Title page

Full title: Identifying Human Encounters that Shape the Transmission of *Streptococcus Pneumoniae* and Other Respiratory Infections

Authors: Olivier le Polain de Waroux[1], Stefan Flasche[1], Adam J Kucharski[1], Celine Langendorf[2], Donny Ndazima[3], Juliet Mwanga-Amumpaire[3,4], Rebecca F Grais[2], Sandra Cohuet[5], W John Edmunds[1]

Affiliations:

1. Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, Keppel Street, WC1E 7HT, London, United Kingdom
2. Department of Research, Epicentre, 8 Rue Saint-Sabin, 75011 Paris, France
3. Epicentre Mbarara Research Centre, PO Box 1956, Mbarara, Uganda
4. Mbarara University of Science and Technology, Mbarara University, PO Box 1410, Mbarara, Uganda
5. Department of Field Epidemiology, Epicentre, 8 Rue Saint-Sabin, 75011 Paris, France

Corresponding author: Olivier Le Polain de Waroux. Email: Olivier.lepolain@gmail.com

Abstract

Although patterns of social contacts are believed to be an important determinant of infectious disease transmission, there is little empirical evidence to back this up. Indeed, no previous study has linked individuals' risk of respiratory infection with their current pattern of social contacts. We explored whether the frequency of different types of social encounters were associated with current pneumococcal carriage and self-reported acute respiratory symptoms (ARS), though a survey in Uganda in 2014. In total 566 participants were asked about their daily social encounters and about symptoms of ARS in the last two weeks. A nasopharyngeal specimen was also taken from each participant. We found that the frequency of physical (i.e. skin-to-skin), long (≥1h) and household contacts – which capture some measure of *close* (i.e. relatively intimate) contact –, was higher among pneumococcal carriers than non-carriers, and among people with ARS compared to those without, irrespective of their age. With each additional physical encounter the age-adjusted risk of carriage and ARS increased by 6% (95%CI 2-9%) and 9% (1-18%) respectively. In contrast, the number of *casual* contacts (<5 minutes long) was not associated with either pneumococcal carriage or ARS. A detailed analysis by age of contacts showed that the number of close contacts with young children (<5 years) was particularly higher among older children and adult carriers than non-carriers, while the higher number of contacts among people with ARS was more homogeneous across contacts of all ages. Our findings provide key evidence that the frequency of *close* interpersonal contact is important for transmission of respiratory infections, but not that of *casual* contacts. Such results strengthen the evidence for public health measures based upon assumptions of what contacts are important for transmission,

and are important to improve disease prevention and control efforts, as well as inform research on infectious disease dynamics.

Author summary

Although social contacts are an important determinant for the transmission of many infectious diseases it is not clear how the nature and frequency of contacts shape individual infection risk. We explored whether frequency, duration and type of social encounters were associated with someone's risk of respiratory infection, using nasopharyngeal carriage (NP) of *Streptococcus pneumoniae* and acute respiratory symptoms as endpoints. To do so, we conducted a survey in South-West Uganda collecting information on people's social encounters, respiratory symptoms, and pneumococcal carriage status. Our results show that both pneumococcal carriage and respiratory symptoms are independently associated with a higher number of social encounters, irrespective of a person's age. More specifically, our findings strongly suggest that the frequency of *close* contacts is important for transmission of respiratory infections, particularly pneumococcal carriage. In contrast, our study showed no association with the frequency of short *casual* contacts. Those results are essential for both improving disease prevention and control efforts as well as informing research on infectious disease dynamics and transmission models.

Introduction

The transmission of respiratory infections is likely to depend on the frequency and age structure of human social contacts, as well as other factors including pre-existing immunity from prior infection or vaccination [1-3]. To understand the dynamics of such infections, studies to quantify social mixing patterns have been conducted in various settings, under the assumption that self-reported encounters reflect transmission probabilities of pathogens transmitted through close contact [2, 4-11]. Combined with disease transmission models, these data are increasingly being used to inform infection control policies [12, 13].

There is evidence from population-based models that self-reported social mixing patterns can reproduce observed aggregated seroprevalence data for chickenpox [14], mumps [4], parvovirus [15], influenza [4, 16, 17]) and whooping cough [18]. Moreover, it has been suggested that age-stratified social mixing patterns can capture individual influenza risk, as measured by a four-fold rise in neutralization titres over the course of an epidemic (17, 18). However, it remains unclear precisely how risk of infection is related to the frequency and nature of an individual's social encounters around the time of infection.

To establish how social behaviour shapes individual-level infection, we explored whether the frequency and duration of different types of social encounters were associated with an individual's risk of respiratory infection, using nasopharyngeal (NP) carriage of *Streptococcus pneumoniae* (the pneumococcus) and self-reported acute respiratory symptoms (ARS) as endpoints. *S.pneumoniae* is one of the main causes of pneumonia and sepsis globally [19], disproportionally so in low-income settings [19-21]. Colonization of the nasopharynx is a precondition to disease, and the main source of human-to-human transmission. Given that

105 most episodes of carriage remain asymptomatic, social behaviour is unlikely to change as a
106 result of carriage, making pneumococcal carriage a more suitable endpoint than
107 symptomatic illness to explore the association between social behaviour and infection risk,
108 given that people tend to limit their contacts during symptomatic illness [22]. In addition, as
109 natural immunity to carriage is weak [23], pre-existing immunity is less likely to confound
110 associations between disease and social contact patterns than in studies using immunizing
111 infections as biological endpoint [16, 17].

112 We analysed data from a social contact study nested within a survey of NP carriage of
113 *S.pneumoniae* conducted across a rural South-West Uganda in 2014, before the introduction
114 of the pneumococcal conjugate vaccine (PCV). Participants across age groups were
115 randomly selected from 56 villages and one small town. Participants were asked to name all
116 individuals with whom they had a conversational encounter lasting ≥ 5 minutes between
117 wake-up on the day prior to the survey day to wake-up on the survey day (~ 24 hours).
118 Those encounters are further referred to as *'ordinary'* contacts. For very short contacts (i.e.
119 <5 minutes), which we here define as *'casual'* contacts, an estimate of the number of
120 contacts was asked (<10, 10-19, 20-29, ≥30) without further details. ARS were defined as
121 any of the following symptoms in the previous two weeks: sore throat, sneezing, difficulty
122 breathing, and runny nose. A nasopharyngeal specimen was taken from each survey
123 participant and was processed and analysed as per WHO guidelines [24]. Using these
124 individually-matched data, we examined whether the type and frequency of social contacts
125 differed by carriage status and between individuals with and without ARS.

126

Results

Study population

Of the 687 individuals initially targeted for inclusion, 568 (83%) individuals from unique households responded to the survey, with data on both social contacts and carriage or ARS available from 566 participants. Participant's age spanned across age groups and the sex distribution was reasonably balanced (58% female). The majority (98%) of children aged 6 – 15 years attended school or college. More than a third of all adults (36%) worked in agriculture and 22% were homemakers/housewife.

On average, people made seven *'ordinary'* contacts (defined as contacts ≥5 minutes long), ranging from 0 to 25, the majority of which were physical (i.e. 'skin-to-skin' contact or indirect physical contact through utensils passed from mouth-to-mouth). There was no evidence that the average number of contacts differed by weekday or between weekdays and weekends (P=0.623). Children aged 5-9 years reported most contacts and children <5 years the fewest (Figure 1). The majority of contacts made by children were physical.

Figure 1: Number of reported contacts by age group and type of contact. Legend: Boxplots showing the median (horizontal line), interquartile range (boxes) and 95% confidence interval around the mean (whiskers). The outliers are shown as bubbles above or under the whiskers

148 The most intense mixing tended to be between individuals of the same age group (i.e.
149 assortative mixing), but there was also substantial mixing between age groups. Contact from
150 and with children aged <10 years involved proportionally more physical touch than contacts
151 between older children and adults (Figure 2A and 2B). The proportion of non-physical
152 contacts, contacts outside the household and contacts of short (<1h) duration was higher
153 among teenagers and adults than among younger children and infants.

154 There was no difference in contact patterns by sex, for all types of contacts considered.

155 Four hundred and ninety (87%) participants estimated how many casual contacts (i.e. <5
156 minutes in duration) they had the day before the survey. Over a third (36%) of these
157 reported 10 or more casual contacts, and 11% reported more than 20 casual contacts.
158 Among the 13% who could not estimate how many casual contacts they had, there were
159 proportionally more children with over half (56%) under 10 years. The mean number of
160 reported 'ordinary' contacts was higher among individuals reporting ≥10 casual contacts
161 than those reporting fewer than 10 contacts (mean 8.9 vs 5.9, P<0.001), as well as among
162 the 78 individuals who did not know how many casual contacts they had (mean 8.8,
163 P<0.001).

164 The prevalence of pneumococcal carriage was strongly age dependent, decreasing from 75%
165 in children <5 years to 46% among 5 – 9 year olds, 17% in 10 – 19 year olds, and further
166 decreasing to 8% and 7% among 20-39 years and ≥40 years old respectively (Figure 2C).
167 Overall, 72 (13%) people reported having suffered from ARS in the two weeks prior to the
168 survey. The prevalence of ARS varied much less with age than that of carriage (Figure 2D),
169 ranging from 20% among 5 to 9 year olds to 8% among ≥40 years old. There was no sex

difference in the prevalence of carriage (age-adjusted RR for males 0.97, P=0.814) or ARS (age-adjusted RR for males 1.40, P=0.134). Carriage and ARS were poorly correlated (Pearson's correlation coefficient R=0.09, ranging from -0.08 to 0.29 by age group), and there was no evidence that the risk of ARS was higher among carriers compared to non-carriers (age-adjusted relative risk (RR) 1.06 (95%CI 0.85 – 1.33)).

Figure 2: Contact matrices and prevalence of *S.pneumoniae* carriage and Acute Respiratory Symptoms (ARS), by age. Legend: Contact matrices and prevalence of *S.pneumoniae* carriage (Panel A) and Acute Respiratory Symptoms (Panel B). In panels C and D, the height of each bar corresponds to the point prevalence and the error bar represents the 95% Confidence Interval (CI).

Social contacts as a risk for pneumococcal carriage or ARS

Overall, the mean number of contacts among carriers was significantly higher than non-carriers, and this observation was consistent across age groups, although most differences were not statistically significant due to small numbers (Figure 3). In particular, carriers had more physical contacts (Fig 3A). This pattern was also consistent for ARS, with the exception of 5 – 9 year olds in whom the mean number of contacts among individuals with ARS was lower for physical contacts (Fig 3B).

Figure 3: Mean number of contacts by age group and nasopharyngeal carriage status (Panel A) or ARS status (Panel B). Legend: The graph shows the mean number of contacts among carriers (red, panel A), non-carriers (blue, panel A), individuals reporting ARS (red, panel B), and individuals without ARS (blue, panel B). This is shown by age group and for all contacts, physical contacts, household contacts and contacts lasting over 1 hour.

In univariable analysis, the risk of carriage increased with all contacts, household contacts, contacts ≥1 hour long and physical contacts. The latter had the largest effect size, with a 13% increased risk for each additional contact reported by participants (Table 1). Physical contacts and contacts ≥1 hour were strongly correlated (R = 0.76), particularly among children <5 years (R=0.85) and children aged 5 – 9 years (R=0.87), hence their effect could not be disentangled, whereas the correlation between physical contacts and household contacts was moderate (R=0.61, ranging from 0.47 to 0.70 between age groups). After age adjustment, physical contacts or contacts ≥1 hour remained most significantly associated with carriage, with a 6% increased risk (95%CI 2 – 10%) for each unit increase in the number of reported contacts (Table 1). We found good evidence that the number of household contacts increased the risk as well (Table 1). There was no confounding effect by other covariates and models were therefore only adjusted for age (Supporting Information S1).

An increase in physical, household and long (≥ 1h) contacts, were also associated with an increased risk of ARS in univariable analysis. There was little or no evidence of a confounding effect of age, given the more constant prevalence of ARS across age groups (Table 1). Unlike pneumococcal carriage, however, the relative risk was more constant across types of contacts, and household contacts rather than physical contacts were most

strongly associated with a risk for ARS, with a risk increase of 9% (1 – 18%) for each additional reported contact.

Table 1: Relative risk of ***S.pneumoniae*** **carriage and ARS** by frequency of contact

Contact types	Mean contacts		Crude RR (95%CI)	Adjusted RR (95%CI)
NP carriage	NP+	NP-		
All contacts	8.09	7.53	1.04 (1.00; 1.07)*	1.03 (1.00; 1.07)*
Physical contacts	6.83	6.09	1.13 (1.09; 1.17)*	1.06 (1.02; 1.09)*
Non-physical contacts	1.25	1.43	0.78 (0.72; 0.86)*	0.97 (0.91; 1.05)
Household contacts	5.64	5.31	1.09 (1.03; 1.14)*	1.04 (0.99 ; 1.10)
Non-household	2.22	2.45	0.98 (0.93; 1.04)	1.03 (0.98; 1.09)
Contacts≥ 1hour	7.50	6.56	1.08 (1.05; 1.12)*	1.06 (1.02; 1.10)*
Contacts <1h	2.38	2.62	0.81 (0.73; 0.92)*	0.94 (0.85; 1.03)
ARS	ARS+	ARS-		
All contacts	8.18	7.63	1.07 (1.02; 1.12)*	1.07 (1.02; 1.13)*
Physical contacts	6.10	6.26	1.09 (1.03; 1.15)*	1.07 (1.00; 1.14)*
Non-physical contacts	2.07	1.36	1.00 (0.89; 1.10)	1.06 (0.95; 1.18)
Household contacts	5.70	5.31	1.10 (1.02; 1.19)*	1.09 (1.01; 1.18)*
Non-household	2.47	2.31	1.04 (0.98; 1.11)	1.05 (0.98; 1.13)
Contacts ≥ 1hour	7.33	6.74	1.07 (1.01; 1.13)*	1.07 (1.00; 1.14)*
Contacts <1 hour	3.14	2.58	1.05 (0.96; 1.15)	1.08 0.99; 1.18)

Legend: *=associations significant at P<0.05. RR= Risk Ratio. CI= Confidence Interval. NP+= pneumococcal carrier, NP - = non-carrier, ARS+ = with ARS, ARS- = without ARS

Next, we analysed whether the number of casual contacts (i.e. contacts lasting < 5 minutes) was associated with either pneumococcal carriage or ARS. We found no evidence that the prevalence of pneumococcal carriage or the risk of ARS were associated with reporting higher levels of casual contacts, as shown in Table 2. Given the small numbers of individuals reporting ≥30 casual contacts, we pooled the 56 individuals reporting ≥20 casual contacts into one category.

In univariable analysis the risk of pneumococcal carriage was higher in the group of 78 participants who did not know how many casual contacts they may have had. This was due to the higher proportion of children <5 years and aged 5 – 9 years in that group, however, after age-adjustment, there was no evidence that the risk of pneumococcal carriage was higher in that group. Similarly, a higher number of casual contacts was associated with increased risk of ARS, and with no confounding effect of age or other variables on the estimates.

We found no confounding effect on casual contacts on the relative risk of carriage or ARS as a function of the frequency of reported 'ordinary' contacts.

238 Table 2: Relative risk (RR) of pneumococcal carriage and ARS by level of reported number of

239 daily casual contacts

		Pneumococcal carriage		Acute Respiratory Symptoms (ARS)	
Number of casual contacts	Number of individuals	Crude RR (95%CI)	Age-adjusted RR (95%CI)	Crude RR (95%CI)	Age-adjusted RR (95%CI)
0 – 9	315	ref	ref	ref	ref
10 – 19	119	0.82 (0.52; 1.27)	0.80 (0.55; 1.75)	1.02 (0.65; 1.49)	1.00 (0.65; 1.54)
≥20	56	0.81 (0.44; 1.49)	1.19 (0.70; 2.04)	0.51 (0.21; 1.27)	0.56 (0.22; 1.39)
Not known	76	1.52 (1.00; 2.31)*	1.07 (0.77; 1.49)	1.01 (0.45; 2.28)	0.88 (0.38; 2.00)

240 Legend: * = associations significant at P<0.05. RR= Risk Ratio. CI= Confidence Interval

241

242 Finally, we explored the characteristics of age-specific mixing patterns by ARS or carriage
243 status in greater detail, with a focus on physical contacts and household contacts. We
244 computed the ratio of the mean number of contacts within and between age groups among
245 carriers compared to non-carriers, and among individuals with ARS compared to those
246 without. The average number of physical encounters within and between age groups
247 tended to be higher for carriers than non-carriers in most instances, albeit with substantial
248 uncertainty owing to small numbers (Figure 4). Carriers reported more contacts with
249 children < 5 years, and particularly adult carriers who reported on average more than twice
250 as many physical contacts with children under five than non-carriers (Figure 4). Given that
251 most of such contacts occurred within the household, the effect of physical and household
252 contacts with children <5 years was indistinguishable, whereas for older age groups, the

253 association with physical contact was stronger than that with household contacts (Figure 4).
254 Similar findings were seen for ARS, however, unlike for pneumococcal carriage,
255 symptomatic adults did not have more contacts with young children than asymptomatic
256 ones. Results based on absolute differences in the mean number of contacts rather than
257 relative means are displayed in Figure S1 (Supporting Information), showing the similar
258 associations than the reported ratios, but providing a quantified difference in mean number
259 of contacts instead.

260

261 Figure 4: Ratio of the mean reported number of *physical* contacts between pneumococcal
262 carriers and non-carries (panel A) and individuals with ARS compared to those without ARS
263 (Panel B). Legend: Matrices of the ratio of the mean number of contacts among
264 pneumococcal carriers compared to non-carriers (Panel A) and individuals with ARS
265 compared to non-ARS (panel B). The numbers represent the point estimate of the ratio and
266 in brackets the 95% confidence bounds. Blue are cells with a point estimate <1 and in red
267 with point estimates >1.

268

269 **Discussion**

270 Our study provided a unique opportunity to explore whether and how social contact
271 patterns are associated with someone's risk of acute respiratory infection in a mostly rural
272 East African setting. Our results show that people who tend to have more frequent close
273 contacts are more likely to be pneumococcal carriers or to report acute respiratory
274 symptoms, irrespective of their age. In contrast, we found that less intimate or short casual

contacts were not associated with someone's infection risk, suggesting that social contacts important for transmission are close interpersonal encounters.

Existing evidence on the association between contact patterns and risk of infection is mostly 'ecological', with very few studies based on individual-level data. Using data of influenza A/H1N1 seroconversion in Hong Kong coupled to social contact data in the same population, Kwok et al. [16] showed that age rather than social contact patterns were the main driver of the individual risk of infection in that setting, and further work by Kucharski et al. [17] on the same data further supported the finding that someone's risk of infection is related to the average mixing pattern within their age group rather than their reported number of contacts. However, the validation of the 'social contact hypothesis' with such data remains difficult, due to challenges in accounting for acquired immunity in the population, assumptions around stability of behaviour over relatively long time periods as well as challenges in capturing influenza infection events based on serological data only [16, 17].

Studying nasopharyngeal carriage of *S.pneumoniae* as the main biological endpoint enabled us to address many of these issues, for several reasons. First, the high prevalence of carriage provided the statistical power required to study individually-matched acquisition risk in our study. Second, a person's individual behaviour is unlikely to be affected by carriage given that the vast majority of episodes remain asymptomatic, in contrast to symptomatic respiratory infections during which social behaviour might change as a result of illness [22]. Most carriage episodes in our study were asymptomatic, and our decision to assess the risk for ARS and NP carriage separately stemmed from that observation. In addition, given that duration of carriage is relatively short – at most 3 months in young children and no longer

298 than a few weeks in adults [25] –, and as no difference in our survey was observed in the
299 mean number of contacts between days of the weeks and between survey weeks, it is
300 reasonable to assume that contact patterns measured on a given day reflect contact
301 patterns around the time of pneumococcal acquisition. Finally, as colonisation results in
302 weak protective immune responses and limited reduction in serotype-specific reacquisition
303 risk [23, 26], and with over 90 circulating serotypes, natural acquired immunity is unlikely to
304 distort the association between carriage and social contact patterns, unlike immunizing
305 infection for which individuals with more frequent risk of infection due to their social
306 contact patterns are also more likely to be immune.

307 In contrast, self-reported respiratory symptoms may be influenced by factors such as
308 behaviour change in illness [22] or immunity. The definition itself also presents limitations;
309 although respiratory viruses such as the Respiratory Syncytial Virus (RSV), adenovirus,
310 parainfluenza and influenza viruses are likely to account for a large proportion of ARS cases
311 [27], the definition was used as a non-specific proxy for acute respiratory infection and may
312 have captured other infectious and non-infectious conditions. Notwithstanding such
313 caveats, results for ARS were similar to those for nasopharyngeal carriage, but also showed
314 a more consistent association with all types of non-casual contacts, rather than physical or
315 contacts of long duration only. This suggests that the definition may have mostly captured
316 acute infections, and also provides further evidence that the number of close interpersonal
317 contacts, and particularly household contacts, plays a role in the transmission of acute
318 respiratory pathogens.

319 One of the striking features of our analysis of the relative number of mean contacts
320 between age groups is that adults colonised with *S.pneumoniae* reported more than twice

as many close contacts with children under five than non-carriers. This is in agreement with observational and modelling studies showing that pneumococcal carriage risk increases with household size and with the number of children <5 years in the household [28]. It also supports the finding that carriage acquisition in adults occurs mostly within the household as a result of contact with young children who are drivers of infection [29]. In contrast, however, we found that adults with ARS reported an equal or lower number of contacts with children <5 years than asymptomatic adults. This likely reflects the epidemiological differences between carriage and ARS, given the very high prevalence of carriage among <5 year olds and the much less marked difference in age-specific prevalence of ARS across age groups, in addition to other potential factors such as acquired immunity among adults more frequently exposed to very young children. Some of the specific details of the contact patterns and differences between *S.pneumoniae* and ARS are harder to interpret, due in part to wide statistical uncertainty, particularly among adults in whom the number of carriers or symptomatic individuals is small. However, overall the results from our analysis support the general finding that increased close contacts are associated with higher risk of ARS or pneumococcal carriage.

There is still much debate about how respiratory pathogens are transmitted from person to person; whether through close direct or indirect physical contact, through large droplet transmission at close range (<1 meter), or through aerosolized particles floating over longer distances, particularly in poorly ventilated indoor settings [30, 31]. It is likely that many pathogens can be transmitted through a combination of these routes. Yet the contribution of each mechanism remains uncertain [30, 31]. It is generally assumed that the main transmission route of *S.pneumoniae* is through direct contact [32], as well as indirectly

344 through shared glasses or bottles [33]. Analogously, for other colonising bacteria such as
345 *N.meningitidis*, close contact and intimate kissing are known risk factors among teenagers
346 and young adults [34]. Similarly, it is believed that influenza and other respiratory viruses
347 are primarily transmitted through direct contact or contact with large droplet transmission
348 at close range rather than aerosolized particles [30].

349 Although our objective was not to demonstrate transmission, our findings strongly support
350 that direct close interpersonal contact is an important mode of transmission for pathogenic
351 bacteria of the nasopharynx as well as respiratory viruses, and strengthens the scientific
352 evidence for public health measures based upon these assumptions, such as hand washing
353 campaigns or chemoprophylaxis of close contacts of cases of meningococcal meningitis to
354 name a few. The strong association of ARS with household and other close contacts,
355 independently from physical touch, might suggest that other mechanisms such as indirect
356 transmission through fomites or aerosol transmission may play a role. Elucidating the
357 contribution of such factors in this context would be an important question for future
358 research.

359 Our findings have also several implications for infectious disease research. Contact
360 structures are central to transmission models, and appropriate assumptions about what
361 type of contact drives infectious disease transmission are essential. Our results suggest that
362 the parameterisation of transmission models of *S.pneumoniae* and similar pathogens using
363 mixing matrices based on physical or another measure of *close* interpersonal contact would
364 more likely capture relevant contact patterns than those based on any type of social
365 encounter. This has also implications for the design of contact studies, particularly in low-
366 income settings given the scarcity of published data currently available and the need to

collect additional data from many more settings [7, 8]. In contrast to diary-based approaches that have been used by many [2, 4, 7, 8], the study design here was relatively simple and only involved a single face-to-face interview. A drawback of such retrospective approach is the lack of detailed information about very short (i.e. 'casual') contacts, as such information was deemed unreliable – and this is further supported by evidence that short contacts tend to be inconsistently recorded even in prospective diary-based approached [1]. However, given that very short contacts may not account for much of the transmission, as our study suggests, a more simple retrospective design is a potential attractive option in other settings where contact data are lacking, and in which data collection through more comprehensive diary-based approaches may be difficult to implement.

There are some additional limitations to our study. We were unable to explore other potential confounding factors, such as bed share, ventilation, indoor smoke or hand washing [35], and the contribution of such factors should be explored further. Moreover, it remains unclear to what extent our results can be generalised to any acute respiratory infection. For example, factors such as ventilation and airflow may be of particular importance for aerosolized transmission of pathogens such as mycobacteria [7], compared to influenza or *S.pneumoniae* [30, 31], and whether household or physical contacts reflect contact patterns important for aerosolized transmission remains uncertain [7]. Finally, while we found a strong association at the individual level, our study does not demonstrate causality.

However, in the absence of detailed longitudinal data on acquisition events between all individuals' contacts —which would be challenging and possibly unrealistic to obtain— our findings provide consistent evidence of a 'dose-response' association at the individual level between close social encounters and acquisition risk for respiratory pathogens, and

391 therefore provides robust support both for the social contact hypothesis, and for research
392 and policy work based upon this hypothesis.

393

394 **Materials and Methods**

395 Data collection

396 The study was conducted in Sheema North Sub-District (Sheema district, South-West
397 Uganda) between January and March 2014. Sixty clusters were randomly selected from the
398 215 villages and two small district towns (Kabwohe and Itendero) in the study area,
399 proportionally to the population size of each village and town. In each cluster 11 or 12
400 individuals were randomly sampled from different households to both answer questions
401 about their social contacts and their history of respiratory illness in the last two weeks, as
402 well as having a nasopharyngeal swab taken. A household was defined as the group of
403 individuals living under the same roof and sharing the same kitchen on a daily basis.

404 For the social contact questionnaire, participants were first asked to list all the individuals
405 with whom they had a two-way conversational contact lasting for ≥5 minutes during a
406 period of approximately 24 hours prior to the survey day (from wake up the previous day
407 until wake up on the survey day). Such encounters were defined as 'ordinary contacts'. For
408 each reported ordinary contact, participants (or their parent/guardian) were asked to
409 estimate the contact's age (or estimated age), how long the encounter lasted for and
410 whether it involved skin-to-skin touch or utensils passed from mouth to mouth (either of
411 those defining 'physical contacts'). For very short social encounters (<5 minutes), which

412 were defined as 'casual contacts' (e.g. seeing someone on the way, encounter in a shop
413 etc.), participants were asked to estimate the number of encounters based on pre-defined
414 categories (<10 contacts, 10-19 contacts, 20-29 contacts, ≥30 contacts), but not to provide
415 further details about each contact.

416 Next, participants were asked about respiratory symptoms experienced in the two weeks
417 prior to the survey, including any of the following: cough, runny nose, sneezing, sore throat,
418 difficulty breathing.

419 Finally, after the interview was completed, a nasopharyngeal swab was taken from each
420 participant. NP samples were collected, transported and analysed as per WHO guidelines
421 [24]. NP swabs (flocked nylon swabs, COPAN, Italy) were inoculated in a skim milk tryptone-
422 glucose-glycerol (STGG) medium, transported in cool boxes and frozen at the research
423 laboratory at -20°C within 8 hours of collection. Specimens were inoculated onto a selective
424 agar plate of 5 mg/L gentamicin-Columbia agar with 5% sheep blood and incubated at 37°C
425 in 5% CO_2 atmosphere overnight. Pneumococcal identification was based on optochin
426 susceptibility testing of all alpha-hemolytic colonies and bile solubility testing in case of
427 intermediate susceptibility to optochin.

428 Statistical analysis

429 We first performed descriptive analyses, with age-specific probability weights to account for
430 different inclusion probabilities by age at the design stage, and adjusted for the clustering
431 by village through the use of clustered 'sandwich' variance estimators to account for
432 possible correlation within each of the sixty clusters [36]. We explored social contact
433 patterns within and between age groups through contact matrices. In such matrices we

434 report the mean number of ordinary contacts of participants in age group j with contacts
435 in age group i (m_{ij}), adjusted for reciprocity, as in Melegaro et al. [15].

436 We then analysed whether and how the frequency distribution of contacts was associated
437 with pneumococcal carriage or self-reported ARS.

438 We modelled the effect of 'ordinary' contacts (defined as contacts ≥5 minutes long) on
439 carriage or respiratory symptoms as a function of contact frequency, through a Poisson
440 model with a robust variance estimator, and inclusion probability weights by age group. We
441 treated contacts as continuous variables, but assessed departure from linearity through
442 likely ratio tests and model comparisons of Bayesian Information Criterion (BIC). In
443 multivariable analysis, we considered for inclusion any covariate significantly associated
444 with the outcome at P<0.10 in univariable analysis. Model improvement was considered for
445 any decrease in the BIC. Further details are provided in the Supporting Information S1.

446 Next, we used the same analytical approach to assess whether carriage and ARS were
447 associated with the level of self-reported 'casual' contacts (i.e. <5 minutes long).

448 Finally, we explored differences in the social mixing matrices by status of pneumococcal
449 carriage and ARS. To do so, we computed the ratio of the mean number of reported
450 contacts by participants j with contacts i among carriers compared to non-carriers (R_{ij}^{C}) or
451 symptomatic compared to asymptomatic individuals (R_{ij}^{R}), such that $R_{ij}^{C} = \frac{m_{ij}^{C}}{m_{ij}^{N-C}}$ and
452 $R_{ij}^{R} = \frac{m_{ij}^{R}}{m_{ij}^{N-R}}$, where C = carriers, R = participants with ARS and N = the total number of
453 participants. Estimates were not adjusted for reciprocity given that subpopulations were

not closed (e.g. contacts of carriers may not be carriers), and that our aim was to compare reported number of contacts rather than calculate a contact matrix. The uncertainty in reported values and ratios was obtained through resampling techniques, drawing random samples from each m_{ij}, with the number of draws equal to the study population in each age group j. Ratios and uncertainty around them was obtained from the ratio of bootstrapped matrices.

The same approach as described above was used to compute the differences in m_{ij} with $D_{ij}^{C} = m_{ij}^{C} - m_{ij}^{N-C}$ and $D_{ij}^{C} = m_{ij}^{R} - m_{ij}^{N-R}$.

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

Ethics

Approval was obtained from the Ethical review boards of the London School of Hygiene and Tropical Medicine, Médecins Sans Frontières, the Faculty of Medicine Research & Ethics Committee of the Mbarara University of Science and Technology, the Institutional Ethical Review Board of the MUST, and the Uganda National Council for Science and Technology.

Acknowledgments

We thank all the surveyors for their work and all the individuals who took part in this study, Rinah Arinaitwe for her help in managing the data entry, Dan Nyehangane for his support with the laboratory analyses, and Dr Yap Boum II for his help overseeing field activities. We

are also grateful to Prof Anthony Scott, Dr Fabienne Nackers and Prof Francesco Checchi for their valuable input in the design and protocol of the field surveys.

Conflict of interest

The authors declare they have no conflict of interest.

Funding

This work was supported by Médecins Sans Frontières International Office, Geneva, Switzerland. Epicentre received core funding from Médecins Sans Frontières. OLP was supported by the AXA Research Fund through a Doctoral research fellowship. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. Smieszek T, Burri EU, Scherzinger R, Scholz RW. Collecting close-contact social mixing data with contact diaries: reporting errors and biases. Epidemiology and infection. 2012;140(4):744-52. doi: 10.1017/S0950268811001130. PubMed PMID: 21733249.
2. Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. PLoS medicine. 2008;5(3):e74. doi: 10.1371/journal.pmed.0050074. PubMed PMID: 18366252; PubMed Central PMCID: PMC2270306.
3. Eubank S, Guclu H, Kumar VS, Marathe MV, Srinivasan A, Toroczkai Z, et al. Modelling disease outbreaks in realistic urban social networks. Nature. 2004;429(6988):180-4. doi: 10.1038/nature02541. PubMed PMID: 15141212.
4. Wallinga J, Teunis P, Kretzschmar M. Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents. American journal of epidemiology. 2006;164(10):936-44. doi: 10.1093/aje/kwj317. PubMed PMID: 16968863.
5. Horby P, Pham QT, Hens N, Nguyen TT, Le QM, Dang DT, et al. Social contact patterns in Vietnam and implications for the control of infectious diseases. PloS one. 2011;6(2):e16965. doi: 10.1371/journal.pone.0016965. PubMed PMID: 21347264; PubMed Central PMCID: PMC3038933.
6. Grijalva CG, Goeyvaerts N, Verastegui H, Edwards KM, Gil AI, Lanata CF, et al. A household-based study of contact networks relevant for the spread of infectious diseases in the highlands of Peru. PloS one. 2015;10(3):e0118457. doi: 10.1371/journal.pone.0118457. PubMed PMID: 25734772; PubMed Central PMCID: PMC4348542.
7. Johnstone-Robertson SP, Mark D, Morrow C, Middelkoop K, Chiswell M, Aquino LD, et al. Social mixing patterns within a South African township community: implications for respiratory

disease transmission and control. American journal of epidemiology. 2011;174(11):1246-55. doi: 10.1093/aje/kwr251. PubMed PMID: 22071585; PubMed Central PMCID: PMC3224253.

8. Kiti MC, Kinyanjui TM, Koech DC, Munywoki PK, Medley GF, Nokes DJ. Quantifying age-related rates of social contact using diaries in a rural coastal population of Kenya. PloS one. 2014;9(8):e104786. doi: 10.1371/journal.pone.0104786. PubMed PMID: 25127257; PubMed Central PMCID: PMC4134222.

9. Fu YC, Wang DW, Chuang JH. Representative contact diaries for modeling the spread of infectious diseases in Taiwan. PloS one. 2012;7(10):e45113. doi: 10.1371/journal.pone.0045113. PubMed PMID: 23056193; PubMed Central PMCID: PMC3463600.

10. Read JM, Lessler J, Riley S, Wang S, Tan LJ, Kwok KO, et al. Social mixing patterns in rural and urban areas of southern China. Proceedings Biological sciences / The Royal Society. 2014;281(1785):20140268. doi: 10.1098/rspb.2014.0268. PubMed PMID: 24789897; PubMed Central PMCID: PMC4024290.

11. Edmunds WJ, O'Callaghan CJ, Nokes DJ. Who mixes with whom? A method to determine the contact patterns of adults that may lead to the spread of airborne infections. Proceedings Biological sciences / The Royal Society. 1997;264(1384):949-57. doi: 10.1098/rspb.1997.0131. PubMed PMID: 9263464; PubMed Central PMCID: PMC1688546.

12. Halloran ME, Ferguson NM, Eubank S, Longini IM, Jr., Cummings DA, Lewis B, et al. Modeling targeted layered containment of an influenza pandemic in the United States. Proceedings of the National Academy of Sciences of the United States of America. 2008;105(12):4639-44. doi: 10.1073/pnas.0706849105. PubMed PMID: 18332436; PubMed Central PMCID: PMC2290797.

13. Baguelin M, Flasche S, Camacho A, Demiris N, Miller E, Edmunds WJ. Assessing optimal target populations for influenza vaccination programmes: an evidence synthesis and modelling study. PLoS medicine. 2013;10(10):e1001527. doi: 10.1371/journal.pmed.1001527. PubMed PMID: 24115913; PubMed Central PMCID: PMC3793005.

14. Santermans E, Goeyvaerts N, Melegaro A, Edmunds WJ, Faes C, Aerts M, et al. The social contact hypothesis under the assumption of endemic equilibrium: Elucidating the transmission potential of VZV in Europe. Epidemics. 2015;11:14-23. doi: 10.1016/j.epidem.2014.12.005. PubMed PMID: 25979278.

15. Melegaro A, Jit M, Gay N, Zagheni E, Edmunds WJ. What types of contacts are important for the spread of infections?: using contact survey data to explore European mixing patterns. Epidemics. 2011;3(3-4):143-51. doi: 10.1016/j.epidem.2011.04.001. PubMed PMID: 22094337.

16. Kwok KO, Cowling BJ, Wei VW, Wu KM, Read JM, Lessler J, et al. Social contacts and the locations in which they occur as risk factors for influenza infection. Proceedings Biological sciences / The Royal Society. 2014;281(1789):20140709. doi: 10.1098/rspb.2014.0709. PubMed PMID: 25009062; PubMed Central PMCID: PMC4100506.

17. Kucharski AJ, Kwok KO, Wei VW, Cowling BJ, Read JM, Lessler J, et al. The contribution of social behaviour to the transmission of influenza A in a human population. PLoS pathogens. 2014;10(6):e1004206. doi: 10.1371/journal.ppat.1004206. PubMed PMID: 24968312; PubMed Central PMCID: PMC4072802.

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

19. O'Brien KL, Wolfson LJ, Watt JP, Henkle E, Deloria-Knoll M, McCall N, et al. Burden of disease caused by Streptococcus pneumoniae in children younger than 5 years: global estimates. Lancet. 2009;374(9693):893-902. doi: 10.1016/S0140-6736(09)61204-6. PubMed PMID: 19748398.

20. Black RE, Cousens S, Johnson HL, Lawn JE, Rudan I, Bassani DG, et al. Global, regional, and national causes of child mortality in 2008: a systematic analysis. Lancet. 2010;375(9730):1969-87. doi: 10.1016/S0140-6736(10)60549-1. PubMed PMID: 20466419.

21. Cohen AL, Hyde TB, Verani J, Watkins M. Integrating pneumonia prevention and treatment interventions with immunization services in resource-poor countries. Bulletin of the World Health

Organization. 2012;90(4):289-94. doi: 10.2471/BLT.11.094029. PubMed PMID: 22511825; PubMed Central PMCID: PMC3324866.

22. Van Kerckhove K, Hens N, Edmunds WJ, Eames KT. The impact of illness on social networks: implications for transmission and control of influenza. American journal of epidemiology. 2013;178(11):1655-62. doi: 10.1093/aje/kwt196. PubMed PMID: 24100954; PubMed Central PMCID: PMC3842903.

23. Prevaes SM, van Wamel WJ, de Vogel CP, Veenhoven RH, van Gils EJ, van Belkum A, et al. Nasopharyngeal colonization elicits antibody responses to staphylococcal and pneumococcal proteins that are not associated with a reduced risk of subsequent carriage. Infection and immunity. 2012;80(6):2186-93. doi: 10.1128/IAI.00037-12. PubMed PMID: 22451514; PubMed Central PMCID: PMC3370583.

24. Satzke C, Turner P, Virolainen-Julkunen A, Adrian PV, Antonio M, Hare KM, et al. Standard method for detecting upper respiratory carriage of Streptococcus pneumoniae: updated recommendations from the World Health Organization Pneumococcal Carriage Working Group. Vaccine. 2013;32(1):165-79. doi: 10.1016/j.vaccine.2013.08.062. PubMed PMID: 24331112.

25. Abdullahi O, Nyiro J, Lewa P, Slack M, Scott JA. The descriptive epidemiology of Streptococcus pneumoniae and Haemophilus influenzae nasopharyngeal carriage in children and adults in Kilifi district, Kenya. The Pediatric infectious disease journal. 2008;27(1):59-64. doi: 10.1097/INF.0b013e31814da70c. PubMed PMID: 18162940; PubMed Central PMCID: PMC2382474.

26. Lipsitch M, Abdullahi O, D'Amour A, Xie W, Weinberger DM, Tchetgen Tchetgen E, et al. Estimating rates of carriage acquisition and clearance and competitive ability for pneumococcal serotypes in Kenya with a Markov transition model. Epidemiology. 2012;23(4):510-9. doi: 10.1097/EDE.0b013e31824f2f32. PubMed PMID: 22441543; PubMed Central PMCID: PMC3670084.

27. Berkley JA, Munywoki P, Ngama M, Kazungu S, Abwao J, Bett A, et al. Viral etiology of severe pneumonia among Kenyan infants and children. Jama. 2010;303(20):2051-7. doi: 10.1001/jama.2010.675. PubMed PMID: 20501927; PubMed Central PMCID: PMC2968755.

28. Farida H, Severin JA, Gasem MH, Keuter M, Wahyono H, van den Broek P, et al. Nasopharyngeal carriage of Streptococcus pneumonia in pneumonia-prone age groups in Semarang, Java Island, Indonesia. PloS one. 2014;9(1):e87431. doi: 10.1371/journal.pone.0087431. PubMed PMID: 24498104; PubMed Central PMCID: PMC3909120.

29. Melegaro A, Gay NJ, Medley GF. Estimating the transmission parameters of pneumococcal carriage in households. Epidemiology and infection. 2004;132(3):433-41. PubMed PMID: 15188713; PubMed Central PMCID: PMC2870123.

30. Brankston G, Gitterman L, Hirji Z, Lemieux C, Gardam M. Transmission of influenza A in human beings. The Lancet Infectious diseases. 2007;7(4):257-65. doi: 10.1016/S1473-3099(07)70029-4. PubMed PMID: 17376383.

31. In: Larson EL, Liverman CT, editors. Preventing Transmission of Pandemic Influenza and Other Viral Respiratory Diseases: Personal Protective Equipment for Healthcare Personnel: Update 2010. Washington (DC)2011.

32. Ferreira DM, Neill DR, Bangert M, Gritzfeld JF, Green N, Wright AK, et al. Controlled human infection and rechallenge with Streptococcus pneumoniae reveals the protective efficacy of carriage in healthy adults. American journal of respiratory and critical care medicine. 2013;187(8):855-64. doi: 10.1164/rccm.201212-2277OC. PubMed PMID: 23370916; PubMed Central PMCID: PMC3707375.

33. Levine H, Balicer RD, Zarka S, Sela T, Rozhavski V, Cohen D, et al. Dynamics of pneumococcal acquisition and carriage in young adults during training in confined settings in Israel. PloS one. 2012;7(10):e46491. doi: 10.1371/journal.pone.0046491. PubMed PMID: 23056322; PubMed Central PMCID: PMC3466294.

34. MacLennan J, Kafatos G, Neal K, Andrews N, Cameron JC, Roberts R, et al. Social behavior and meningococcal carriage in British teenagers. Emerging infectious diseases. 2006;12(6):950-7. PubMed PMID: 16707051; PubMed Central PMCID: PMC3373034.

607 35. Tumwesigire SG, Barton T. Environmental risk factors for acute respiratory infections among
608 children of military personnel in Uganda. East African medical journal. 1995;72(5):290-4. PubMed
609 PMID: 7555884.
610 36. Rogers WH. Regression standard errors in clustered samples. Stata Technical Bulletin
611 1993;13:19–23.

612 .

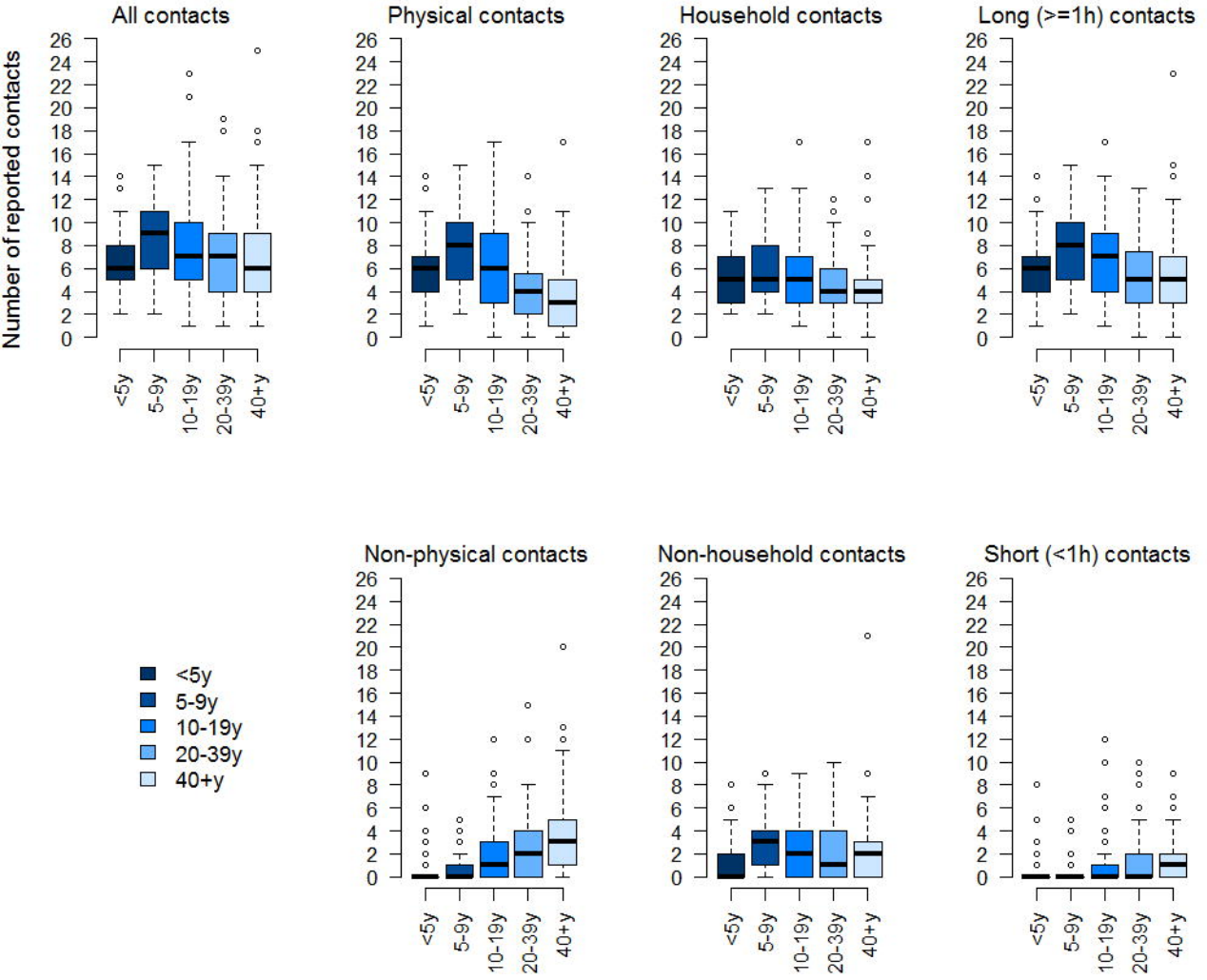
613

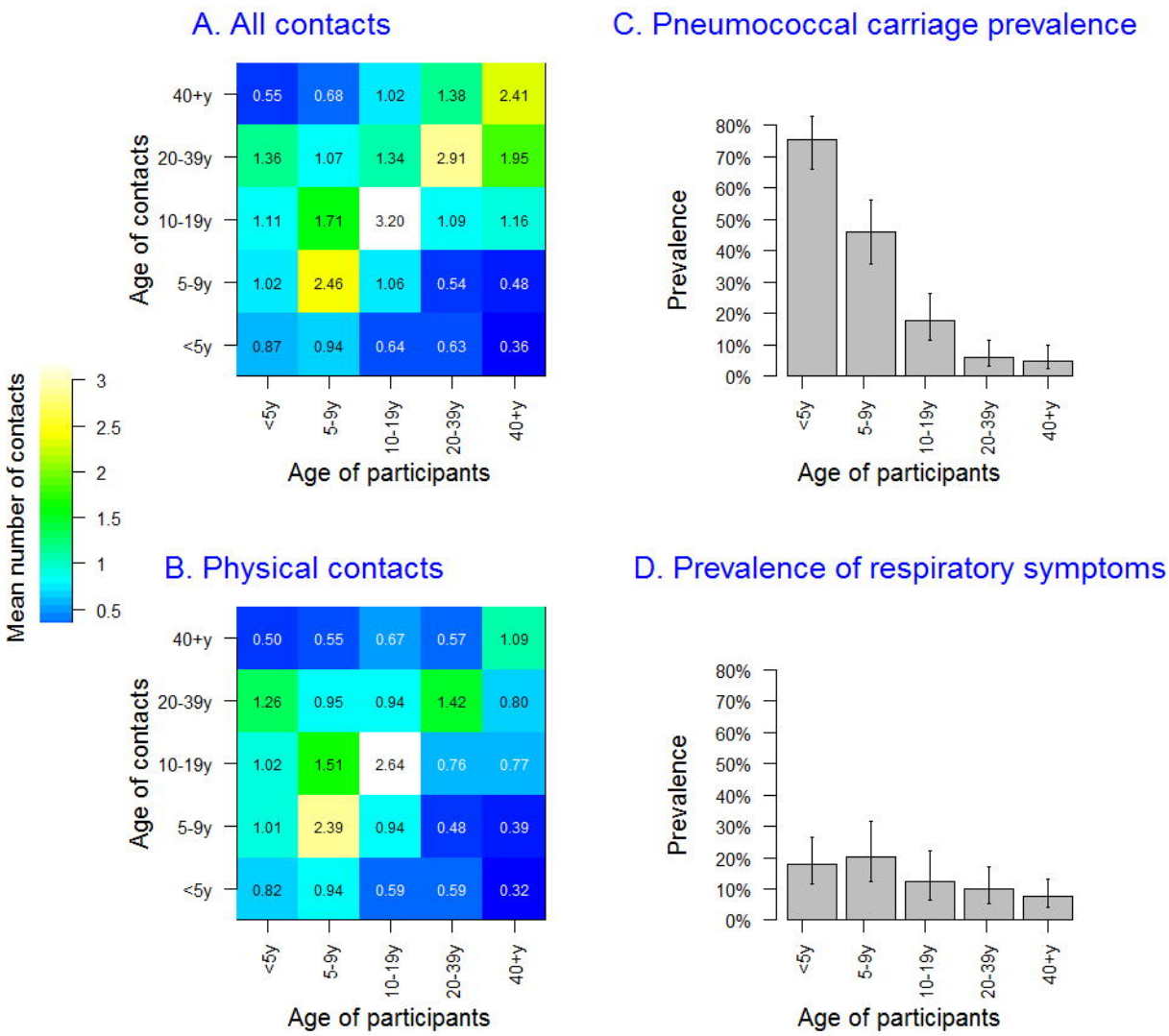
614 **Supporting Information Captions**

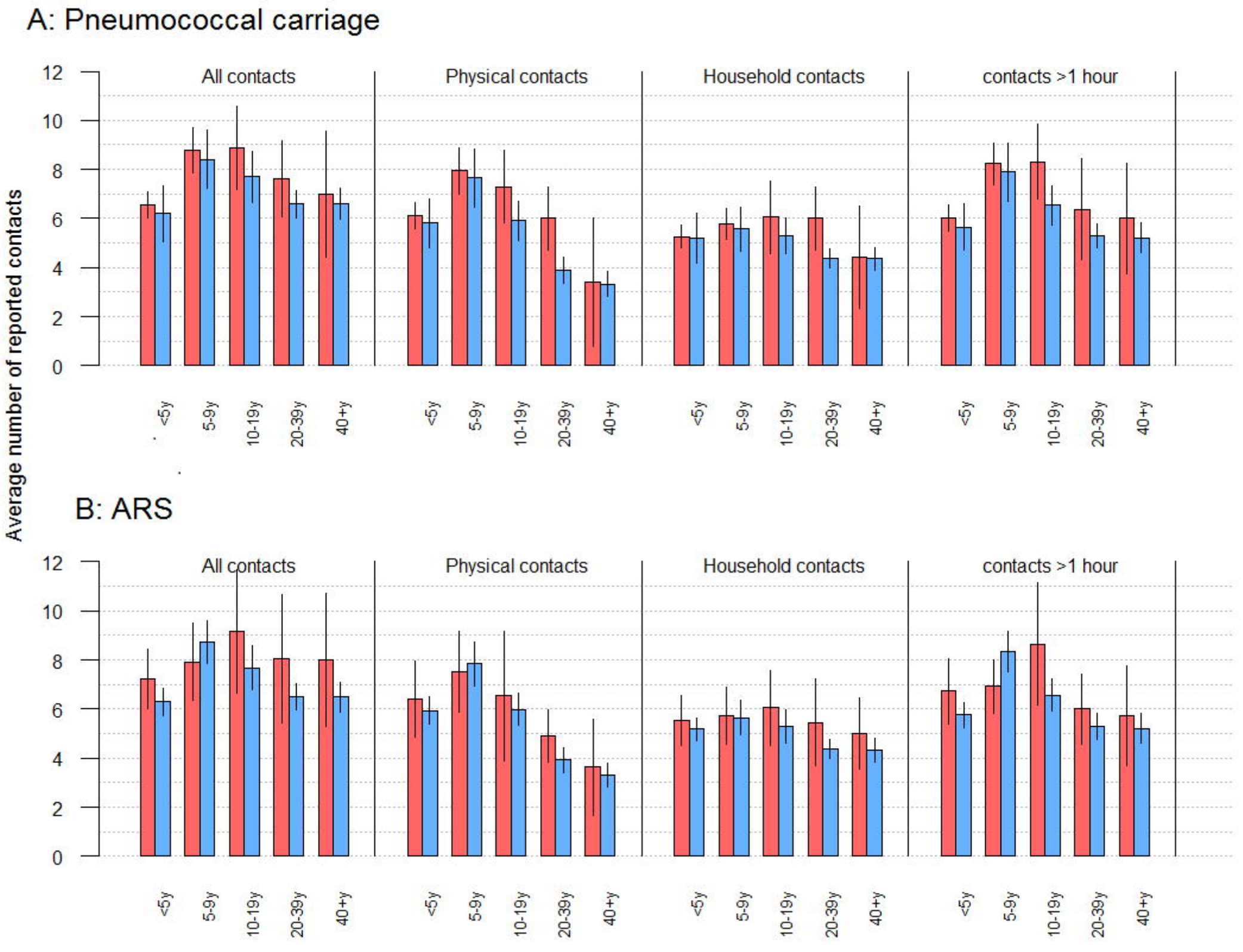
615 **S1 Text.** Analytical approach and model comparison

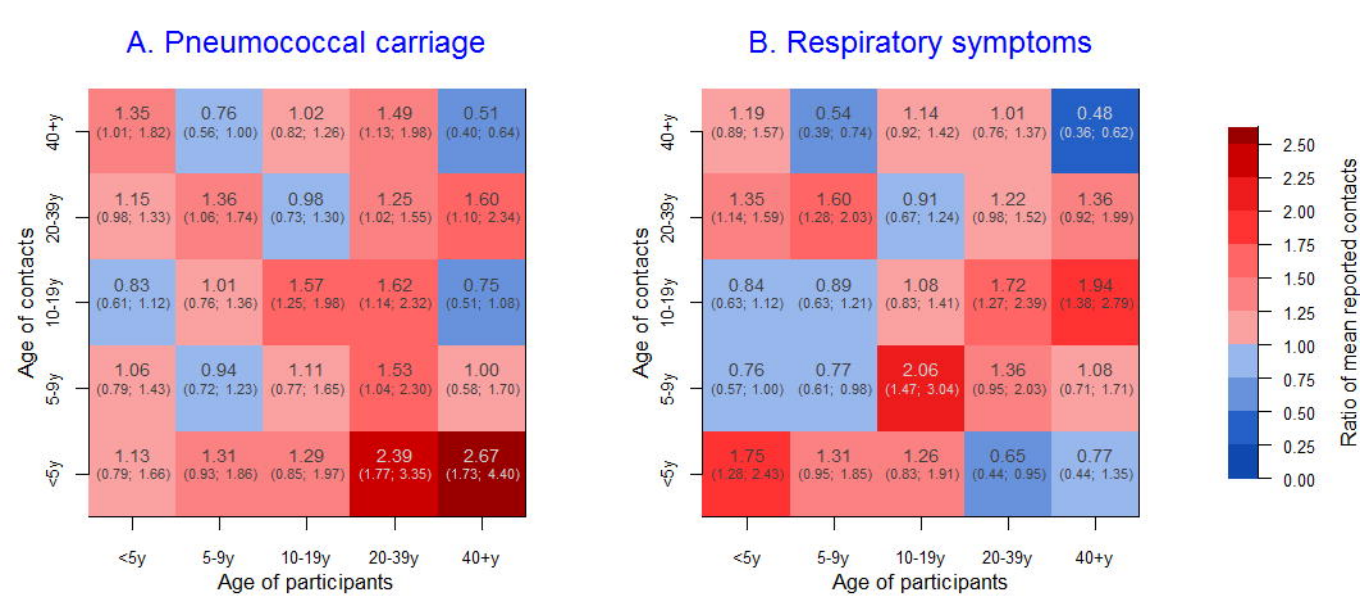
616

617 **S1 Figure**: Difference in the mean number of *physical* contact between pneumococcal
618 carriers and non-carriers (Panel A) and between individuals suffering and not suffering from
619 respiratory symptoms (Panel B).Legend: Matrices of the mean difference between the mean
620 number of contacts among pneumococcal carriers and non-carriers (panel A), and
621 individuals with ARS compared to non-ARS (panel B). The numbers represent the point
622 estimate and the lower and upper bound of the 95% credible interval are shown inside the
623 brackets.
624
625
626 **S2 Text**: Data dictionary
627
628 **S3 Data**: Dataset


All contacts
Physical contacts
Household contacts
Long (>=1h) contacts
Non-physical contacts
Non-household contacts
Short (<1h) contacts
Number of reported contacts
0
2
4
6
8
10
12
14
16
18
20
22
24
26
<5y
5-9y
10-19y
20-39y
40+y


A. All contacts
C. Pneumococcal carriage prevalence
B. Physical contacts
D. Prevalence of respiratory symptoms
Mean number of contacts
3
2.5
2
1.5
1
0.5
Age of contacts
40+y
20-39y
10-19y
5-9y
<5y
0.55 0.68 1.02 1.38 2.41
1.36 1.07 1.34 2.91 1.95
1.11 1.71 3.20 1.09 1.16
1.02 2.46 1.06 0.54 0.48
0.87 0.94 0.64 0.63 0.36
0.50 0.55 0.67 0.57 1.09
1.26 0.95 0.94 1.42 0.80
1.02 1.51 2.64 0.76 0.77
1.01 2.39 0.94 0.48 0.39
0.82 0.94 0.59 0.59 0.32
Age of participants
Prevalence
80%
70%
60%
50%
40%
30%
20%
10%
0%


A: Pneumococcal carriage
All contacts
Physical contacts
Household contacts
contacts >1 hour
Average number of reported contacts
0
2
4
6
8
10
12
<5y
5-9y
10-19y
20-39y
40+y
B: ARS
All contacts
Physical contacts
Household contacts
contacts >1 hour
0
2
4
6
8
10
12
<5y
5-9y
10-19y
20-39y
40+y


A. Pneumococcal carriage
B. Respiratory symptoms
Age of contacts
Age of participants
<5y
5-9y
10-19y
20-39y
40+y
1.35 (1.01; 1.82)
0.76 (0.56; 1.00)
1.02 (0.82; 1.26)
1.49 (1.13; 1.98)
0.51 (0.40; 0.64)
1.15 (0.98; 1.33)
1.36 (1.06; 1.74)
0.98 (0.73; 1.30)
1.25 (1.02; 1.55)
1.60 (1.10; 2.34)
0.83 (0.61; 1.12)
1.01 (0.76; 1.36)
1.57 (1.25; 1.98)
1.62 (1.14; 2.32)
0.75 (0.51; 1.08)
1.06 (0.79; 1.43)
0.94 (0.72; 1.23)
1.11 (0.77; 1.65)
1.53 (1.04; 2.30)
1.00 (0.58; 1.70)
1.13 (0.79; 1.66)
1.31 (0.93; 1.86)
1.29 (0.85; 1.97)
2.39 (1.77; 3.35)
2.67 (1.73; 4.40)
1.19 (0.89; 1.57)
0.54 (0.39; 0.74)
1.14 (0.92; 1.42)
1.01 (0.76; 1.37)
0.48 (0.36; 0.62)
1.35 (1.14; 1.59)
1.60 (1.28; 2.03)
0.91 (0.67; 1.24)
1.22 (0.98; 1.52)
1.36 (0.92; 1.99)
0.84 (0.63; 1.12)
0.89 (0.63; 1.21)
1.08 (0.83; 1.41)
1.72 (1.27; 2.39)
1.94 (1.38; 2.79)
0.76 (0.57; 1.00)
0.77 (0.61; 0.98)
2.06 (1.47; 3.04)
1.36 (0.95; 2.03)
1.08 (0.71; 1.71)
1.75 (1.28; 2.43)
1.31 (0.95; 1.85)
1.26 (0.83; 1.91)
0.65 (0.44; 0.95)
0.77 (0.44; 1.35)
Ratio of mean reported contacts
2.50
2.25
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00