the data entry progressed.

126 **Analysis**

127 Characteristics of social contacts by time, person and place

128 We analysed the frequency distribution of contacts for a set of covariates, including age,
129 sex, and occupation, day of the week, distance travelled, and type of contact. Encounters
130 reported with the same individual more than once counted as one contact only. Distance
131 travelled was measured as straight line distances between the centre point of the
132 participants' home village/town and that of the village/town where each reported
133 encounter took place.

134 We used negative binomial regression to estimate the ratio of the mean contacts per person
135 as a function of the different covariates of interest. Negative binomial was preferred over
136 Poisson regression given evidence of over-dispersion (variance > mean, and likelihood ratio
137 significant ($P < 0.05$) for the over-dispersion parameter). We considered variables associated
138 with contact frequency at $p < 0.10$ for multivariable analysis, and retained them in
139 multivariable models if they resulted in a reduction of the Bayesian Information Criterion
140 (BIC).

141 Next, we explored whether people reporting a high frequency of casual contacts (≥ 10 casual
142 contacts) differed from those reporting fewer contacts with regards to their socio-
143 demographic characteristics. We did so using log-binomial regression to compute crude and
144 adjusted relative risks (RRs) for having a high frequency. In all analyses we accounted for
145 possible within-cluster correlation by using linearized based variance estimators [16].
146 Analyses were also weighted for the unequal probabilities of sampling selection by age
147 group.

148 Age-specific social contact patterns

149 We analysed the age-specific contact patterns through matrices of the mean number of
150 contacts between participants of age group j and individuals in age group i , adjusting for
151 reciprocity, as in Melegaro et al. [6].

152 If x_{ij} denotes the total number of contacts in age group i reported by individuals in age
153 group j , the mean number of reported contacts (m_{ij}) is calculated as x_{ij}/p_j , where p_j is
154 the study population size of age group j . At the population level the frequency of contacts
155 made between age groups should be equal such that $m_{ij}P_j = m_{ji}P_i$. The expected
156 number of contacts between the two groups is therefore $C_{ij} = (m_{ij}P_j + m_{ji}P_i)/2$. Hence, the
157 mean number of contacts corrected for reciprocity (m_{ij}^C) can be expressed as C_{ij}/P_j .

158 Epidemic simulations

159 Finally, in order to explore the infection transmission dynamics resulting from our contact
160 pattern data, we simulated the spread of an immunizing respiratory infection transmitted
161 through close contact in a totally susceptible population. The model contained nine mixing
162 age groups, with a transmission rate at which individuals in age group j come into routine
163 contact with individuals in age group i computed as $\beta_{ij} = m_{ij}^C / \omega_i$, where ω_i is the
164 proportion of individuals in age group i , and qm_{ij}^C is the next generation matrix, with q
165 representing the probability of successful transmission per contact event [17]. We assumed
166 q to be homogeneous and constant across all age groups and conducted a set of
167 simulations for fixed values of q between 25% and 40%, in line with what has been

168 reported with influenza pandemic strains [17, 18]. The basic reproduction number (R_0) –
169 which corresponds to the average number of people infected by one infectious individual in
170 a totally susceptible population – was calculated as the dominant eigenvalue of the next
171 generation matrix. We took uncertainty estimates in the contact matrices (and hence final
172 size outputs) into account by iterating the model on bootstrapped matrices.

173 We then computed the final epidemic size (i.e. the number of individuals who would have
174 been infected during the epidemic) for each specific age group, based on a mass action
175 model adapted to account for multiple age classes, as described in Kucharski et al. [19].

176 Estimates obtained using the contact data from Uganda were compared to that of Great
177 Britain, using data from the POLYMOD study [4] for the latter and a similar approach to
178 compute the mixing matrix. The model was parameterised with social contact data on
179 physical contacts only, lasting ≥ 5 minutes, rather than all contacts, given that physical
180 contacts generally seem to better capture contact structures relevant for the transmission
181 of respiratory infections [6], and that the definition of physical contacts is more similar and
182 comparable between studies than that of overall contacts.

183 All analyses were performed in STATA 13.1 IC and R version 3.2.

184 **RESULTS**

185 **Study population**

186 A total of 568 individuals participated in the survey, but no information about age and
187 contacts was missing for 2 individuals, resulting in 566 included in the analysis. This

188 corresponds to an overall response rate of 83%; higher among ≥ 15 years old (98%), and
189 lower among under 2s (68%), 2-4 year olds (64%), 5 –14y olds (69%).

190 There were more female (58%) than male respondents, but this differed by age group, with
191 fewer females in young age groups and more adult females than males (Table S1 in the
192 Supporting Information).

193 The mean household size was 5.3 (median 5, range 1 – 18). Almost all (98%) school-aged
194 children aged 6 – 14 years attended school or college. Among adults, agriculture was the
195 main occupation and about 27% of the females were homemakers/housewives (Table 1).

196 **Characteristics of contacts**

197 Contacts (i.e. ≥ 5 minutes long)

198 A total of 3,965 contacts with different individuals were reported, corresponding to an
199 average of 7.2 contacts per person (median 7, range 0 - 25) (Figure 1). The majority of
200 contacts were physical (mean 5.1, median 5 (range 0 – 18)).

201 Over half of all contacts ($n = 2,060$ (52%)) were with household members, 627 (16%) with
202 other relatives, 873 (22%) with colleagues/friends/schoolmates and 402 (10%) with other
203 individuals. The duration of contacts is shown in Figure S2 (Supporting Information).

204 Most contacts (82%) were with individuals who would be normally seen daily, 520 (13%)
205 with people normally seen at least weekly, 4% with people met more rarely and 1% of the
206 reported contacts were with people that the participants had never met before.

207 We found marked differences in the number of contacts by age group, but not by sex.
208 School-aged children reported the highest daily number of contacts, while the elderly had
209 the fewest (Table 1). Table 1 provides further details about the population characteristics,
210 the mean number of contacts by socio-demographic and other covariates, as well as the
211 ratio of mean contacts by covariate. Results were adjusted for age, but not other variables,
212 as identified through the results of the negative binomial model.

213 Overall, contacts tended to be assortative, as shown by the strong diagonal feature on
214 Figure 2, with most of the intergenerational mixing occurring within households (Figure 3).
215 Only teenagers and adults reported non-physical contacts (Figure 3). Reciprocity correction
216 accounted for the differential reporting between age groups, particularly higher frequency
217 of contacts reported by small children with older age groups than older age groups reported
218 (Figure S2 in Supporting Information).

219 There was no statistical difference in the average number of contacts between weekend
220 (Sunday) and weekdays (Monday, Tuesday, Thursday and Friday) (Table 1). Given that
221 survey teams had a day off on Sundays and on Thursdays information about contact on
222 Saturdays and Wednesdays was not recorded. About a quarter (n=136 (24%)) of
223 participants reported social encounters outside their village of residence, and about 12% of
224 contacts occurred outside participants' village of residence. The majority (56%) of people
225 who travelled outside their village went to places located within a 5km radius from the
226 centre point of their village of residence, and 90% stayed within 12km (Figure 4). Adult
227 males tended to travel more than females (Figure 4). Most contacts made outside the
228 household as well as those with individuals outside participants' village were mostly
229 assortative (Figure 3), and the proportion of contacts outside the village was different by

230 age group ($P < 0.001$); higher among adults, increasingly so as distance from home increased
231 (Figure 4).

232 'Casual' contacts (<5 minutes long)

233 Information on the number of casual contacts was reported by 490 (87%) participants.
234 Among those, 64% ($n=315$) estimated they had fewer than 10 different contacts, 24%
235 reported between 10 and 19 casual contacts, 6% reported between 20 – 29 contacts and
236 6% reported an estimated 30 contacts or more.

237 Individuals who reported high levels (i.e. ≥ 10 contacts) of social contacts also tended to
238 report more contacts (Table 1). We found no difference between those reporting high
239 number of social contacts (≥ 10) and others, by age, sex or day of the week (Table S2 in the
240 Supporting Information). However, people whose primary activity was at home tended to
241 reported fewer casual contacts than others, and there were about 50% more individuals
242 reporting high levels of casual contacts among those who travelled outside their village.

243 Epidemic simulations

244 Finally, we compared patterns of reported physical contacts in Uganda and Great Britain,
245 and explored differences in the relative and absolute epidemic size by age group, as well as
246 the corresponding R_0 , for a hypothetical respiratory infection in an immune-naive
247 population.

248 The number of reported physical contacts was similar between Uganda and Great Britain,
249 with the average number of contacts by age group ranging from 3.2 (≥ 65 year olds) to 7.3 (2
250 – 4 year olds) in Uganda and from 3.3 (55 – 64 year olds) to 7.3 (10 – 14 year olds) in Great

251 Britain. However contacts were more assortative in Britain than in Uganda (Figures 5A & B),
252 some of which might be related to differences in household structures and number of
253 household contacts, as contacts outside the household were mostly assortative (Figure 3).

254 The computed mean values of R_0 for a per contact infectivity value (q) ranging from 0.25
255 to 0.40 was slightly higher in Great Britain than in Uganda (1.51 to 2.41 vs. 1.40 to 2.24).
256 Figure 5F shows the values for an infectivity parameter of 0.33. The proportion of people
257 infected in younger age groups was also higher in Great Britain, and there were
258 proportionally more adults infected in Uganda. However, given the differences in population
259 structure, the total number of infections in the population was higher in Uganda than in
260 Great Britain (Figures 5 C – E).

261 **DISCUSSION**

262 To our knowledge this is only the third study of its kind in Africa [10, 12], and the first one to
263 specifically explore spatial patterns of social contacts. The quantification of mixing patterns
264 is central to accurately model transmission dynamics and inform infectious disease control
265 strategies [4]. Having such data thus fills an important gap, particularly given the high
266 burden of respiratory infections in low income settings [20, 21], and the risk of emerging
267 and re-emerging diseases transmitted by close interpersonal contact, such as influenza [22],
268 measles [23] or meningitis [24].

269 Our findings share similarities with studies from Africa [10, 12, 13] and other low or lower-
270 middle income settings [15, 25], including the high contact frequency among school-aged
271 children and that most contacts tend to be age-assortative. We also found substantial
272 mixing between age groups, largely driven by intra-household mixing. This may result in a

273 higher force of infection from children to adults than would be seen in other contexts such
274 as Great Britain, as our final size epidemic model suggests. The final size model should be
275 seen as an illustration of how different social mixing patterns impact on disease
276 epidemiology in different settings, rather than a specific quantification of the differences. It
277 shows the importance of using setting-specific data when modelling disease dynamics and
278 evaluates control strategies. Our data could be best applied to evaluate transmission
279 dynamics and the impact of interventions for endemic diseases and current epidemics in
280 non-naïve population in similar rural East African contexts.. In our final size model, it is also
281 likely that our retrospective design resulted in underreporting compared to a prospective
282 diary-based approach [26], which hampers comparisons between countries. In sensitivity
283 analyses we explored the impact of potential underreporting in our retrospective survey
284 design compared to a prospective diary-based approach [26], assuming a 25% under-
285 ascertainment compared to a diary-based study, with homogeneous underreporting across
286 age groups. In such scenario, the proportion of infections across all age groups is predicted
287 to be higher in Uganda than in Britain, disproportionately so in adults, and the R_0 to be higher
288 too (see Figure S3 in the Supporting Material).

289 Our results also provide important insights into the local spatial dynamics of routine daily
290 human interactions, showing that most contacts tend to occur within the vicinity of people's
291 area of residence, that working age adult males travel most and young children and the
292 elderly the least, and that contacts tend to be increasingly age assortative as people travel
293 further away from home. Similar patterns were observed in rural and semi-urban China [11].
294 Such findings have important implications to predict outbreak dynamics and control
295 strategies given that interconnectedness between geographic patches is an essential factor

296 driving epidemic extinction or persistence of epidemics hotspots and the effectiveness of
297 control strategies. Studies of measles in Niger suggest that dynamics differ from that
298 observed in high-income countries in the pre-vaccination era, likely due to different mixing
299 patterns and weaker spatial connectivity [27, 28]. This, together with important variations in
300 vaccination coverage between local geographic patches [29-31], strengthens the need to
301 account for spatial mobility when designing efficient control strategies in those settings.
302 Optimal targeted interventions tailored to specific geographic clusters of high transmission
303 have also been key considerations in recent cholera outbreaks in Africa, given the limited
304 available vaccine doses [32, 33]. Spatially targeted approaches are also central to outbreak
305 control in the recent West African Ebola epidemic [34], and the current measles epidemic in
306 the Democratic Republic of the Congo, which is sustained in part due to

307 In our study the frequency of contacts was about half that of the number of contacts
308 reported in Kenya, [10] or South Africa [12], and lower than in a recent contact study
309 conducted by Melegaro et al. in rural and peri-urban areas in Zimbabwe [13]. Although
310 differences between settings are expected, some of these are likely to be due to the
311 exclusion of 'casual contacts' from our contact count. There might be further differences
312 linked to the definition of social contacts, which was based on conversational encounters in
313 our study but not in the Kenyan study [10]. When defining contacts based on conversational
314 exchanges the household setting tends to dominate over other settings, compared to a
315 more inclusive definition [8].

316 Both our contact definition and the retrospective study design may have also resulted in
317 some level of reporting bias with more stable, regular contacts being reported over others.
318 However, the extent to which a more inclusive definition reflects contact events relevant for

319 transmission remains unclear. Modelling studies suggest that close interpersonal rather
320 than short casual contacts matter more for transmission of respiratory infections [6]. In
321 addition, for modelling purposes the age-specific structure of relative contact frequency
322 matters more than the actual reported frequency, as matrices are scaled to fit
323 epidemiological data. Our retrospective interview-based design thus offers a simpler and
324 easier alternative to prospective diary based approaches, particularly in such setting.
325 Further research should explore what contact information is most relevant and how such
326 data should best be captured.

327 Selection bias may have occurred to some extent, particularly given that more adult women
328 were included than men. However, there was no significant difference in the number of
329 contacts reported between males and females, including at the weekend, suggesting that
330 selection bias was unlikely to be major. We also tried to reduce selection bias by
331 interviewing on Saturdays people who were initially absent on the survey.

332 In conclusion, our study fills an important gap for two main reasons. First, we provide
333 information by detailed age groups about social contacts and mixing patterns relevant to
334 the spread of infectious diseases in a region where such data are scarce. Second, we also
335 provide some insights into spatial characteristics of social encounters. Although this has
336 increasingly being recognized as an important component in evaluating epidemic risk and in
337 the design of efficient control strategies, it has not previously been quantified in low-income
338 settings, and should be explored further. Our study thus provides essential evidence to
339 inform further research and infectious disease modelling work, particularly in similar rural
340 African settings.

341 **FUNDING**

342 This work was supported by Médecins Sans Frontières International Office, Geneva,
343 Switzerland. Epicentre received core funding from Médecins Sans Frontières. OLP was
344 supported by the AXA Research Fund through a Doctoral research fellowship. The funders
345 had no role in the study design, data collection, data analysis, data interpretation or write
346 up of the manuscript.

347 **ACKNOWLEDGMENTS**

348 We would like to thank all the surveyors for their work. We are grateful to Gertrude
349 Ngabirano for her help in translating questionnaires.

350 **CONFLICT OF INTEREST**

351 The authors declare they have no conflict of interest.

352 **REFERENCES**

- 353 1. Heesterbeek H, Anderson RM, Andreasen V, Bansal S, De Angelis D, Dye C, et al.
354 Modeling infectious disease dynamics in the complex landscape of global health. *Science*.
355 2015;347(6227):aaa4339. doi: 10.1126/science.aaa4339. PubMed PMID: 25766240;
356 PubMed Central PMCID: PMC4445966.
- 357 2. Read JM, Edmunds WJ, Riley S, Lessler J, Cummings DA. Close encounters of the
358 infectious kind: methods to measure social mixing behaviour. *Epidemiology and infection*.
359 2012;140(12):2117-30. doi: 10.1017/S0950268812000842. PubMed PMID: 22687447;
360 PubMed Central PMCID: PMC4288744.
- 361 3. Edmunds WJ, O'Callaghan CJ, Nokes DJ. Who mixes with whom? A method to
362 determine the contact patterns of adults that may lead to the spread of airborne infections.
363 *Proceedings Biological sciences / The Royal Society*. 1997;264(1384):949-57. doi:
364 10.1098/rspb.1997.0131. PubMed PMID: 9263464; PubMed Central PMCID: PMC1688546.
- 365 4. Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, et al. Social contacts
366 and mixing patterns relevant to the spread of infectious diseases. *PLoS medicine*.
367 2008;5(3):e74. doi: 10.1371/journal.pmed.0050074. PubMed PMID: 18366252; PubMed
368 Central PMCID: PMC2270306.

- 369 5. Beutels P, Shkedy Z, Aerts M, Van Damme P. Social mixing patterns for transmission
370 models of close contact infections: exploring self-evaluation and diary-based data collection
371 through a web-based interface. *Epidemiology and infection*. 2006;134(6):1158-66. doi:
372 10.1017/S0950268806006418. PubMed PMID: 16707031; PubMed Central PMCID:
373 PMC2870524.
- 374 6. Melegaro A, Jit M, Gay N, Zagheni E, Edmunds WJ. What types of contacts are
375 important for the spread of infections?: using contact survey data to explore European
376 mixing patterns. *Epidemics*. 2011;3(3-4):143-51. doi: 10.1016/j.epidem.2011.04.001.
377 PubMed PMID: 22094337.
- 378 7. Smieszek T, Burri EU, Scherzinger R, Scholz RW. Collecting close-contact social mixing
379 data with contact diaries: reporting errors and biases. *Epidemiology and infection*.
380 2012;140(4):744-52. doi: 10.1017/S0950268811001130. PubMed PMID: 21733249.
- 381 8. Bolton KJ, McCaw JM, Forbes K, Nathan P, Robins G, Pattison P, et al. Influence of
382 contact definitions in assessment of the relative importance of social settings in disease
383 transmission risk. *PloS one*. 2012;7(2):e30893. doi: 10.1371/journal.pone.0030893. PubMed
384 PMID: 22359553; PubMed Central PMCID: PMC3281034.
- 385 9. Kretzschmar M, Mikolajczyk RT. Contact profiles in eight European countries and
386 implications for modelling the spread of airborne infectious diseases. *PloS one*.
387 2009;4(6):e5931. doi: 10.1371/journal.pone.0005931. PubMed PMID: 19536278; PubMed
388 Central PMCID: PMC2691957.
- 389 10. Kiti MC, Kinyanjui TM, Koech DC, Munywoki PK, Medley GF, Nokes DJ. Quantifying
390 age-related rates of social contact using diaries in a rural coastal population of Kenya. *PloS*
391 *one*. 2014;9(8):e104786. doi: 10.1371/journal.pone.0104786. PubMed PMID: 25127257;
392 PubMed Central PMCID: PMC4134222.
- 393 11. Read JM, Lessler J, Riley S, Wang S, Tan LJ, Kwok KO, et al. Social mixing patterns in
394 rural and urban areas of southern China. *Proceedings Biological sciences / The Royal Society*.
395 2014;281(1785):20140268. doi: 10.1098/rspb.2014.0268. PubMed PMID: 24789897;
396 PubMed Central PMCID: PMC4024290.
- 397 12. Johnstone-Robertson SP, Mark D, Morrow C, Middelkoop K, Chiswell M, Aquino LD,
398 et al. Social mixing patterns within a South African township community: implications for
399 respiratory disease transmission and control. *American journal of epidemiology*.
400 2011;174(11):1246-55. doi: 10.1093/aje/kwr251. PubMed PMID: 22071585; PubMed
401 Central PMCID: PMC3224253.
- 402 13. Melegaro A, Del Fava E, Poletti P, Merler S, Nyamukapa C, Williams J, et al. Social
403 Contact Structures and Time Use Patterns in the Manicaland Province of Zimbabwe. *PloS*
404 *one*. 2017;12(1):e0170459. doi: 10.1371/journal.pone.0170459. PubMed PMID: 28099479;
405 PubMed Central PMCID: PMC5242544.
- 406 14. Wesolowski A, Metcalf CJ, Eagle N, Kombich J, Grenfell BT, Bjornstad ON, et al.
407 Quantifying seasonal population fluxes driving rubella transmission dynamics using mobile
408 phone data. *Proceedings of the National Academy of Sciences of the United States of*
409 *America*. 2015;112(35):11114-9. doi: 10.1073/pnas.1423542112. PubMed PMID: 26283349;
410 PubMed Central PMCID: PMC4568255.
- 411 15. Horby P, Pham QT, Hens N, Nguyen TT, Le QM, Dang DT, et al. Social contact
412 patterns in Vietnam and implications for the control of infectious diseases. *PloS one*.
413 2011;6(2):e16965. doi: 10.1371/journal.pone.0016965. PubMed PMID: 21347264; PubMed
414 Central PMCID: PMC3038933.

- 415 16. Rogers WH. Regression standard errors in clustered samples. *Stata Technical Bulletin*
416 1993;13:19–23.
- 417 .
- 418 17. Wallinga J, Teunis P, Kretzschmar M. Using data on social contacts to estimate age-
419 specific transmission parameters for respiratory-spread infectious agents. *American journal*
420 *of epidemiology*. 2006;164(10):936-44. doi: 10.1093/aje/kwj317. PubMed PMID: 16968863.
- 421 18. House T, Inglis N, Ross JV, Wilson F, Suleman S, Edeghere O, et al. Estimation of
422 outbreak severity and transmissibility: Influenza A(H1N1)pdm09 in households. *BMC*
423 *medicine*. 2012;10:117. doi: 10.1186/1741-7015-10-117. PubMed PMID: 23046520;
424 PubMed Central PMCID: PMC3520767.
- 425 19. Kucharski AJ, Kwok KO, Wei VW, Cowling BJ, Read JM, Lessler J, et al. The
426 contribution of social behaviour to the transmission of influenza A in a human population.
427 *PLoS pathogens*. 2014;10(6):e1004206. doi: 10.1371/journal.ppat.1004206. PubMed PMID:
428 24968312; PubMed Central PMCID: PMC4072802.
- 429 20. Nair H, Simoes EA, Rudan I, Gessner BD, Azziz-Baumgartner E, Zhang JS, et al. Global
430 and regional burden of hospital admissions for severe acute lower respiratory infections in
431 young children in 2010: a systematic analysis. *Lancet*. 2013;381(9875):1380-90. doi:
432 10.1016/S0140-6736(12)61901-1. PubMed PMID: 23369797; PubMed Central PMCID:
433 PMC3986472.
- 434 21. Black RE, Cousens S, Johnson HL, Lawn JE, Rudan I, Bassani DG, et al. Global, regional,
435 and national causes of child mortality in 2008: a systematic analysis. *Lancet*.
436 2010;375(9730):1969-87. doi: 10.1016/S0140-6736(10)60549-1. PubMed PMID: 20466419.
- 437 22. Radin JM, Katz MA, Tempia S, Talla Nzussouo N, Davis R, Duque J, et al. Influenza
438 surveillance in 15 countries in Africa, 2006-2010. *The Journal of infectious diseases*.
439 2012;206 Suppl 1:S14-21. doi: 10.1093/infdis/jis606. PubMed PMID: 23169960.
- 440 23. Maurice J. Measles outbreak in DR Congo an "epidemic emergency". *Lancet*.
441 2015;386(9997):943. doi: 10.1016/S0140-6736(15)00115-4. PubMed PMID: 26369457.
- 442 24. Burki T. Meningitis outbreak in Niger is an urgent warning. *The Lancet Infectious*
443 *diseases*. 2015;15(9):1011.
- 444 25. Grijalva CG, Goeyvaerts N, Verastegui H, Edwards KM, Gil AI, Lanata CF, et al. A
445 household-based study of contact networks relevant for the spread of infectious diseases in
446 the highlands of Peru. *PloS one*. 2015;10(3):e0118457. doi: 10.1371/journal.pone.0118457.
447 PubMed PMID: 25734772; PubMed Central PMCID: PMC4348542.
- 448 26. McCaw JM, Forbes K, Nathan PM, Pattison PE, Robins GL, Nolan TM, et al.
449 Comparison of three methods for ascertainment of contact information relevant to
450 respiratory pathogen transmission in encounter networks. *BMC infectious diseases*.
451 2010;10:166. doi: 10.1186/1471-2334-10-166. PubMed PMID: 20537186; PubMed Central
452 PMCID: PMC2893181.
- 453 27. Grais RF, Ferrari MJ, Dubray C, Bjornstad ON, Grenfell BT, Djibo A, et al. Estimating
454 transmission intensity for a measles epidemic in Niamey, Niger: lessons for intervention.
455 *Transactions of the Royal Society of Tropical Medicine and Hygiene*. 2006;100(9):867-73.
456 doi: 10.1016/j.trstmh.2005.10.014. PubMed PMID: 16540134.
- 457 28. Ferrari MJ, Grais RF, Bharti N, Conlan AJ, Bjornstad ON, Wolfson LJ, et al. The
458 dynamics of measles in sub-Saharan Africa. *Nature*. 2008;451(7179):679-84. doi:
459 10.1038/nature06509. PubMed PMID: 18256664.
- 460 29. Le Polain de Waroux O, Schellenberg JR, Manzi F, Mrisho M, Shirima K, Mshinda H, et
461 al. Timeliness and completeness of vaccination and risk factors for low and late vaccine

- 462 uptake in young children living in rural southern Tanzania. *International health*.
 463 2013;5(2):139-47. doi: 10.1093/inthealth/iht006. PubMed PMID: 24030114.
- 464 30. Babirye JN, Engebretsen IM, Makumbi F, Fadnes LT, Wamani H, Tylleskar T, et al.
 465 Timeliness of childhood vaccinations in Kampala Uganda: a community-based cross-
 466 sectional study. *PloS one*. 2012;7(4):e35432. doi: 10.1371/journal.pone.0035432. PubMed
 467 PMID: 22539972; PubMed Central PMCID: PMC3335141.
- 468 31. Satzke C, Turner P, Virolainen-Julkunen A, Adrian PV, Antonio M, Hare KM, et al.
 469 Standard method for detecting upper respiratory carriage of *Streptococcus pneumoniae*:
 470 updated recommendations from the World Health Organization Pneumococcal Carriage
 471 Working Group. *Vaccine*. 2013;32(1):165-79. doi: 10.1016/j.vaccine.2013.08.062. PubMed
 472 PMID: 24331112.
- 473 32. Azman AS, Luquero FJ, Rodrigues A, Palma PP, Grais RF, Banga CN, et al. Urban
 474 cholera transmission hotspots and their implications for reactive vaccination: evidence from
 475 Bissau city, Guinea bissau. *PLoS neglected tropical diseases*. 2012;6(11):e1901. doi:
 476 10.1371/journal.pntd.0001901. PubMed PMID: 23145204; PubMed Central PMCID:
 477 PMC3493445.
- 478 33. Luquero FJ, Banga CN, Remartinez D, Palma PP, Baron E, Grais RF. Cholera epidemic
 479 in Guinea-Bissau (2008): the importance of "place". *PloS one*. 2011;6(5):e19005. doi:
 480 10.1371/journal.pone.0019005. PubMed PMID: 21572530; PubMed Central PMCID:
 481 PMC3087718.
- 482 34. Zinszer K, Morrison K, Anema A, Majumder MS, Brownstein JS. The velocity of Ebola
 483 spread in parts of west Africa. *The Lancet Infectious diseases*. 2015;15(9):1005-7. doi:
 484 10.1016/S1473-3099(15)00234-0. PubMed PMID: 26333328.

485

486

487

488

489 **Table 1.** Mean Number of Reported Contacts and Ratio of Means By Socio-demographic
 490 Characteristic of The Study Population, Sheema, Uganda, January – March 2014.

Variables	N	Mean number of contacts (95%CI)	Crude RoM (95%CI)	Age adjusted RoM (95%CI)
Age groups				
<2y	61	6.11 (5.42 ,6.81)	0.99 (0.83 ,1.17)	
2-4y	57	6.70 (6.08 ,6.81)	1.08 (0.95 ,1.23)	
5-9y	74	8.50 (7.79 ,9.20)	1.37 (1.18 ,1.59)	

10-14y	51	8.70 (7.52 ,9.90)	1.40 (1.19 ,1.66)	
15-24y	91	6.20 (5.51 ,6.88)	ref	
25-34y	55	6.89 (5.99 ,7.80)	1.11 (0.93 ,1.33)	
35-44y	54	7.73 (6.99 ,8.46)	1.25 (1.09 ,1.43)	
45-54y	46	7.74 (6.37 ,9.11)	1.25 (1.01 ,1.54)	
55-64y	26	6.27 (5.01 ,7.53)	1.01 (0.80 ,1.29)	
65+y	48	4.85 (4.18 ,5.53)	0.78 (0.64 ,0.95)	
Sex				
Female	328	7.05 (6.66 ,7.44)	ref	
Male	235	7.47 (6.82 ,8.11)	1.06 (0.96 ,1.18)	
Occupation/daily activity				
Pre-school child	93	7.00 (6.30 ,7.70)	1.06 (0.90 ,1.23)	1.26 (1.02 ,1.55)
Student	166	8.27 (7.56 ,8.98)	1.27 (1.10 ,1.46)	1.30 (1.05 ,1.63)
Office worker	4	11.34 (9.64 ,13.03)	1.81 (1.20 ,2.73)	1.70 (1.32 ,2.18)
Shop worker	34	6.82 (5.56 ,8.08)	1.03 (0.84 ,1.25)	1.03 (0.83 ,1.29)
Agriculture	105	7.30 (6.58 ,8.01)	1.12 (0.97 ,1.30)	1.12 (0.96 ,1.30)
Other manual worker	40	5.37 (4.36 ,6.38)	0.85 (0.70 ,1.04)	0.85 (0.68 ,1.06)
At home	60	6.43 (5.62 ,7.24)	ref	ref
Unemployed	11	6.44 (2.92 ,9.96)	0.83 (0.60 ,1.15)	1.22 (0.77 ,1.94)
Retired	8	4.77 (3.75 ,5.80)	0.71 (0.48 ,1.05)	0.89 (0.70 ,1.13)
Other/unreported	41	6.65 (5.80 ,7.51)	1.02 (0.85 ,1.23)	1.18 (0.98 ,1.43)
Day of the week				
Weekday	439	7.14 (6.79 ,7.49)	ref	
Sunday	124	7.50 (6.56 ,8.44)	1.05 (0.92 ,1.20)	1.04 (0.92 ,1.18)
Travel outside village/town in previous 24 hours				
No	427	6.57 (6.20 ,6.92)	ref	
Yes	139	9.04 (8.35 ,9.73)	1.38 (1.25 ,1.52)	1.35 (1.22 ,1.49)
Number of casual contacts				
<10	315	5 (1 -15)	Ref	Ref
10 -19	119	8 (2-23)	1.43 (1.28, 1.59)	1.39 (1.25, 1.55)
≥20	56	9 (2-25)	1.64 (1.44; 1.86)	1.61 (1.43,1.83)
Don't know	76	8 (0-19)	1.52 (1.37; 1.68)	1.45 (1.29, 1.64)

491 Footnote CI=Confidence Interval; RoM: Ratio of Means

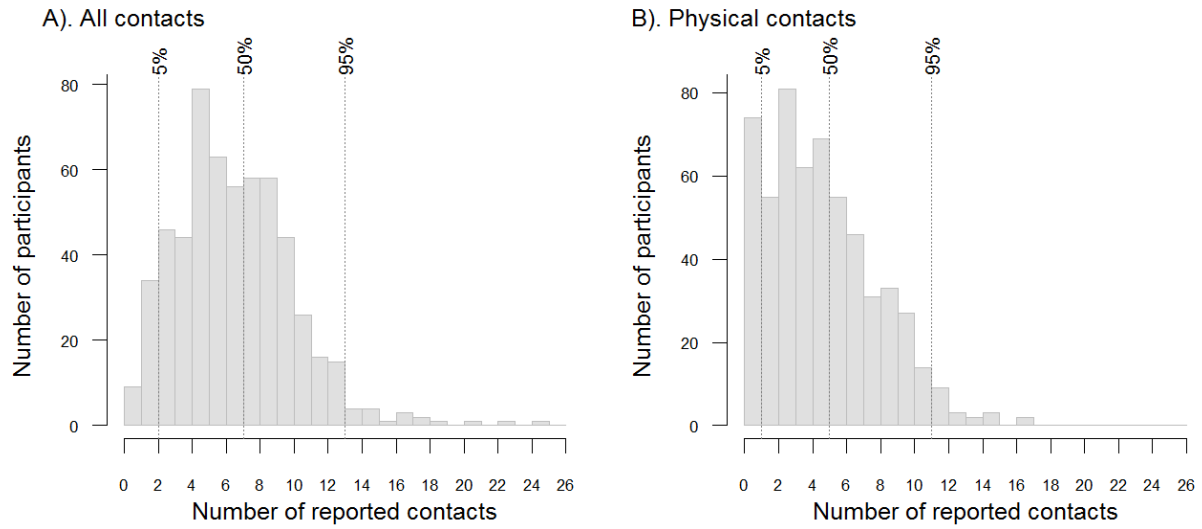
492

493 **Figures**

494 **Figure 1:** Number of Reported Contacts, Including All Contacts (A) and Physical contacts (B),

495 Sheema, Uganda, January – March 2014. Legend: the vertical dotted lines represent the 5%

496 centile, the median and 95% centile of the total number of reported contacts



497

498

499

500

501

502

503

504

505 **Figure 2:** Average Number of Reported Contacts By Age Group, Sheema, Uganda, January –

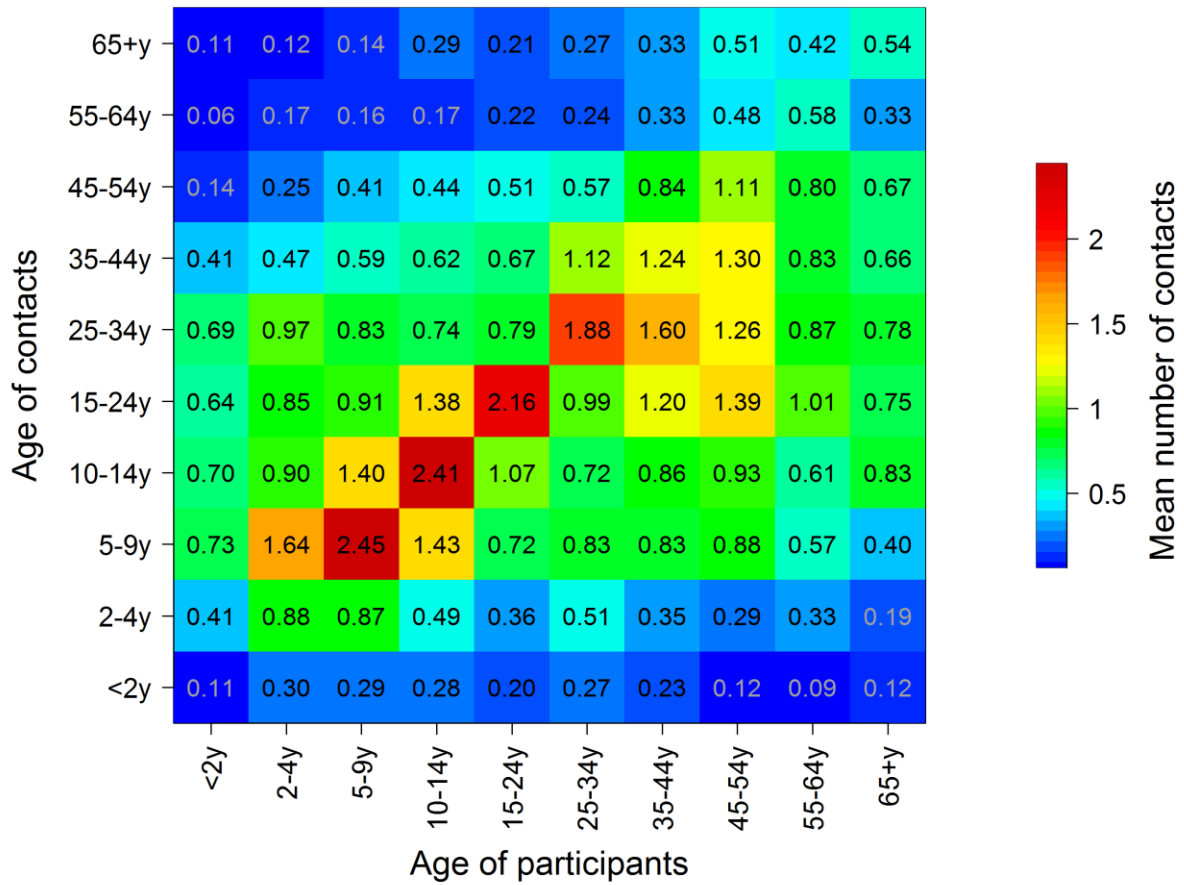
506 March 2014 Legend: Numbers in each cell represent the average number of contacts

507 between between age groups corrected for reciprocity, and 95% confidence intervals are

508 shown

in

brackets



509

510

511

512

513

Figure 3: Contact Matrices With Household members and Non-Household Members (Left upper and lower panel), for Physical and Non-Physical Contacts (Middle upper and lower panel), and for Contacts Made Within and Outside the Village (Right upper and lower panel), Sheema, Uganda, January – March 2014

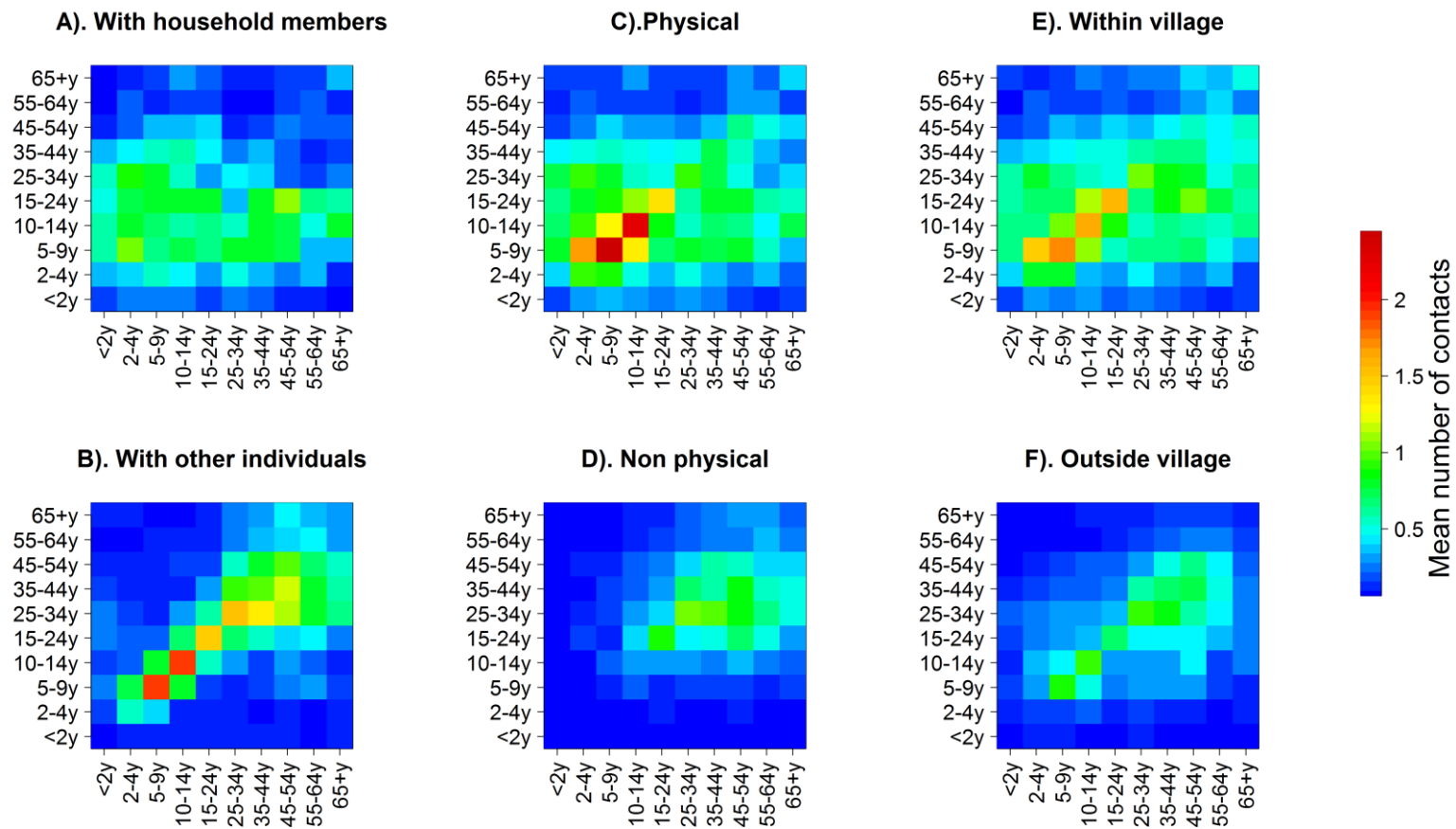
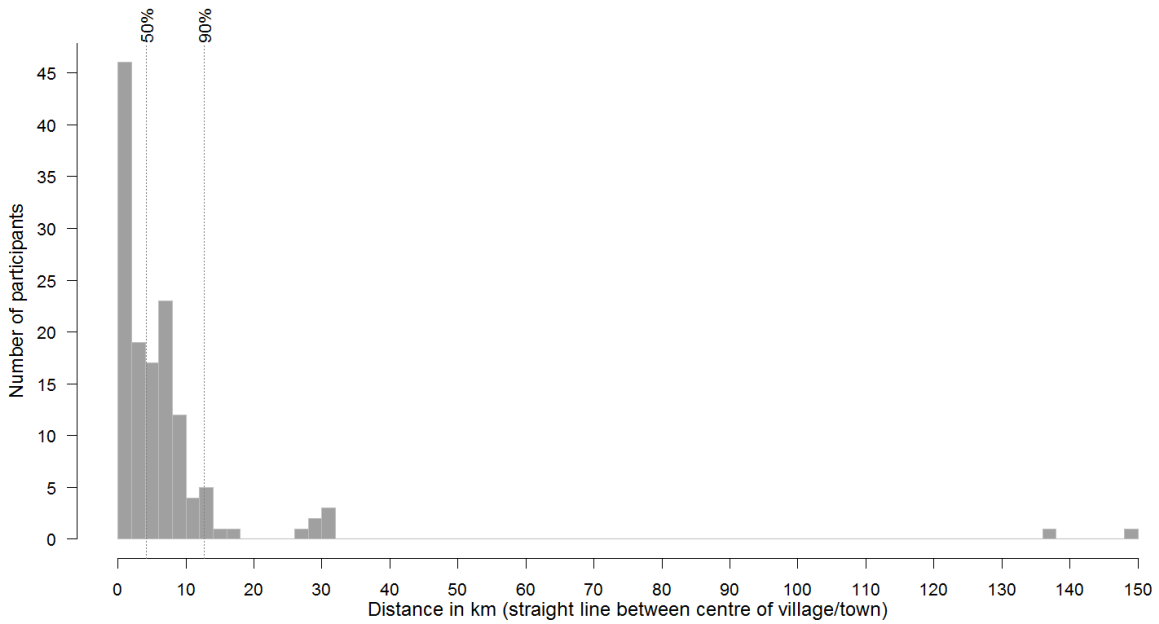


Figure 4: Distance Travelled By Study Participants in the 24 Hours Preceding the Survey, Overall (A) and By Categories of Distance, Age and Sex (B), Sheema, Uganda, January – March 2014

A). Maximum distance (straight line) participants travelled in the previous 24 hours



B). Categories by age and sex

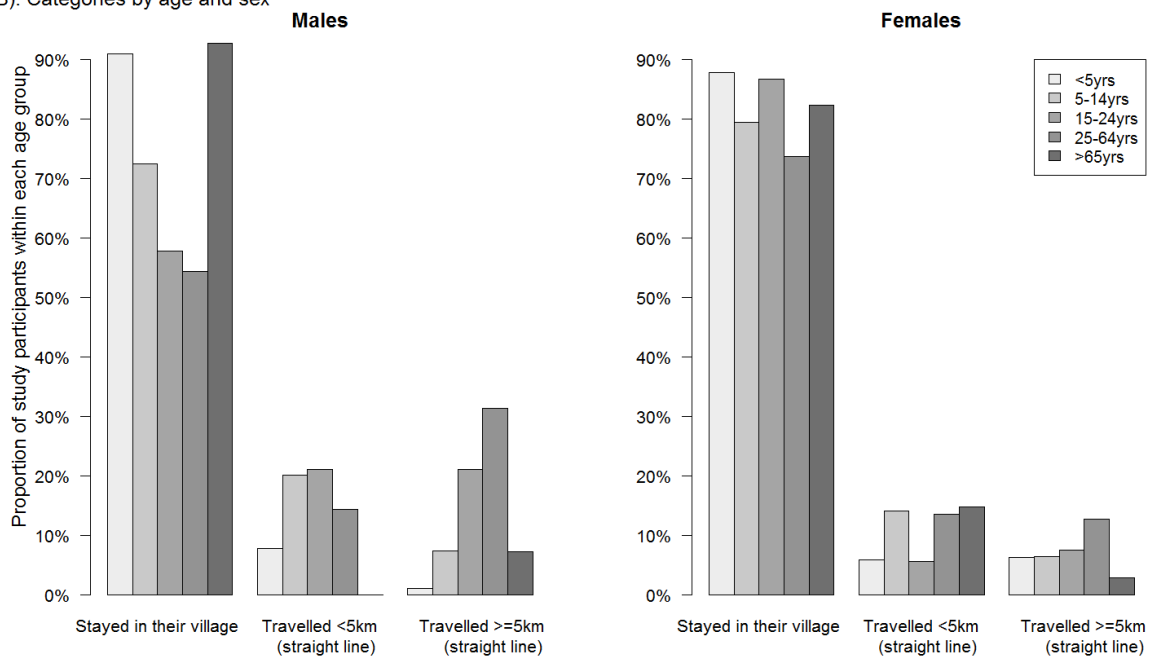
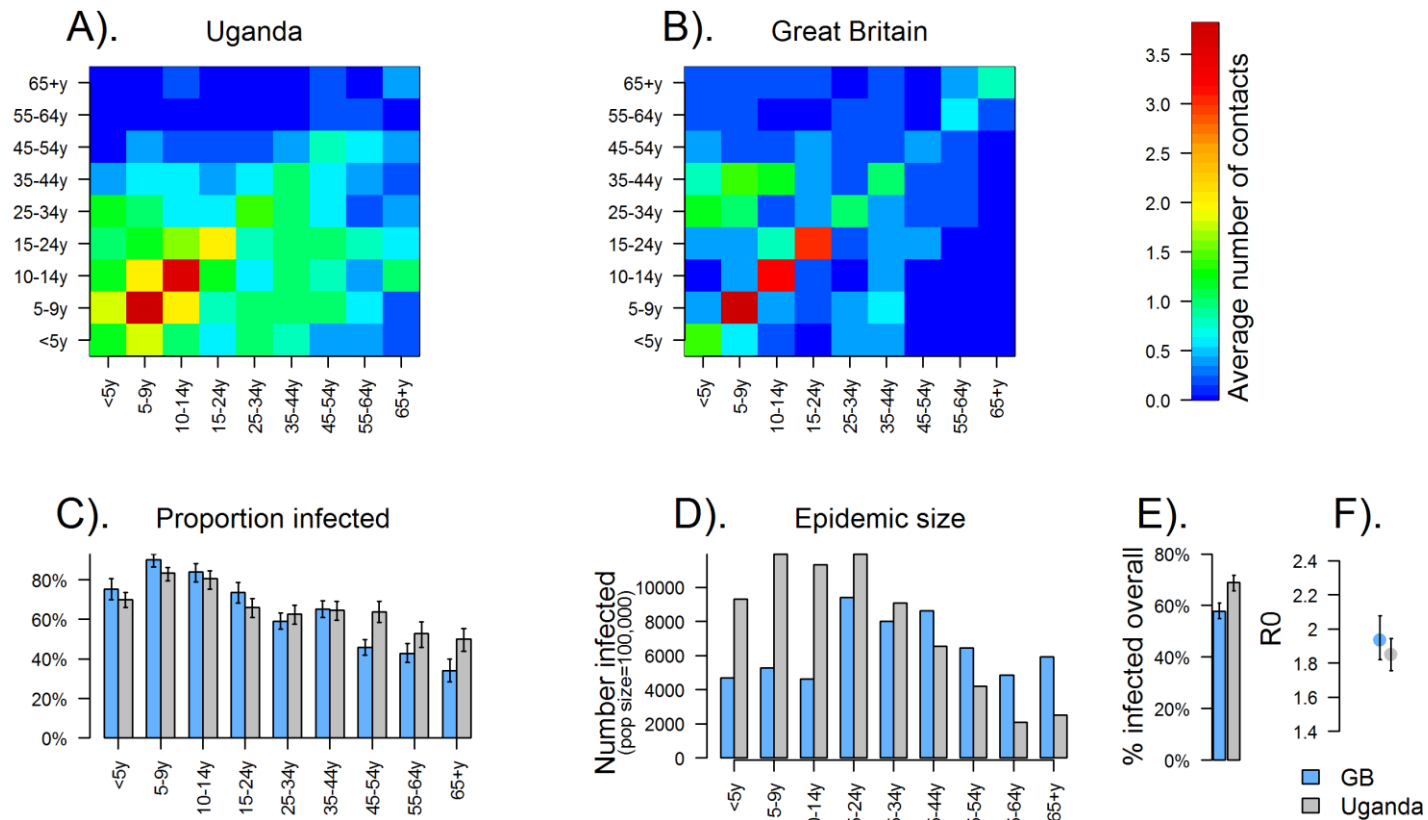


Figure 5: Epidemic Simulation Using Matrices on Physical Contacts from Uganda (A) and Great Britain (B), for a Hypothetical Respiratory Infection In An Immune-Naïve Population, with the Proportion Infected by Age Group (C), the Epidemic Size by Age group (D), the Overall Proportion Infected (D) and the Basic Reproduction Number R_0 (F).



Legend: A: Matrix for physical contacts in Uganda. B: Matrix of physical contacts in Great Britain. C: Epidemic final size simulation: Proportion of individuals infected by age group in Great Britain (blue) and Uganda (grey), with error bars representing the 95% confidence interval. The results are presented for a q value of 33%. D: Epidemic size by age group, based on a total population size of 100,000 in Great Britain and in Uganda. E: Total proportion of people who were infected at the end of the epidemic in each setting. F: Estimates of R_0 for each setting, based on a q value of 33%, with dots showing the mean value and the bars showing the 95% CI

Web material

Tables

Table S1. Age and Sex Distribution of Study Participants

Age category	Number (%) female	Number (%) male	Total number
<2 years	23 (38%)	38 (62%)	61
2 – 4 years	26 (45%)	31 (55%)	57
5 – 9 years	37 (50%)	37 (50%)	74
10 – 14 years	22 (43%)	29 (57%)	51
15 - 24 years	53 (58%)	38 (42%)	91
25 – 34 years	45 (79%)	12 (21%)	57
35 – 44 years	35 (64%)	20 (36%)	55
45 – 54 years	35 (76%)	11 (24%)	46
55 – 64 years	20 (77%)	6 (23%)	26
65+ years	34 (71%)	14 (29%)	48

Table S2. Association Between Socio-demographic Variables and Level of Social Contacts

Variables	N	N (%) with high frequency of casual contacts (≥ 10)	Crude Risk Ratio (RR) (and 95%CI)	Adjusted RR (95%CI)
Age groups				
<2y	50	10 (20%)	0.57 (0.30 ,1.09)	
2-4y	47	14 (30%)	0.85 (0.50 ,1.45)	
5-9y	52	25 (48%)	1.37 (0.86 ,2.20)	
10-14y	43	19 (44%)	1.26 (0.80 ,1.99)	
15-24y	83	29 (35%)	ref	
25-34y	53	14 (26%)	0.76 (0.45 ,1.27)	
35-44y	49	27 (55%)	1.58 (1.11 ,2.25)	
45-54y	43	20 (47%)	1.33 (0.86 ,2.06)	
55-64y	23	7 (30%)	0.87 (0.42 ,1.80)	
65+y	47	10 (21%)	0.61 (0.32 ,1.14)	
Sex				
Female	297	98 (33%)	ref	
Male	193	77 (40%)	1.23 (0.96 ,1.59)	
Occupation/daily activity				
Pre-school child	81	22 (27%)	ref	ref
Student	132	53 (40%)	1.22 (0.80 ,1.86)	1.14 (0.74 ,1.74)
Office/Shop worker	34	19 (56%)	1.68 (1.11 ,2.54)	1.41 (0.90 ,2.21)
Agriculture/Manual work	132	47 (36%)	1.07 (0.65 ,1.76)	1.00 (0.61 ,1.62)
At home	60	11 (18%)	0.55 (0.30 ,1.00)	0.51 (0.27 ,0.94)
Other	51	23 (45%)	1.38 (0.84 ,2.26)	1.27 (0.77 ,2.08)
Day of the week				
Weekday	385	138 (36%)	ref	
Sunday	105	37 (35%)	0.92 (0.67 ,1.26)	1.04 (0.92 ,1.18)
Travel outside village/town in				
No	374	118 (32%)	ref	
Yes	116	57 (49%)	1.58 (1.23 ,2.04)	1.54 (1.18 ,2.00)

Footnote: CI: Confidence Interval; RR=Risk Ratio

Supplementary figures

Figure S1. The Reported Duration of Contact By Age Group Among Study Participants, Sheema District, January – March 2014

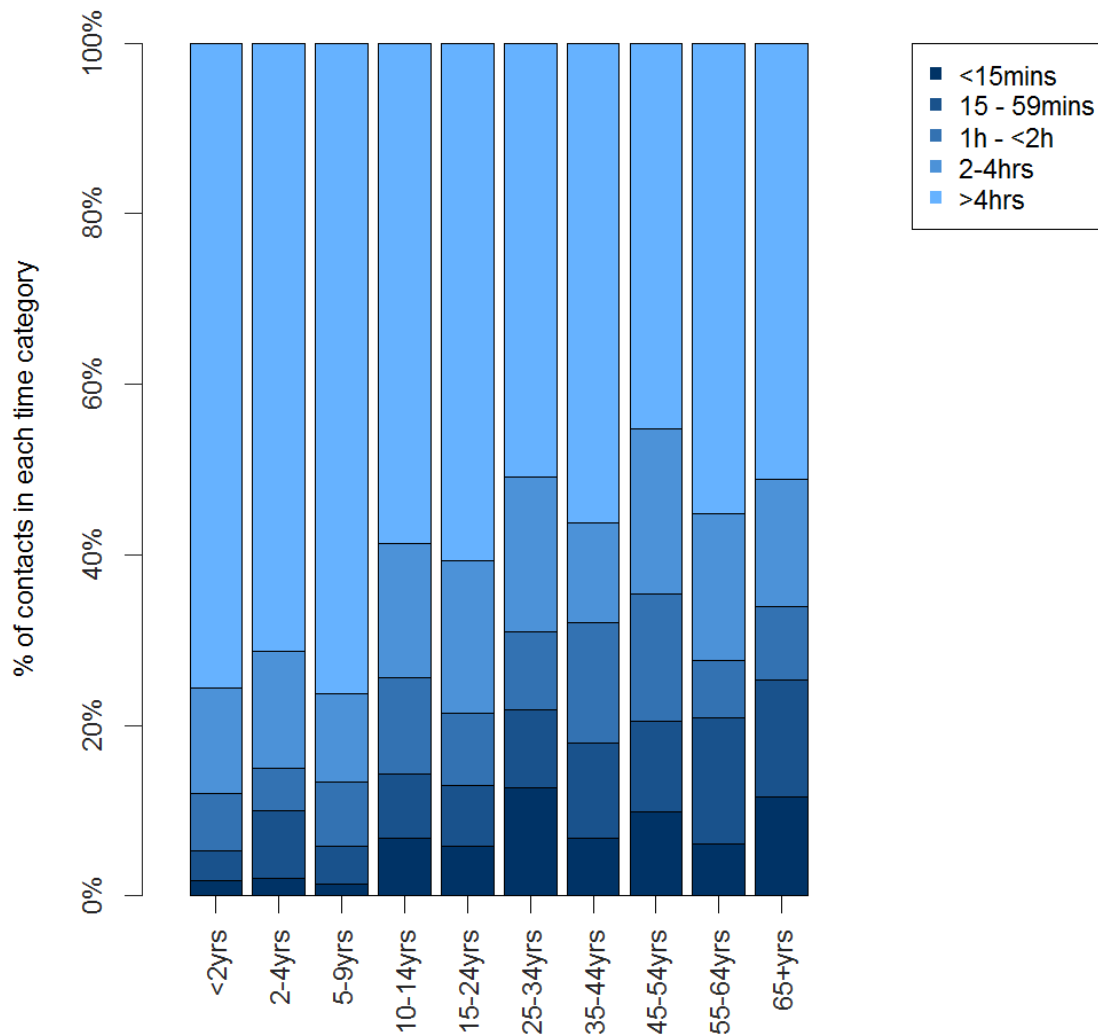


Figure S2. Reciprocity Correction of the Contact Matrices for All Contacts, Sheema District, January – March 2014. Legend: A) matrix for all reported contacts, not corrected. B) matrix for all reported contacts, corrected for reciprocity. C) Ratio of corrected over uncorrected matrices. Red cells illustrate where age-specific contacts were over-reported before correction, and blue cells under-reported. D) shows where participants significantly over-reported the number of age-specific contacts they had (upper 95% confidence bound) in red, significantly under-reported contacts in blue (lower 95% confidence bound), or where no significant adjustment was made (grey)

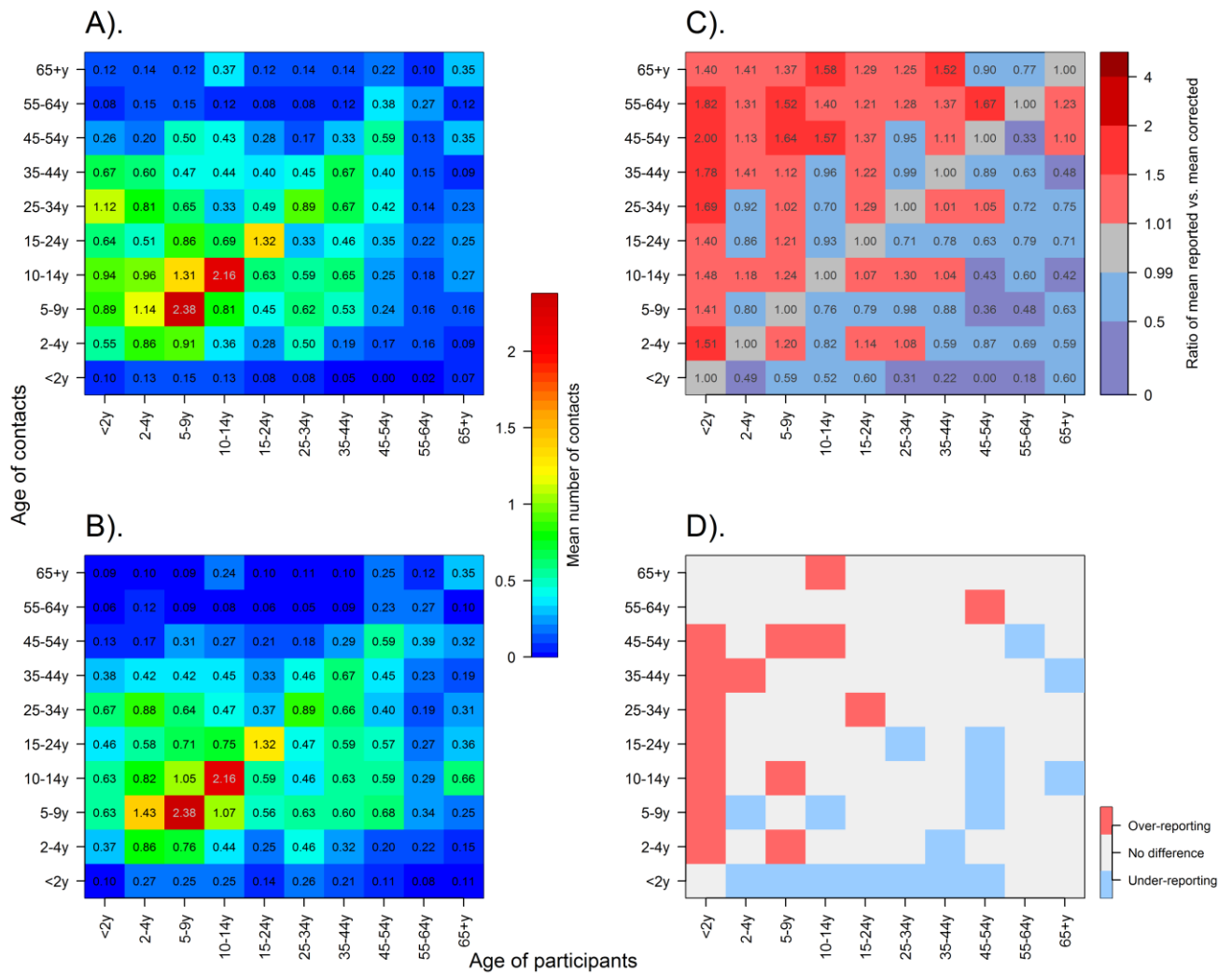


Figure S3. Epidemic Simulations Using Comparing Uganda and Great Britain, Assuming a 25% Underreporting Of Contacts In Uganda. Legend: A) Physical Contacts from Uganda, B) Physical Contacts from Great Britain, C) Proportion Infected by Age Group, D) Epidemic Size by Age group, E) Overall Proportion Infected, and F) the Basic Reproduction Number R0.

