

# Improving Estimates of Social Contact Patterns for Airborne Transmission of Respiratory Pathogens

Nicky McCreesh, Mbali Mohlamonyane, Anita Edwards, Stephen Olivier, Keabetswe Dikgale, Njabulo Dayi, Dickman Gareta, Robin Wood, Alison D. Grant, Richard G. White, Keren Middelkoop

Data on social contact patterns are widely used to parameterize age-mixing matrices in mathematical models of infectious diseases. Most studies focus on close contacts only (i.e., persons spoken with face-to-face). This focus may be appropriate for studies of droplet and short-range aerosol transmission but neglects casual or shared air contacts, who may be at risk from airborne transmission. Using data from 2 provinces in South Africa, we estimated age mixing patterns relevant for droplet transmission, nonsaturating airborne transmission, and *Mycobacterium tuberculosis* transmission, an airborne infection where saturation of household contacts occurs. Estimated contact patterns by age did not vary greatly between the infection types, indicating that widespread use of close contact data may not be resulting in major inaccuracies. However, contact in persons  $\geq 50$  years of age was lower when we considered casual contacts, and therefore the contribution of older age groups to airborne transmission may be overestimated.

Mathematical models of infectious disease transmission are widely used to help develop infectious disease policy, estimate the potential effect of interventions, and provide insight into disease dynamics and natural history. Many models incorporate patterns of mixing between different sections of the population, most commonly between different age groups.

Author affiliations: London School of Hygiene and Tropical Medicine TB Centre, London, UK (N. McCreesh, A.D. Grant, R.G. White); The Desmond Tutu HIV Centre, University of Cape Town, Cape Town, South Africa (M. Mohlamonyane, R. Wood, K. Middelkoop); University of KwaZulu-Natal Africa Health Research Institute, Durban, South Africa (A. Edwards, A.D. Grant); Africa Health Research Institute, Durban (S. Olivier, K. Dikgale, N. Dayi, D. Gareta); University of Cape Town Department of Medicine, Cape Town (R. Wood, K. Middelkoop); University of the Witwatersrand School of Public Health, Johannesburg, South Africa (A.D. Grant)

DOI: <https://doi.org/10.3201/eid2810.212567>

Simulated mixing patterns can have a considerable effect on model dynamics (1), underscoring the importance of simulating realistic mixing patterns. Mixing patterns are frequently shaped by social contact data (i.e., empirical data collected from respondents about the persons with whom they had contact during a set period) (2).

Most social contact data collection has focused on close contacts, using a definition of contacts that required a 2-way face-to-face conversation of  $\geq 3$  words, close proximity (e.g., within 2 meters), physical contact, or some combination of those criteria (2). Those types of contact may approximate reasonably well the types of contact that are relevant for infections that are transmitted primarily through direct contact, short range aerosols, droplets, or some combination of these modes. For obligate, preferential, or opportunistic airborne infections such as measles, *Mycobacterium tuberculosis*, and SARS-CoV-2, however, this definition probably excludes many potentially effective contacts because transmission of airborne infections can occur between anybody sharing air in inadequately ventilated indoor spaces, regardless of whether conversation occurs, and over distances  $> 2$  meters (3). For airborne infections, estimates of casual contact time may therefore be more appropriate, calculated as the time spent in indoor locations multiplied by the number of other persons present.

Tuberculosis also differs from most respiratory infections in terms of the long periods during which persons are potentially infectious; an estimated 9–36 months elapses between disease development and diagnosis (or notification) in 11 countries with high tuberculosis incidences (4). Therefore, transmission to repeated contacts can partially saturate (even allowing for reinfection), making the relationship between contact time and infection risk nonlinear (5). This effect is most pronounced for contact between

household members (5). Household membership and repeated contacts are rarely explicitly simulated in mathematical models, and therefore the effects of contact saturation need to be incorporated into the mixing matrices used to parameterize the models.

In this article, we describe methods for estimating age-mixing patterns relevant for nonsaturating airborne transmission and *M. tuberculosis* transmission by using a novel weighted approach to incorporate the effects of household contact saturation into our estimates for *M. tuberculosis*. We generate estimates of age mixing using data on close and casual contacts from 2 communities in South Africa and compare the estimated mixing patterns with those typically used in mathematical modeling studies (i.e., generated using close contact numbers, and more suitable for droplet or short range aerosol transmission).

## Methods

We collected social contact data in 2 study communities in South Africa: 1 in KwaZulu-Natal Province and 1 in Western Cape Province. Both communities have high rates of unemployment, high prevalence of HIV, and high incidence of tuberculosis compared with the other provinces as a whole. The study community in KwaZulu-Natal consisted of a population of  $\approx 46,000$ , living in the predominantly rural and peri-urban areas in the catchment areas of 2 primary care clinics and within a demographic surveillance area (DSA). The study community in Western Cape was a peri-urban community of  $\approx 27,000$  and was an established research site with biennial censuses.

## Data Collection

We collected the KwaZulu-Natal data during March–December 2019. We sampled 3,093 adults ( $\geq 18$  years of age) at random from an estimated population of 33,288, stratified by residential area (small-scale divisions with  $\approx 350$  households per area) and with probability proportional to the number of eligible persons in each area, based on the most recent DSA census conducted before area entry. We made up to 3 attempts to contact sampled persons.

We collected the Western Cape data during May–October 2019. In total, we selected 1,530 adults ( $\geq 15$  years of age) from an estimated population of 20,633, by using age- and sex-stratified random sampling, based on a census conducted in the study population in February and March 2019. We made up to 5 attempts to contact selected persons on different days of the week (including weekends).

For both surveys, we conducted interviews face-to-face at the respondents' homes, by using interview

administered questionnaires on tablet computers. We conducted interviews in isiZulu in KwaZulu-Natal and in English or isiXhosa in Western Cape. We asked respondents about their movements on a randomly assigned day during the preceding week in KwaZulu-Natal, and on the day before the interview in Western Cape. To allow casual contact time (defined as time spent "sharing air" indoors or on transport) to be estimated, we asked respondents to list the places they had visited (including their own home) and transport they had used. For each location, questions asked included:

- What type of location was it? (Appendix Figure 5, <https://wwwnc.cdc.gov/EID/article/28/10/21-2567-App1.pdf>)
- How long did you spend there? (recorded in hours and minutes)
- How many persons were there halfway through the time you were there?

We did not ask respondents for the ages of persons present because it was thought that respondents would not be able to accurately remember and estimate the ages of all persons present in all indoor locations visited and transport used. We also asked respondents about their close contacts, defined as persons with whom the respondent had a face-to-face conversation. We first asked respondents to make a numbered list of all their contacts, with help from the interviewer. We then asked respondents questions about 10 contacts (selected at random by number by the tablet computers) or all of their contacts if they reported  $< 10$ . Questions included:

- Is this contact a member of your household?
- How old do you think they are?
- How much time did you spend with them in total?

We also collected respondents' basic demographic information. For the KwaZulu-Natal community, we obtained data on household size and residency (i.e., urban, peri-urban, or rural) from the most recent DSA census. We collected all other data directly from the respondents.

## Data Analysis

We estimated close contact numbers and times by using data on persons with whom the respondents reported having a face-to-face conversation. We generated 95% plausible intervals for the age-mixing matrices by using bootstrapping.

We estimated casual contact time in a location as the duration of time the respondent reported

spending there multiplied by the reported number of persons present. We generated central estimates for casual contact time age-mixing matrices by using the method outlined in McCreesh et al (6). In brief, because data were collected on numbers of total persons and children present in indoor locations only, and not the ages of adults, we need to estimate the age distribution of adult casual contacts. We therefore assumed that the age distribution of adult contacts in each location type matched the weighted age distribution of respondents who reported visiting locations of that type. Again, we generated 95% plausible ranges by using bootstrapping.

We adjusted the age-mixing matrices to be symmetric by using the study community age structures. We used data on adult contact numbers and time with children to estimate child contact numbers and time with adults, assuming that overall contact numbers and time between children and adults in each age group is equal to overall contact numbers and time between adults in each age group and children. To enable comparison between the 2 study communities, the lowest respondent age group was set at 15–19 years for both surveys. Because persons 15–17 years of age were not interviewed in KwaZulu-Natal, we assumed that contact patterns in persons 18–19 years of age were representative of contact patterns in all persons 15–19 years of age (Appendix).

**Generating Age-Mixing Matrices for Droplet and Nonsaturating Airborne Transmission and *Mycobacteria tuberculosis***

We set age-mixing matrices relevant for droplet transmission to be equal to age-mixing matrices

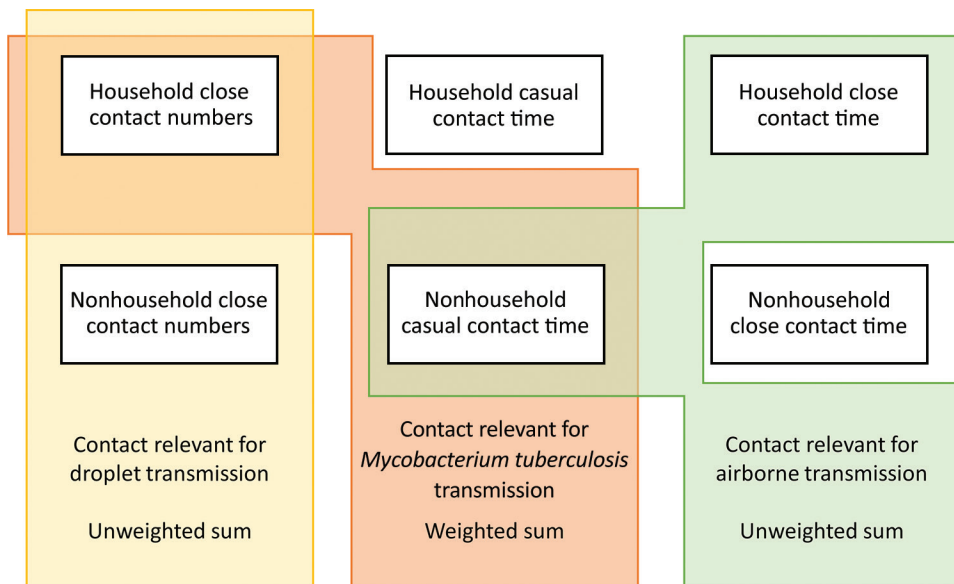
calculated using close contact numbers (Figure 1). We set age-mixing matrices relevant for nonsaturating airborne transmission to be equal to the unweighted sum of the household close contact time matrices and the nonhousehold casual contact time matrices. We used close contact time between household members for household estimates, as opposed to casual contact time occurring in households. We did so because most contact between household members is likely to meet the definition of close contact, and because this approach enabled the age structures of households to be more accurately reflected in the age-mixing matrices. We set age-mixing matrices relevant for *M. tuberculosis* transmission to be equal to the sum of the household close contact number matrices and the nonhousehold casual contact time matrices. We weighted these matrices to reflect empirical estimates of the proportion of tuberculosis that results from household transmission (central estimate 12% [range 8%–16%]) (5).

To enable direct comparisons to be made between the different age-mixing matrices, we adjusted the matrices for nonsaturating airborne transmission and *M. tuberculosis* transmission to give the same mean contact intensity between adults as the matrices for droplet transmission. We used bootstrapping to generate plausible ranges (Appendix).

**Results**

**Recruitment**

Of the 3,093 persons sampled in KwaZulu-Natal, 1,723 (56%) were successfully contacted, 299 (10%) were dead or reported to have out-migrated, and



**Figure 1.** Summary of data used to estimate age-mixing matrices for a study of social contact patterns for airborne transmission of respiratory pathogens, KwaZulu Natal and Western Cape Provinces, South Africa, 2019. Diagram showing how age-mixing matrices relevant for the transmission of droplet infections, airborne infections, and *Mycobacterium tuberculosis* were estimated using empirical data on close contact numbers, close contact time, and casual contact time.

**Table.** Characteristics of respondents and target population for study of social contact patterns for airborne transmission of respiratory pathogens, KwaZulu Natal and Western Cape Provinces, South Africa, 2019

Characteristic	KwaZulu Natal		Western Cape	
	Sample, no. (%)	Target population, %*	Sample, no. (%)	Target population, %*
Sex				
M	751 (44)	41	553 (50)	52
F	953 (56)	59	562 (50)	48
Age group, y				
15–17	0	9.1	56 (5)	4.5
18–19	118 (6.9)	5.6	84 (7.5)	4.5
20–29	495 (29)	26	412 (37)	33
30–39	308 (18)	21	358 (32)	37
40–49	227 (13)	13	142 (13)	15
>50	556 (33)	25	63 (5.7)	6.5
Residence				
Rural	867 (51)	59	0	0
Peri-urban	716 (42)	33	1,115 (100)	100
Urban	121 (7.1)	8	0	0
Monthly household income, South African rands				
<1,000	416 (24)		111 (10)	
1,000–2,500	785 (46)		261 (23)	
2,500–5,000	302 (18)		374 (34)	
5,000–10,000	125 (7.3)		179 (16)	
>10,000	65 (3.8)		61 (5.5)	
Unknown/missing	11 (0.65)		129 (12)	
Employment				
Full-time	329 (19)		403 (36)	
Part-time/casual	68 (4)		213 (19)	
None	1299 (76)		492 (44)	
Missing	8 (0.5)		7 (0.6)	
Household size				
1	115 (6.7)	4.1	203 (18)	19
2–4	287 (17)	26	683 (61)	66
5–7	488 (29)	33	195 (17)	13
8–10	375 (22)	20	26 (2.3)	1.6
>11	439 (26)	17	8 (0.72)	0.4
Day reported				
Monday	239 (14)		203 (18)	
Tuesday	242 (14)		202 (18)	
Wednesday	239 (14)		187 (17)	
Thursday	251 (15)		138 (12)	
Friday	261 (15)		80 (7.2)	
Saturday	245 (14)		98 (8.8)	
Sunday	227 (13)		207 (19)	
<b>Total</b>	<b>1,704</b>	<b>33,288</b>	<b>1,115</b>	<b>20,633</b>

\*Target population refers to persons in the populations  $\geq 15$  years of age.

†In KwaZulu-Natal, urban is defined as KwaMsane Municipality, peri-urban as other areas with a population density  $\geq 400/\text{km}^2$ , and rural as areas with a population density  $< 400/\text{km}^2$ .

1,071 (35%) could not be contacted. Of those successfully contacted, 1,704 (99%) completed an interview.

Of the 1,530 persons sampled in Western Cape, 1,214 (93%) were successfully contacted, 117 (8%) had moved or died, 193 (13%) had had incorrect information listed in the census, and 6 were uncontactable. Of the 1,214 persons contacted, 77 (6%) refused to be interviewed and 14 were ineligible (because of disability or lack of fluency with English and isiXhosa). Of 1,123 persons interviewed, unexplained technical issues meant that data from 8 interviews were lost between collection and transfer to the database, leaving 1,115 (92%) completed interviews.

For both populations, the recruited sample was a reasonable match to the target population in terms

of sex, age, and residence type (urban, peri-urban, or rural) (Table). Respondents in Kwa-Zulu-Natal also were a close match to the target population in terms of employment status (Appendix). No data on employment status for the target population were available for Western Cape.

#### Contact Numbers and Time

We stratified household and nonhousehold close contact numbers and time and casual contact time in KwaZulu-Natal and Western Cape, by sex, age, and household size (Figure 2, 3; Appendix Tables 1–6). Overall, close contact numbers and time, as well as casual contact time, were higher for women than for men in both communities; however, the

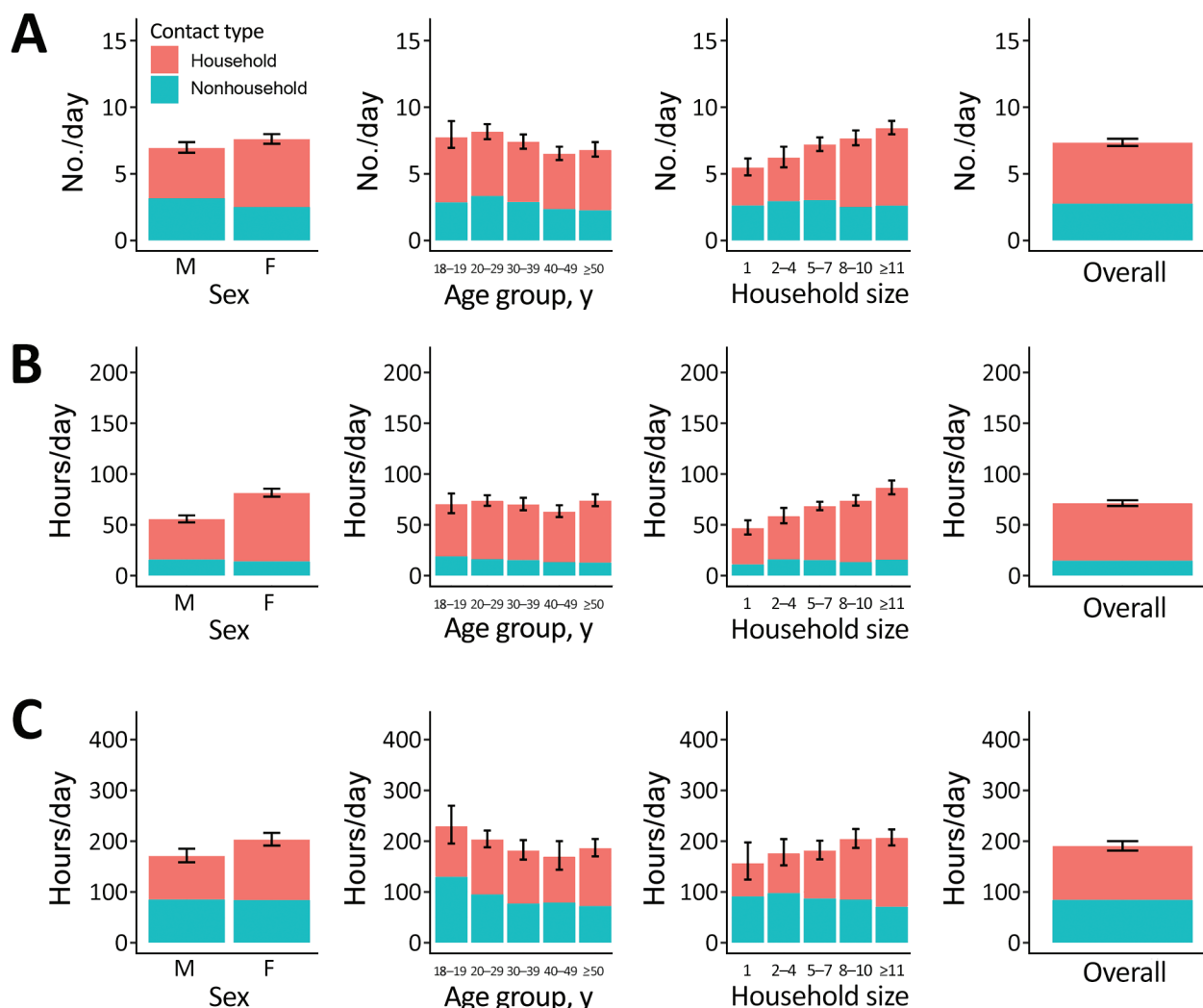


differences were generally not large (close contact time 46% higher for women in KwaZulu-Natal and 8%–22% higher for other contact measures and settings) and not significant for close contact numbers or casual contact time in Western Cape. We observed a tendency for casual contact time to decrease slightly with age in both communities, and close contact numbers and time were substantially higher in persons 15–19 years of age than in older age groups in Western Cape only (Western Cape close contact numbers: 11 in persons 15–19 years of age, 7.7–8.7 in older age groups [ $p < 0.001$ ]; close contact time: 80 hours in persons 15–19 years of age, 49–63 in older age groups [ $p < 0.001$ ]). Close contact numbers and time, as well as casual contact time, increased with increasing

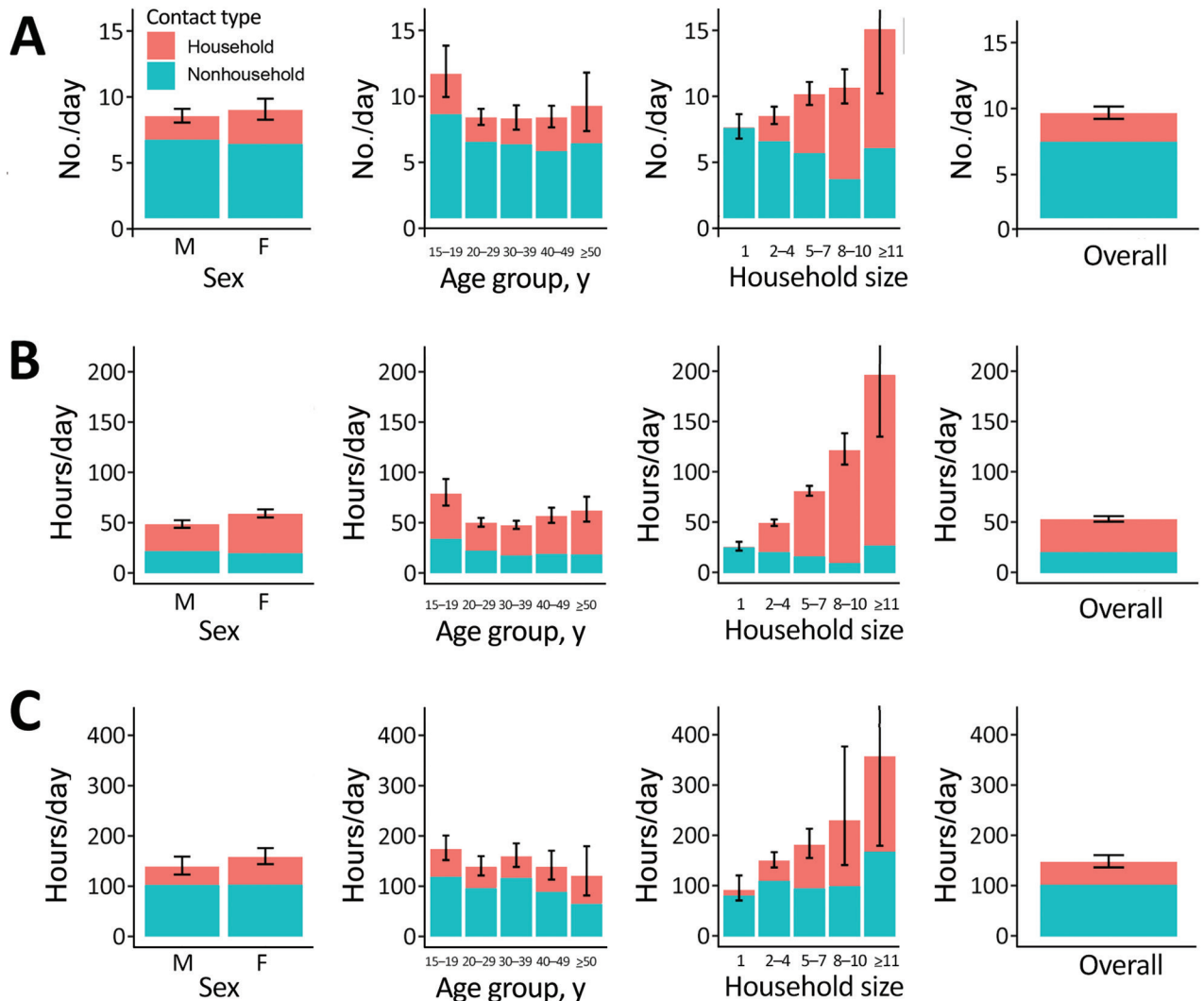
household size in both communities, driven by increases in contact with household members. Contact between household members made up a higher proportion of total contact in KwaZulu-Natal than in Western Cape for all types of contact (close contact numbers: 62% in KwaZulu-Natal, 27% in Western Cape; close contact time: 79% in KwaZulu-Natal, 60% in Western Cape; casual contact time: 55% in KwaZulu-Natal, 31% in Western Cape).

### Age Mixing

We generated estimated age-mixing matrices for droplet transmission non-saturating airborne transmission, and *M. tuberculosis* transmission for KwaZulu-Natal and Western Cape (Figure 4, 5). We also



**Figure 2.** Household and nonhousehold close contact numbers (A), close contact time (B), and casual contact time (C) for study of social contact patterns for airborne transmission of respiratory pathogens, KwaZulu-Natal Province, South Africa, by sex, age group, and household size. Error bars show 95% CIs for total contact numbers or time. For KwaZulu-Natal, household size data were taken from census data and did not always correspond exactly with respondents' views of who they considered to be household members. For this reason, some contact with household members was reported by respondents who we recorded as having a household size of 1.



**Figure 3.** Household and nonhousehold close contact numbers (A), close contact time (B), and casual contact time (C) in Western Cape Province, South Africa, by sex, age, and household size, for study of social contact patterns for airborne transmission of respiratory pathogens. Error bars show 95% CIs for total contact numbers or time. In Western Cape, contact with household members was reported by a small proportion of respondents who had reported having no household members, most likely reflecting errors in the data.

generated 95% plausible ranges for these matrices (Appendix Figure 1, 2).

Estimated contact patterns by age did not vary greatly between the infection types in either community. However, age-mixing patterns were less assortative in the nonsaturating airborne and *M. tuberculosis* matrices compared with the droplet matrices in both communities (Appendix). The exception to this pattern was contact between persons 15–19 years of age in KwaZulu-Natal, which was more intense in the nonsaturating airborne and *M. tuberculosis* matrices than the droplet matrices. In both communities, relative to other adult age groups, overall contact intensities were lower in persons ≥50 years of age when considering contact relevant for

nonsaturating airborne transmission or the transmission of *M. tuberculosis* compared with contact relevant for droplet transmission.

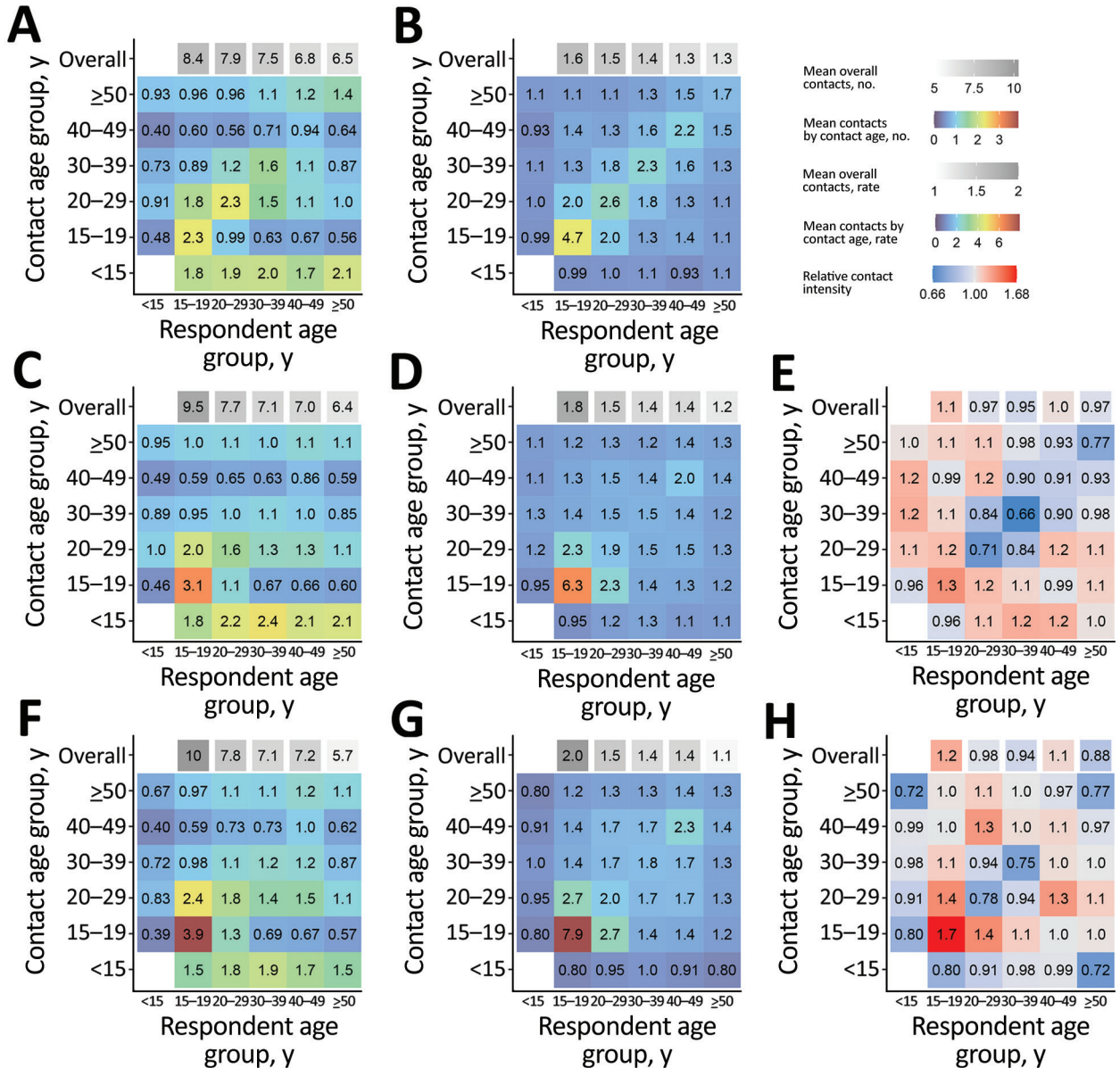
## Discussion

Using data from 2 provinces in South Africa, we estimated contact and age-mixing patterns relevant for the transmission of droplet infections, nonsaturating airborne infections, and *M. tuberculosis*. In our communities, contact patterns did not vary greatly between contacts relevant for droplet infections and those relevant for nonsaturating airborne or *M. tuberculosis* transmission. However, using close contact data in models of the transmission of *M. tuberculosis* or other airborne infections in our study communities

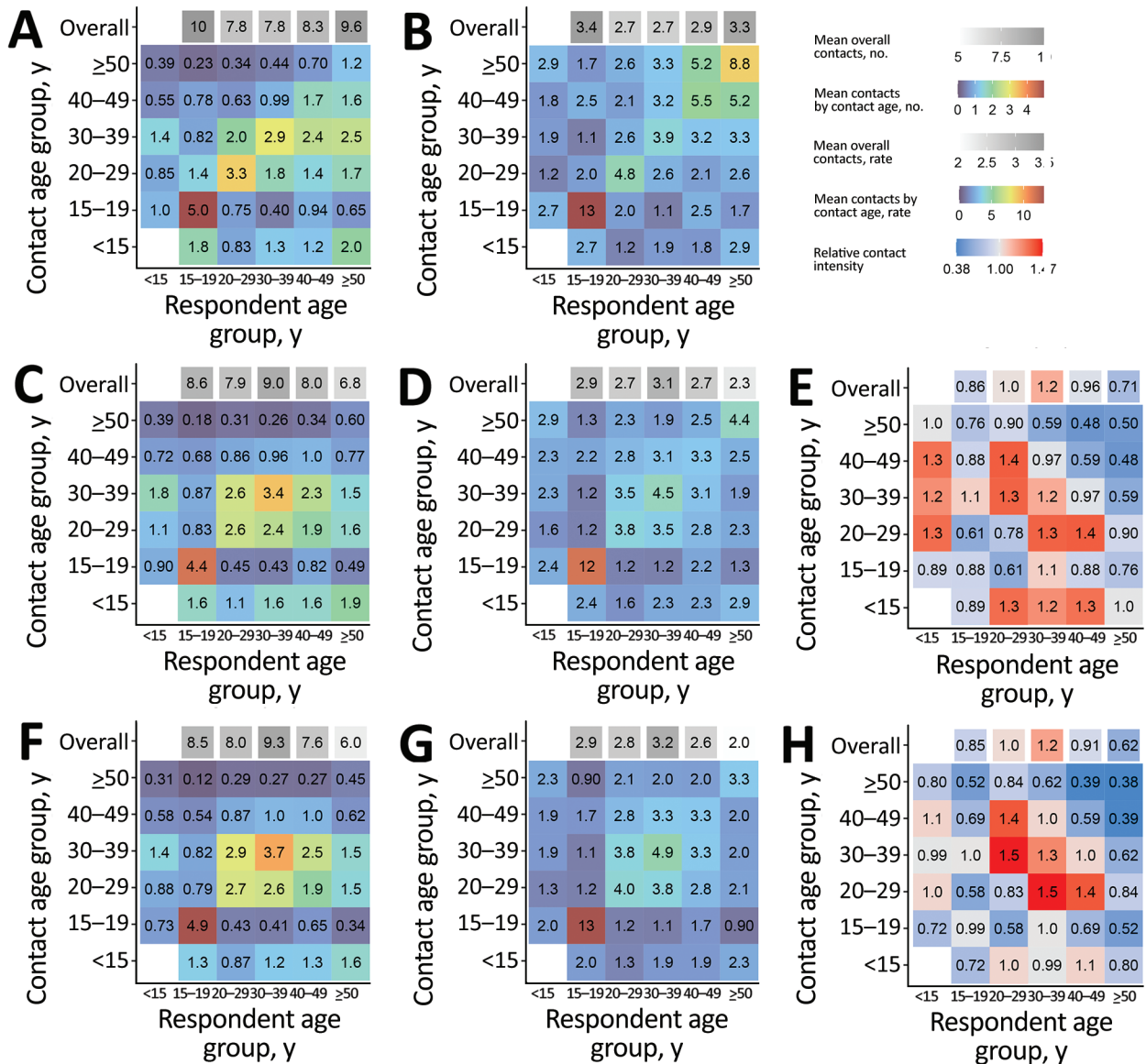
may mean that the importance of adults  $\geq 50$  years of age to transmission is overestimated.

Very few data are available on casual contact patterns from any setting. Previous studies in the same community in Western Cape have found greater drops in casual contact time than in close contact

numbers in older age groups (6) and decreases in indoor casual contact numbers with age (7). Another study in the same community found high levels of age-assortative mixing with respect to casual contact time in schools and workplaces (8). More data are needed on casual contact patterns, and age-mixing



**Figure 4.** Age-mixing matrices relevant for droplet transmission (A, B), nonsaturating airborne transmission (C, D), and *Mycobacterium tuberculosis* transmission (F, G) for study of social contact patterns for airborne transmission of respiratory pathogens, KwaZulu-Natal Province, South Africa. Panels A, C, and F show absolute contact intensities between respondents and contacts in each age group; panels B, D, and G show intensities of contact between each member of each age group; panels E and H show intensities for airborne infections and *M. tuberculosis* compared with intensities for droplet infections, respectively. Numbers shown in panel A are the mean number of contacts respondents in each age group have with contacts in each age group per day. Numbers shown in panel B are the rate of contact between each person in the population per day, expressed as rates  $\times 10^5$ . Numbers and rates in panels C, D, F, and G are standardized so that the mean overall contact intensity by reported by adult respondents is the same as the mean number of overall close contacts reported by adult respondents (panel A). Contact numbers between child respondents and contacts in each age group were estimated from data on contact between adult respondents and child contacts.



**Figure 5.** Age-mixing matrices relevant for droplet transmission (A, B), nonsaturating airborne transmission (C, D), and *Mycobacterium tuberculosis* transmission (F, G) for study of social contact patterns for airborne transmission of respiratory pathogens, Western Cape Province, South Africa. Panels A, C, and F show absolute contact intensities between respondents and contacts in each age group; panels B, D, and G show intensities of contact between each member of each age group; panels E and H show intensities for airborne infections and *Mycobacterium tuberculosis* compared with intensities for droplet infections, respectively. Numbers shown in panel A are the mean number of contacts respondents in each age group have with contacts in each age group per day. Numbers shown in panel B are the rate of contact between each person in the population per day, expressed as rates  $\times 10^5$ . Numbers and rates in panels C, D, F, and G are standardized so that the mean overall contact intensity by reported by adult respondents is the same as the mean number of overall close contacts reported by adult respondents (panel A). Contact numbers between child respondents and contacts in each age group were estimated from data on contact between adult respondents and child contacts.

patterns in particular, to determine whether the findings of this study are generalizable to other settings and to improve the predictions from mathematical models of the transmission of *M. tuberculosis* and other airborne infections.

Our approaches to generating the separate droplet and airborne transmission matrices are necessarily

simplifications, and many infections will not fit neatly into these 2 categories. In addition, considerable uncertainty exists about the role of different transmission routes to the spread of many infections. Droplets have traditionally been considered to be the main transmission route for most respiratory viruses; however, there is evidence that airborne transmission



can occur for a wide range of pathogens, including influenza, respiratory syncytial virus, Middle East respiratory syndrome coronavirus, and SARS-CoV-2 (9). One model using data on household transmission of influenza A suggested that airborne transmission was responsible for about half of infections (10). For infections where both airborne and droplet or short range aerosol transmission are thought to play an important role in transmission, an intermediate matrix may be preferable.

There are 2 main differences between our droplet and airborne or *M. tuberculosis* age-mixing matrices. The first is the type of nonhousehold contacts considered: close (face-to-face conversation) or casual (sharing space indoors). The second is that the airborne and (nonhousehold component of the) *M. tuberculosis* matrices are based on contact time, rather than unique contact numbers. The primary reason for using contact time for casual contacts is that respondents are unlikely to be able to estimate unique casual contact numbers for many locations they visit, necessitating the use of contact time or assumptions about the rate of turnover of unique persons in a location. For our droplet transmission matrices, we chose to use unique contact numbers in a 24-hour period because that is the most commonly used method (2) and therefore enables comparisons to be made with what is typically done. However, we should note that both the choice of a 24-hour time period and the lack of any weighting or restrictions by contact duration or other measures of closeness are relatively arbitrary choices.

Robust evidence as to the types of contact most relevant to transmission are limited for respiratory infections. Several studies have compared the fit to data on varicella, parvovirus B19, or influenza A seroprevalence by age of models parameterized by using contact patterns generated from close contact data in a range of different ways (11–13). Overall, those studies suggest that analysis methods that give greater weight to more intimate contacts may be preferable in some circumstances; for instance, restricting what counts as a contact to those involving physical touch or a minimum contact duration or using contact time rather than contact numbers. Approaches based on contact numbers may be more suitable for more highly transmissible infections such as measles, where only a short duration of contact is needed for transmission, whereas approaches based on contact time may be more suitable for less transmissible infections, where repeated or longer contacts are needed (14).

Fewer studies have considered expanding the pool of contacts beyond close contacts only, to also include casual contacts. However, a study that had

paired individual-level contact data and pandemic influenza A serologic data found that models that included a variable for number of locations visited were strongly supported over those that only included variables for age and close contact numbers (15). This finding suggests that airborne transmission may play a role in the spread of influenza A, or that the standard close contact definition misses a substantial proportion of contacts at risk for droplet transmission.

Other factors may also influence airborne and *M. tuberculosis* transmission risk, which are not accounted for in the analyses. Ventilation rates play a large role in determining airborne infection risk (16), and giving less weight to contact occurring in better ventilated settings would improve our airborne and *M. tuberculosis* matrices. Unfortunately, few data on ventilation rates by location type are available, and they show large amounts of variation between locations and between the same location on different days (17). Saturation of contacts may occur for infections other than *M. tuberculosis*, particularly highly transmissible pathogens such as measles virus. An approach based on casual contact numbers may be preferable for these infections but would be highly dependent on assumptions made about how unique contact numbers are related to estimates of cross-sectional numbers of persons present.

There are several limitations when using casual contact data to estimate mixing patterns. First, estimates of contact time in places where large numbers of persons are present are likely to be less reliable because a person's estimates of the number of persons present are likely to be poor and because the assumption that a risk for transmission exists between all persons present in the space may not be true in larger spaces. Estimates may be poorer when asking about a random day in the past week (as we did in KwaZulu-Natal) than when asking about the day before the interview (as we did in Western Cape). In our main analysis, when estimating contact time, we cap the number of persons at risk for transmission at 100. In our sensitivity analyses, we show that using a cap of 20 persons or not capping the numbers of persons has a moderate effect on casual contact time age-mixing matrices (Appendix). Conducting similar sensitivity analyses may be necessary when using age-mixing matrices calculated using casual contact time in mathematical models.

A second limitation is that the approach we use to determining the ages of adults present in locations other than respondents' own homes is indirect and relies on the assumption that the age distribution of adults present in a location type reflects the duration

of time respondents of different ages reported spending in that location type. This assumption may not always be reasonable if different age groups tend to visit different locations of the same type (or at different times) or substantial mixing occurs with persons from outside the study community. These issues are discussed further in McCreesh et al. (6).

An additional limitation of our estimates for KwaZulu-Natal only is that we did not recruit persons 15–17 years of age and instead assumed in the analysis that contact by persons 18–19 years of age was representative of contact by all persons 15–19 years of age. This assumption is unlikely to be true given that contacts by persons 15–17 and 18–19 years of age differ greatly in Western Cape (Appendix Figure 9). For this reason, our estimates for persons 15–19 years of age in KwaZulu-Natal should be treated with caution.

To conclude, our estimated age-mixing matrices for droplet transmission, nonsaturating airborne transmission, and *M. tuberculosis* transmission were not substantially different from each other for either community. This finding provides some reassurance that the widespread use of close contact data to parameterize age-mixing matrices for transmission models of airborne infections may not be resulting in major inaccuracies. Some differences were observed, however, particularly in the oldest age group, and our data were from 2 communities in South Africa only. We recommend that future social contact surveys collect data on casual contacts as well as close contacts to determine whether the similarity between different types of contact pattern is true across other settings. We would also urge mathematical modelers to consider whether unique close contact numbers in a 24-hour period are the most appropriate contacts for the infection and scenario they are simulating and to consider performing sensitivity analyses when uncertainty exists as to the most appropriate contact definition.

### Acknowledgments

We thank the social contact survey respondents and all of the fieldworkers involved in the data collection (in KwaZulu-Natal: Nkosingiphile Buthelezi, Zilethile Khumalo, Sifundesihle Malembe, Zodwa Mkwanazi, Sanele Mthiyane; in Western Cape: Sinayo Dyubhele, Mzukisa Diniso, Olwethu Kemele, Zilungile Majola Nyembezi, Nomonde Vungama & Nontyatyambo Bangani).

This research is jointly funded by the UK Medical Research Council (MRC) and the UK Department for International Development under an MRC/Department

for International Development concordat (agreement no MR/P002404/1). The support of the Economic and Social Research Council is gratefully acknowledged. The project is partly funded by the Antimicrobial Resistance Cross Council Initiative supported by the 7 research councils in partnership with other funders, including support from the Global Challenges Research Fund (grant no. ES/P008011/1). N.M. is additionally funded by the Wellcome Trust (218261/Z/19/Z). R.G.W. is funded by the Wellcome Trust (grant no. 218261/Z/19/Z), the US National Institutes of Health (grant no. 1R01AI147321-01), the European and Developing Countries Clinical Trials Partnership (grant no. RIA208D-2505B), UK MRC (grant no. CCF17-7779 through Bloomsbury SET), the Economic and Social Research Council (grant no. ES/P008011/1), the Bill and Melinda Gates Foundation (grant nos. OPP1084276, OPP1135288, and INV-001754), and the World Health Organization (grant no. 2020/985800-0).

Analysis code and data are available at <https://github.com/NickyMcC/CasualAgeMixing>.

### About the Author

Dr. McCreesh is an assistant professor in Infectious Disease Modelling at the London School of Hygiene and Tropical Medicine. Her research interests include using social contact data and mathematical modelling to understand the transmission patterns of, and inform the control of, tuberculosis and other infectious diseases.

### References

1. Prem K, Cook AR, Jit M. Projecting social contact matrices in 152 countries using contact surveys and demographic data. *PLOS Comput Biol*. 2017;13:e1005697. <https://doi.org/10.1371/journal.pcbi.1005697>
2. Van Hoang T, Coletti P, Melegaro A, Wallinga J, Grijalva C, Edmunds J, et al. A systematic review of social contact surveys to inform transmission models of close contact infections. *Epidemiology*. 2019;30:723–36.
3. Raffalli J, Sepkowitz KA, Armstrong D. Community-based outbreaks of tuberculosis. *Arch Intern Med*. 1996;156:1053–60. <https://doi.org/10.1001/archinte.1996.0040041053002>
4. Ku C-C, MacPherson P, Khundi M, Nzawa R, Feasey HR, Nliwasa M, et al. Estimated durations of asymptomatic, symptomatic, and care-seeking phases of tuberculosis disease. *BMC Med*. 2021;19:298. <https://doi.org/10.1186/s12916-021-02128-9>
5. McCreesh N, White RG. An explanation for the low proportion of tuberculosis that results from transmission between household and known social contacts. *Sci Rep*. 2018;8:5382. <https://doi.org/10.1038/s41598-018-23797-2>
6. McCreesh N, Morrow C, Middelkoop K, Wood R, White RG. Estimating age-mixing patterns relevant for the transmission of airborne infections. *Epidemics*. 2019;28:100339. <https://doi.org/10.1016/j.epidem.2019.03.005>
7. Wood R, Racow K, Bekker L-G, Morrow C, Middelkoop K, Mark D, et al. Indoor social networks in a South African township: potential contribution of location to tuberculosis

- transmission. *PLoS One*. 2012;7:e39246. <https://doi.org/10.1371/journal.pone.0039246>
8. Andrews JR, Morrow C, Walensky RP, Wood R. Integrating social contact and environmental data in evaluating tuberculosis transmission in a South African township. *J Infect Dis*. 2014;210:597–603. <https://doi.org/10.1093/infdis/jiu138>
  9. Wang CC, Prather KA, Sznitman J, Jimenez JL, Lakdawala SS, Tufekci Z, et al. Airborne transmission of respiratory viruses. *Science*. 2021;373:eabd9149. <https://doi.org/10.1126/science.abd9149>
  10. Cowling BJ, Ip DKM, Fang VJ, Suntarattiwong P, Olsen SJ, Levy J, et al. Aerosol transmission is an important mode of influenza A virus spread. *Nat Commun*. 2013;4:1935. <https://doi.org/10.1038/ncomms2922>
  11. Ogunjimi B, Hens N, Goeyvaerts N, Aerts M, Van Damme P, Beutels P. Using empirical social contact data to model person to person infectious disease transmission: an illustration for varicella. *Math Biosci*. 2009;218:80–7. <https://doi.org/10.1016/j.mbs.2008.12.009>
  12. 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;3:143–51. <https://doi.org/10.1016/j.epidem.2011.04.001>
  13. Kucharski AJ, Kwok KO, Wei VWI, Cowling BJ, Read JM, Lessler J, et al. The contribution of social behaviour to the transmission of influenza A in a human population. *PLoS Pathog*. 2014;10:e1004206. <https://doi.org/10.1371/journal.ppat.1004206>
  14. Iozzi F, Trusiano F, Chinazzi M, Billari FC, Zagheni E, Merler S, et al. Little Italy: an agent-based approach to the estimation of contact patterns – fitting predicted matrices to serological data. *PLOS Comput Biol*. 2010;6:e1001021. <https://doi.org/10.1371/journal.pcbi.1001021>
  15. 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. *Proc Biol Sci*. 2014;281:20140709.
  16. Li Y, Leung GM, Tang JW, Yang X, Chao CY, Lin JZ, et al. Role of ventilation in airborne transmission of infectious agents in the built environment – a multidisciplinary systematic review. *Indoor Air*. 2007;17:2–18. <https://doi.org/10.1111/j.1600-0668.2006.00445.x>
  17. Taylor JG, Yates TA, Mthethwa M, Tanser F, Abubakar I, Altamirano H. Measuring ventilation and modelling *M. tuberculosis* transmission in indoor congregate settings, rural KwaZulu-Natal. *Int J Tuberc Lung Dis*. 2016;20:1155–61. <https://doi.org/10.5588/ijtld.16.0085>

Address for correspondence: Nicky McCreesh, London School of Hygiene & Tropical Medicine, Keppel St, London WC1E 7HT, UK; email: [nicky.mccreesh@lshtm.ac.uk](mailto:nicky.mccreesh@lshtm.ac.uk)

## EID Podcast

### Laboratory-Associated Zika Virus, United States

Since the 2015 Zika virus outbreak in the Americas, transmission of this vectorborne disease has substantially decreased. But Zika virus doesn't spread only through mosquito bites...it also spreads through sexual transmission, blood transfusions, breastfeeding, and even needlestick injuries in laboratories.

Stringent safety protocols minimize the risk of laboratory-associated exposures. But on rare occasions, researchers are accidentally exposed to the disease they are trying to solve.

In this EID podcast, Dr. Susan Hills, a medical epidemiologist at CDC in Fort Collins, Colorado, describes the biosafety lessons exemplified by four cases of laboratory-associated Zika infection.

Visit our website to listen:  
<https://go.usa.gov/xFZU2>

**EMERGING  
 INFECTIOUS DISEASES®**



# Improving Estimates of Social Contact Patterns for Airborne Transmission of Respiratory Pathogens

## Appendix

### Methods

#### Ethics

Ethical approval for the data collection in KwaZulu-Natal was granted by the Biomedical Research Ethics Committee (BREC) of the University of KwaZulu-Natal (BE662/17) and the London School of Hygiene & Tropical Medicine Observational / Interventions Research Ethics Committee (14640). Ethical approval for the data collection in Western Cape was granted by the Human Research Ethics Committee at the University of Cape Town (HREC/REF: 008/2018) and the London School of Hygiene & Tropical Medicine Observational / Interventions Research Ethics Committee (14520). Informed consent was obtained from all participants.

#### Comparison of employment status with target population, KwaZulu-Natal

To investigate the whether we may have under-recruited people who were employed in KwaZulu-Natal, we compared data on employment from the most recent DSA survey between respondents in the social contact survey and the target population aged 18+ years as a whole. No data on employment status for the target population were available for Western Cape.

#### Weighting

All analyses of contact numbers and contact time were weighted. For KwaZulu-Natal, they were weighted to the study population composition by age group (18–19, 20–29, 30–39, 40–49, 50+) and sex. For Western Cape, they were weighted to the study population composition by age group (15–17, 18–19, 20–29, 30–39, 40–49, 50+) and sex. As fewer respondents were asked about Fridays and Saturdays, the Western Cape data were also weighted by the day of the week.



### **Estimating close contact numbers**

Respondents were first asked to make a numbered list of all their contacts, with help from the interviewer. The total number of contacts was recorded on the tablet computers, along with the number of those contacts who were members of the respondent's household. Close contact numbers by respondent characteristic were estimated using the total number of close contacts that the respondent reported, and the number of those contacts who were household members. Respondents who reported more household contacts than total contacts were excluded from the analysis.

Close contact age mixing matrices were generated using data on close contacts on which more detailed information were available (all contacts if respondents reported  $\leq 10$ , a random 10 if they reported more). When generating the central estimate, contact numbers by age group for respondents were multiplied by  $T/D_a$ , where  $T$  was the total number of contacts that they reported, and  $D_a$  was the total number of contacts whose age was provided. Respondents who reported a nonzero number of close contacts, but who did not give the age of any of their contacts were excluded. 95% plausible intervals were generated using 10,000 bootstrapped samples, re-sampling respondents with replacement within age categories, and re-sampling  $T$  contacts with replacement from the set of all contacts of the respondent on which detailed information were collected ( $I$ ). Only sampled contacts who were or weren't reported to be a member of the respondent's household were included when estimating household and 'other' age-mixing patterns respectively.

The age mixing matrices were adjusted to be symmetric, using the study community age structures. Data on adult contacts with children were used to estimate child contact with adults. To allow comparison between the two study communities, and between close and casual age mixing patterns, the lowest respondent age group was set at 15–19 years for both surveys. As 15–17 year olds were not interviewed in KwaZulu-Natal, we assumed that contact patterns in 18–19 years olds were representative of contact patterns in all 15–19 year olds, and adjusted the weights accordingly.

### **Estimating close contact time**

The approach used for estimating close contact time was the same as that used for estimating close contact numbers, except that contact numbers were multiplied by the amount of time that respondents reported spending with each contact that day.

Contacts with the contact duration missing were excluded when generating bootstrap samples for estimating close contact time age mixing patterns. When generating the central estimate, contact numbers by age group for respondents were multiplied by  $T / D_{ad}$ , where T was the total number of contacts that they reported, and  $D_{ad}$  was the total number of contacts whose age and duration were both provided.

### **Estimating casual contact time**

For each location visited, respondents were asked to identify the location type, from a list of frequently visited location types identified by local researchers and fieldworkers before the start of data collection. If the interviewer could not identify the correct location type on the list, they selected 'Other' and gave details. In the analysis, locations were excluded if it was considered likely that most all or of the time would have been spent outdoors (e.g., if the details given were 'gardening'). Several responses were re-coded, if it was considered plausible from the details given that the location was covered by one of the original options (e.g., 'domestic worker' was changed to 'House off plot'). Several new location categories were added, if reported by multiple respondents (e.g., 'factory'). Finally, remaining responses in the 'Other' category were recorded as 'Other' if the type of location could be determined from the free text variable, and 'Missing' if it could not be.

Respondents were excluded from all casual contact time analyses if:

- 1) They reported visiting no locations (including their own home) and using no transport
- 2) The variable giving the total number locations visited was missing, and no information was provided on any locations visited
- 3) No information was available on any of the locations visited or transport used

Casual contact time was estimated as the duration of time that respondents reported spending in a location, multiplied by the number of people that they estimated were present at

the location, halfway through the time that they were there. In the analysis, total numbers of people present were capped at a maximum of 100, as above this value, it is unlikely that the respondent had sufficient contact with each person present to allow transmission. Estimates of numbers of adults and children present were reduced by the same proportion for each location, to give a maximum total number of people present of 100.

If the estimated number of people or children present was missing for a location, or if the estimated number of children present was greater than the estimated total number of people present, then the numbers of adults and children present were set equal to the mean reported number of adults and children present at locations of that type in the same community (KwaZulu-Natal or Western Cape). These numbers were rounded to the nearest whole number when generating bootstrap samples. If the duration of time spent in the location was missing, the duration was set equal to the mean duration for locations of that type.

Age mixing matrices for casual contact time were generated using the data on locations visited, the duration of time spent in the location, and the estimated number of adults and children present. Central estimates were generated using the method outlined in McCreesh *et al* (2). As data were collected on numbers of people and children present in indoor locations only, and not on the ages of adults present, the age distribution of adult casual contacts needed to be estimated. To do so, we assumed that the age distribution of adults present in each location type matched the age distribution of respondents who reported visiting locations of that type, weighted by the duration of time they reported spending in that location type and weighted to the sampled population age and sex distribution. To generate plausible ranges, 10,000 bootstrap samples were generated, re-sampling respondents with replacement within age categories, and sampling the ages of adults present by resampling with replacement respondents who reported visiting locations of that type (weighted by duration, and to the sampled age and sex distribution). The number of children present was set equal to number of children present reported by the respondent. Contact times were then estimated by multiplying the duration of time the respondent reported spending in each location by the sampled number of people of each age group present.

The age mixing matrices were made symmetric, using data on adult contact time with children to estimate child contact time with adults. As respondents were asked to estimate

numbers of children present who were aged under 15 years, we set the lowest respondent age group to be 15–19 years. As 15–17 year olds were not interviewed in KwaZulu-Natal, we assumed that contact patterns in 18–19 years olds were representative of contact patterns in all 15–19 year olds, and adjusted the weights accordingly.

#### **Estimating age-mixing patterns relevant for the transmission of droplet infections**

Age-mixing patterns relevant for the transmission of droplet infections were assumed to be equal to age mixing patterns calculated from close contact numbers.

#### **Estimating age-mixing patterns relevant for the transmission of airborne infections**

To generate age-mixing patterns relevant for the transmission of airborne infections, we summed estimated close contact time between household members, and estimated casual contact time occurring in locations other than the respondents' own houses. 95% plausible ranges were generated by pairing each of the 10,000 close contact household bootstrapped matrices with one of the 10,000 outside-household casual contact time bootstrapped matrices.

To allow direct comparisons to be made between the different age mixing matrices, the matrices for airborne infections (central estimate and 10,000 individual bootstrapped matrices) were then adjusted to give the same mean contact intensity between adults as the matrices for close contacts (using the central estimate matrices and 10,000 individual bootstrapped matrices respectively).

#### **Estimating age-mixing patterns relevant for the transmission of *Mycobacterium tuberculosis***

It is estimated that, in high incidence settings, only 8%–19% of tuberculosis comes from household transmission (2). Long durations of disease (3) also mean that transmission to household members may partially saturate, meaning that the relationship between contact time and transmission is nonlinear for household contacts. We therefore estimated age mixing matrices relevant to the transmission of *Mtb* by creating weighted averages of close contact numbers with household members, and casual contact time occurring outside respondents' own households.

For each pair of bootstrapped household close contact number and non-household casual contact time matrices, a proportion of contact that should occur in households was sampled from a uniform distribution between 8%–16% (the range of values found by different studies), with



12% used for the central estimate. A weighted average was then generated, given the desired proportion of overall ‘contact’ by adults occurring in households.

To allow direct comparisons to be made between the different age mixing matrices, the matrices for *Mtb* (central estimate and 10,000 individual bootstrapped matrices) were then adjusted to give the same mean contact intensity between adults as the matrices for close contacts (using the central estimate matrices and 10,000 individual bootstrapped matrices respectively).

### **Assortative mixing by age group**

We quantified the assortativeness of mixing by age group using the index  $Q$ , which takes the value of 1 when all contact occurs within age groups, and 0 when there is homogeneous mixing by age (4). 95% plausible intervals and p-values for the difference between assortativeness in the airborne and *Mtb* transmission matrices compared to the droplet transmission matrices were generated using the bootstrapped samples.

### **Sensitivity analyses**

In our main analysis, we cap the number of people at risk of infection in locations (i.e., casual contact numbers) at a maximum of 100. In the sensitivity analysis, we explore the effect of setting the cap at 20, or not having a cap.

When generating the non-saturating airborne and *Mtb* age-mixing matrices, we used close contact time and close contact number data respectively to estimate contact time between household members. In the sensitivity analysis, we explore the effect of using casual contact time.

For all sensitivity analysis age-mixing matrices, we rescaled contact times to give the same overall mean contact hours per adult in the sensitivity analysis as in the main analysis. This was done because the primary use of age mixing matrices is in mathematical modeling, where it is usually the relative values of the cells in the matrices that has an impact on model dynamics, not the absolute values.

## Results

### Comparison of employment status with target population, KwaZulu-Natal

19.3% (329/1704) of social contact survey respondents gave their employment status as 'Full time' in the DSA census, compared to 19.5% (5898/30259) of the target population; and 4.0% (69/1704) of respondents gave their employment status as 'Part time', compared to 3.8% (1139/30259) of the target population.

### Missing/incomplete data

#### Close contact data

- The reported number of household contacts was higher than the reported total number for two respondents in KwaZulu-Natal. They were excluded from all analyses of close contacts.
- Three respondents in KwaZulu-Natal had the contact age missing for all of their contacts, and were excluded from the age-mixing analysis.
- The number of household contacts was unknown for 11 respondents in KwaZulu-Natal. They were not included in the analysis of household or non-household contacts (Appendix Figures 10–13).
- Contact ages were missing for 54 contacts in KwaZulu-Natal and 16 contacts in Western Cape
- Whether a contact is a household member was missing for 16 contacts in KwaZulu-Natal and two in Western Cape
- Contact durations were missing for 16 contacts in KwaZulu-Natal and three contacts in Western Cape. No respondents were missing the duration of time for all contacts.

#### Casual contact data

- One person in KwaZulu-Natal and eight in Western Cape reported that they visited no locations (including their own house) and used no transport used. This was considered to be implausible, and they were excluded from the casual contact time analyses.

- The ‘total number of locations visited’ variable was missing for four respondents in Western Cape, and they gave no information on any locations visited. They were excluded from the casual contact time analyses.
- 43 respondents in Western Cape reported the total number of locations visited, but did not give any information on any of the locations visited. They did not differ from other respondents in Western Cape in respect to the reported number of locations visited (no details given: mean locations = 3.0, 95% CI 2.5–3.5, detail given: 3.0 (2.9–3.1), and were excluded from the casual contact time analyses.
- The number of people and/or number of children present was missing for 102 locations in KwaZulu-Natal and 220 in Western Cape. The reported number of children present was greater than reported total number of people present for seven locations in KwaZulu-Natal and none in Western Cape.
- The duration of time spent in a location was missing for two locations in Western Cape and none in KwaZulu-Natal.

#### **Assortative mixing by age group**

Using the data from KwaZulu-Natal, the index of assortativeness  $Q$  took the value of 0.14 (95% plausible range 0.11 – 0.17) for the droplet transmission matrix, 0.10 (0.08–0.12) for the airborne transmission matrix ( $p = 0.01$ , when compared to the droplet transmission matrix), and 0.12 (0.09 – 0.13) for the Mtb transmission matrix ( $p = 0.09$ ). Using the data from Western Cape, the index of assortativeness  $Q$  took the value of 0.24 (95% plausible range 0.20 – 0.28) for the droplet transmission matrix, 0.21 (0.17–0.25) for the airborne transmission matrix ( $p = 0.13$ ), and 0.23 (0.19 – 0.26) for the Mtb transmission matrix ( $p = 0.24$ ).

#### **Sensitivity analyses**

Changing the cap on people at risk in locations

Overall casual contact time was lower when the number of people at risk was capped at 20, and higher when it was not capped, compared to when it was capped at 100 people (Appendix Figure 1). Changing the cap had a moderate effect on casual contact time age-mixing patterns, although most changes were not large compared to the breadths of the 95% plausible ranges (Appendix Figure 2).

## Generating age-mixing matrices using casual contact time for household contact

Calculating age-mixing matrices relevant for non-saturating airborne and *Mtb* transmission using casual contact time data for contact in between household members had very little effect on estimated age-mixing patterns (Supporting information, Appendix Figures 5, 6). The exception to this was contact relevant to airborne transmission between 15–19 year olds in KwaZulu-Natal, which was lower when casual contact time data were used than in the main analysis.

## References

1. Funk S. socialmixr: social mixing matrices for infectious disease modelling. Version 0.1.8. 2020 [cited 2021 Jan 1]. <https://cran.r-project.org/web/packages/socialmixr/index.html>
2. McCreesh N, Morrow C, Middelkoop K, Wood R, White RG. Estimating age-mixing patterns relevant for the transmission of airborne infections. *Epidemics*. 2019;28:100339. [PubMed](#) <https://doi.org/10.1016/j.epidem.2019.03.005>
3. Ku C-C, MacPherson P, Khundi M, Nzawa R, Feasey HR, Nliwasa M, et al. Estimated durations of asymptomatic, symptomatic, and care-seeking phases of tuberculosis disease. *BMC Med*. 2021;19:298. [PubMed](#) <https://doi.org/10.1186/s12916-021-02128-9>
4. Gupta S, Anderson RM, May RM. Networks of sexual contacts: implications for the pattern of spread of HIV. *AIDS*. 1989;3:807–17. [PubMed](#) <https://doi.org/10.1097/00002030-198912000-00005>



**Appendix Table 1.** Mean overall, household, and non-household close contact numbers in KwaZulu-Natal. P-values were calculated using a Wald test.

Characteristic		Overall		Household members		Non-household members	
		Mean (95% CI)	p-value	Mean (95% CI)	p-value	Mean (95% CI)	p-value
Sex	Male	7.0 (6.6–7.4)	0.02	3.8 (3.6–4.0)	<0.001	3.2 (2.8–3.5)	0.007
	Female	7.6 (7.3–8.0)		5.1 (4.9–5.3)		2.5 (2.2–2.9)	
Age	15–19	7.9 (7.0–9)	<0.001	4.9 (4.3–5.5)	0.08	2.9 (2.1–3.8)	0.01
	20–29	8.2 (7.6–8.7)		4.8 (4.5–5.1)		3.3 (2.9–3.9)	
	30–39	7.4 (6.9–8.0)		4.5 (4.2–4.9)		2.9 (2.5–3.4)	
	40–49	6.5 (6.0–7.0)		4.2 (3.8–4.6)		2.4 (2.0–2.8)	
	50+	6.8 (6.3–7.4)		4.5 (4.2–4.8)		2.3 (1.8–2.8)	
Residence	Peri-Urban	7.0 (6.6–7.5)	0.1	4.3 (4.0–4.5)	<0.001	2.7 (2.4–3.1)	0.1
	Urban	7.6 (7.3–8.0)		5.0 (4.8–5.2)		2.6 (2.4–2.9)	
	Rural	7.4 (6.0–9.1)		3.5 (3.1–3.9)		3.9 (2.7–5.8)	
Monthly household income	<R1000	6.8 (6.3–7.4)	0.003	4.2 (4.0–4.5)	<0.001	2.5 (2.1–3.1)	0.09
	R1000–R2500	7.6 (7.2–7.9)		5.0 (4.8–5.3)		2.6 (2.3–2.9)	
	R2500–R5000	7.2 (6.5–7.9)		4.2 (3.8–4.5)		3.1 (2.5–3.8)	
	R5000–R10000	7.1 (6.3–8.0)		4.2 (3.7–4.9)		2.9 (2.3–3.6)	
	>R10000	9.7 (7.5–13)		4.6 (3.8–5.7)		5.1 (3–8.6)	
	Unknown/missing	5.5 (4.3–7.0)		2.5 (1.8–3.4)		3.0 (2.0–4.3)	
Employment	Full-time	7.4 (6.7–8.2)	0.9	3.8 (3.5–4.2)	<0.001	3.6 (2.9–4.4)	0.007
	Part-time	7.1 (6.1–8.2)		4.0 (3.3–4.8)		3.0 (2.4–3.8)	
	None/missing	7.4 (7.1–7.6)		4.8 (4.6–5.0)		2.5 (2.3–2.8)	
Household size	1	5.5 (4.9–6.2)	<0.001	2.9 (2.5–3.3)	<0.001	2.6 (2.1–3.3)	0.5
	2–4	6.2 (5.5–7.0)		3.3 (2.9–3.6)		3.0 (2.3–3.8)	
	5–7	7.2 (6.7–7.7)		4.2 (4.0–4.4)		3.0 (2.6–3.6)	
	8–10	7.7 (7.1–8.3)		5.2 (4.9–5.4)		2.5 (2.1–3.0)	
	11+	8.5 (8.0–9.0)		5.8 (5.4–6.2)		2.6 (2.3–3.0)	
Day reported	Sunday	7.6 (6.8–8.4)	0.2	4.5 (4.1–4.9)	0.9	3.2 (2.5–4.0)	0.005
	Monday	7.5 (6.7–8.4)		4.6 (4.2–5.1)		2.9 (2.2–3.8)	
	Tuesday	6.9 (6.4–7.4)		4.8 (4.4–5.2)		2.0 (1.7–2.4)	
	Wednesday	7.2 (6.7–7.7)		4.6 (4.2–5.0)		2.6 (2.2–3.0)	
	Thursday	8.1 (7.3–9.1)		4.6 (4.2–5.1)		3.4 (2.7–4.4)	
	Friday	7.0 (6.5–7.7)		4.5 (4.1–4.9)		2.6 (2.1–3.1)	
	Saturday	7.2 (6.5–7.9)		4.5 (4.1–4.9)		2.7 (2.2–3.4)	
Total		7.4 (7.1–7.6)		4.6 (4.4–4.7)		2.8 (2.5–3.0)	

**Appendix Table 2.** Mean overall, household, and non-household close contact numbers in Western Cape. P-values were calculated using a Wald test.

Characteristic		Overall		Household members		Non-household members	
		Mean (95% CI)	p-value	Mean (95% CI)	p-value	Mean (95% CI)	p-value
Sex	Male	7.9 (7.3–8.4)	0.3	1.8 (1.7–2.0)	<0.001	6.0 (5.5–6.6)	0.5
	Female	8.3 (7.6–9.2)		2.6 (2.5–2.8)		5.7 (5.0–6.6)	
Age	15–19	11 (9.3–13)	0.004	3.1 (2.8–3.5)	<0.001	8.0 (6.3–10)	0.05
	20–29	7.8 (7.2–8.4)		1.9 (1.7–2.1)		5.9 (5.3–6.5)	
	30–39	7.7 (6.8–8.7)		2.0 (1.8–2.2)		5.7 (4.8–6.7)	
	40–49	7.8 (7.0–8.6)		2.6 (2.3–3.0)		5.2 (4.4–6.0)	
	50+	8.7 (6.7–11)		2.9 (2.4–3.5)		5.8 (3.8–8.7)	
Residence	Peri-Urban	8.1 (7.6–8.6)		2.2 (2.1–2.3)		5.9 (5.4–6.4)	
	Urban	NA		NA		NA	
	Rural	NA		NA		NA	
Monthly household income	<R1000	9.2 (6.5–13)	0.07	2.0 (1.7–2.5)	<0.001	7.2 (4.6–11)	0.03
	R1000–R2500	7.8 (7.1–8.6)		2.1 (1.8–2.3)		5.7 (5.1–6.5)	
	R2500–R5000	8.6 (8.0–9.3)		2.2 (2–2.4)		6.4 (5.8–7.1)	
	R5000–R10000	7.2 (6.3–8.2)		2.0 (1.8–2.3)		5.2 (4.3–6.2)	
	>R10000	8.5 (6.7–11)		2.7 (2.3–3.2)		5.8 (4.1–8.2)	
	Unknown/missing	7.1 (6.2–8.1)		2.7 (2.4–3.1)		4.4 (3.5–5.4)	
Employment	Full-time	8.3 (7.7–9.0)	0.3	2.0 (1.9–2.2)	<0.001	6.3 (5.7–7.0)	0.03
	Part-time	8.5 (7.1–10)		1.9 (1.7–2.2)		6.6 (5.2–8.4)	
	None/missing	7.6 (7.0–8.3)		2.5 (2.4–2.7)		5.1 (4.5–5.8)	

Characteristic		Overall		Household members		Non-household members	
		Mean (95% CI)	p-value	Mean (95% CI)	p-value	Mean (95% CI)	p-value
Household size	1	7.0 (6.1–8.0)	<0.001	0.05 (0.02–0.12)	<0.001	7.0 (6.1–7.9)	<0.001
	2–4	7.9 (7.2–8.6)		1.9 (1.9–2.0)		5.9 (5.3–6.6)	
	5–7	9.5 (8.7–11)		4.5 (4.4–4.7)		5.0 (4.2–6.0)	
	8–10	10 (8.8–11)		7.0 (6.3–7.9)		3.0 (2.0–4.5)	
	11+	15 (9.6–22)		9.1 (6.7–13)		5.4 (1.9–15)	
Day reported	Sunday	8.1 (7.3–8.9)	0.5	2.0 (1.8–2.3)	0.08	6.1 (5.3–6.9)	0.8
	Monday	8.6 (7.7–9.6)		2.4 (2.2–2.7)		6.2 (5.3–7.3)	
	Tuesday	7.8 (7.0–8.8)		2.2 (1.9–2.5)		5.6 (4.8–6.6)	
	Wednesday	8.5 (7.3–9.9)		2.2 (2.0–2.5)		6.2 (5.1–7.7)	
	Thursday	8.0 (7.0–9.1)		2.0 (1.8–2.4)		5.9 (5.0–7.0)	
	Friday	7.1 (6.1–8.3)		2.0 (1.6–2.4)		5.2 (4.2–6.3)	
	Saturday	8.5 (6.5–11)		2.6 (2.2–3.0)		5.9 (4.1–8.7)	
Total		8.1 (7.6–8.6)		2.2 (2.1–2.3)		5.9 (5.4–6.4)	

**Appendix Table 3.** Mean overall, household, and non-household close contact time (hours) in KwaZulu-Natal. P-values were calculated using a Wald test.

Characteristic		Overall		Household members		Non-household members	
		Mean (95% CI)	p-value	Mean (95% CI)	p-value	Mean (95% CI)	p-value
Sex	Male	56 (52–59)	<0.001	40 (37–43)	<0.001	16 (14–18)	0.2
	Female	82 (78–86)		67 (64–71)		14 (12–16)	
Age	15–19	71 (62–81)	0.08	52 (42–63)	0.07	19 (15–25)	0.1
	20–29	74 (69–79)		58 (53–63)		16 (14–18)	
	30–39	70 (64–77)		55 (49–61)		15 (13–18)	
	40–49	63 (58–69)		50 (44–56)		13 (11–16)	
	50+	74 (68–80)		61 (56–66)		13 (10–17)	
Residence	Peri-Urban	70 (66–75)	0.01	54 (50–59)	<0.001	16 (13–18)	0.4
	Urban	74 (70–78)		60 (56–64)		14 (13–15)	
	Rural	60 (52–68)		43 (37–51)		16 (12–22)	
Monthly household income	<R1000	72 (66–78)	<0.001	61 (56–67)	<0.001	11 (9–13)	<0.001
	R1000–R2500	81 (77–86)		69 (65–74)		12 (10–13)	
	R2500–R5000	53 (48–59)		33 (30–37)		20 (16–26)	
	R5000–R10000	56 (50–63)		30 (25–37)		26 (22–31)	
	>R10000	73 (62–85)		41 (32–53)		32 (24–41)	
	Unknown/missing	34 (28–40)		23 (16–34)		10 (6–18)	
Employment	Full-time	64 (58–71)	0.05	45 (40–50)	<0.001	19 (15–25)	0.02
	Part-time	67 (54–82)		51 (38–67)		16 (12–21)	
	None/missing	73 (70–77)		60 (57–63)		14 (12–15)	
Household size	1	47 (40–54)	<0.001	36 (30–43)	<0.001	11 (8–15)	0.2
	2–4	59 (52–67)		43 (37–49)		16 (11–22)	
	5–7	68 (65–73)		53 (49–57)		15 (13–18)	
	8–10	74 (69–79)		61 (56–66)		13 (11–16)	
	11+	87 (80–94)		71 (64–78)		16 (14–18)	
Day reported	Sunday	71 (64–78)	0.8	56 (49–63)	0.9	15 (12–18)	0.4
	Monday	70 (62–79)		53 (47–61)		17 (12–25)	
	Tuesday	70 (64–77)		58 (51–65)		12 (10–15)	
	Wednesday	73 (66–80)		58 (51–66)		15 (12–17)	
	Thursday	75 (67–84)		58 (51–67)		17 (14–21)	
	Friday	67 (61–74)		53 (47–60)		13 (11–16)	
	Saturday	73 (66–81)		59 (52–67)		14 (12–18)	
Total		71 (69–74)		56 (54–59)		15 (14–16)	

**Appendix Table 4.** Mean overall, household, and non-household close contact time (hours) in Western Cape. P-values were calculated using a Wald test.

Characteristic		Overall		Household members		Non-household members	
		Mean (95% CI)	p-value	Mean (95% CI)	p-value	Mean (95% CI)	p-value
Sex	Male	49 (46–53)	<0.001	27 (24–30)	<0.001	22 (20–25)	<0.3
	Female	60 (56–64)		40 (37–43)		20 (18–23)	
Age	15–19	80 (68–95)	<0.001	46 (39–53)	<0.001	35 (25–49)	0.006
	20–29	51 (47–56)		28 (25–32)		23 (19–27)	
	30–39	49 (45–53)		31 (28–34)		18 (16–21)	
	40–49	58 (51–66)		38 (32–46)		20 (16–25)	
	50+	63 (52–77)		44 (35–57)		19 (12–31)	
Residence	Peri-Urban	55 (52–57)	NA	33 (31–35)	NA	21 (19–24)	NA
	Urban	NA		NA		NA	
	Rural	NA		NA		NA	
Monthly household income	<R1000	58 (48–71)	0.001	37 (30–45)	<0.001	22 (16–30)	0.03
	R1000–R2500	54 (49–60)		35 (31–40)		19 (16–23)	
	R2500–R5000	60 (55–66)		34 (31–38)		26 (23–30)	
	R5000–R10000	41 (35–48)		24 (22–28)		16 (12–24)	
	>R10000	51 (42–62)		29 (25–35)		21 (14–33)	
	Unknown/missing	57 (51–64)		39 (33–46)		18 (13–24)	
Employment	Full-time	53 (49–58)	0.3	28 (25–31)	<0.001	26 (22–29)	0.005
	Part-time	52 (46–59)		31 (26–36)		21 (17–25)	
	None/missing	57 (53–62)		40 (36–43)		18 (15–21)	
Household size	1	27 (23–32)	<0.001	0.6 (0.3–1.3)	<0.001	26 (22–31)	0.009
	2–4	51 (48–54)		29 (28–31)		21 (19–24)	
	5–7	83 (78–88)		66 (62–70)		17 (13–22)	
	8–10	124 (109–141)		114 (97–134)		10 (5–21)	
	11+	200 (138–291)		172 (112–264)		28 (7–112)	
Day reported	Sunday	53 (46–60)	0.2	34 (30–52)	0.08	18 (14–23)	0.3
	Monday	60 (54–78)		34 (30–52)		26 (21–29)	
	Tuesday	52 (46–67)		32 (28–49)		20 (16–27)	
	Wednesday	57 (49–52)		32 (28–47)		24 (18–22)	
	Thursday	54 (48–66)		30 (26–43)		23 (19–28)	
	Friday	47 (40–45)		27 (23–33)		20 (14–19)	
	Saturday	60 (52–53)		42 (34–37)		18 (14–22)	
Total		55 (52–57)		33 (31–35)		21 (19–24)	

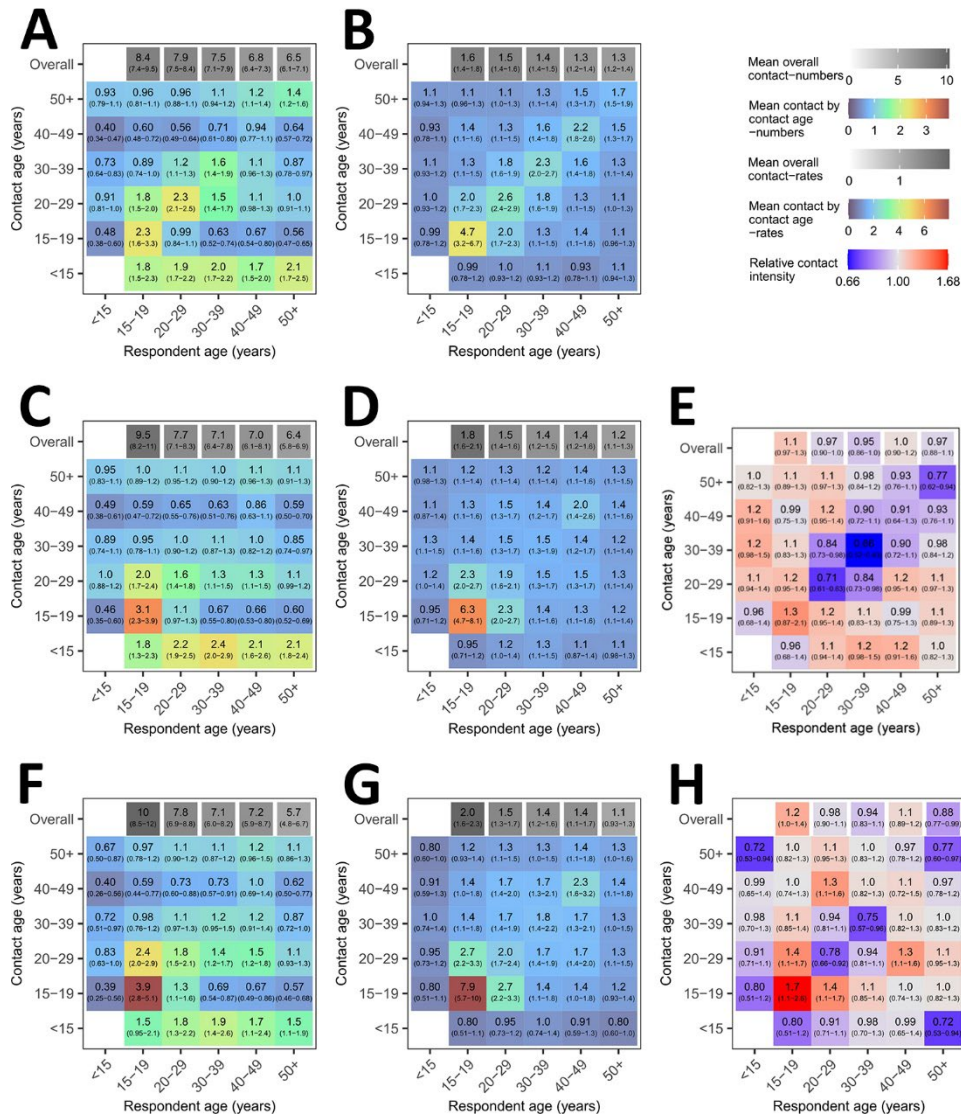
**Appendix Table 5.** Mean overall, household, and non-household casual contact time (hours) in KwaZulu-Natal. P-values were calculated using a Wald test.

Characteristic		Overall		Household members		Non-household members	
		Mean (95% CI)	p-value	Mean (95% CI)	p-value	Mean (95% CI)	p-value
Sex	Male	171 (158–185)	<0.001	86 (81–91)	<0.001	85 (73–99)	0.9
	Female	203 (191–216)		119 (114–125)		84 (73–97)	
Age	15–19	230 (195–270)	0.03	100 (86–115)	0.007	130 (98–172)	0.01
	20–29	204 (188–221)		108 (101–116)		95 (80–113)	
	30–39	182 (164–202)		105 (96–114)		77 (61–98)	
	40–49	170 (144–200)		91 (81–101)		79 (56–112)	
	50+	186 (170–204)		114 (107–122)		72 (57–91)	
Residence	Peri-Urban	177 (164–190)	0.03	98 (92–104)	<0.001	79 (67–92)	0.2
	Urban	203 (189–217)		117 (111–123)		86 (73–100)	
Monthly household income	Rural	186 (154–225)	<0.001	74 (65–85)	<0.001	112 (80–156)	<0.001
	<R1000	150 (136–166)		91 (85–99)		59 (46–74)	
	R1000–R2500	207 (193–222)		126 (119–132)		82 (69–97)	
	R2500–R5000	191 (170–214)		85 (77–94)		106 (85–131)	

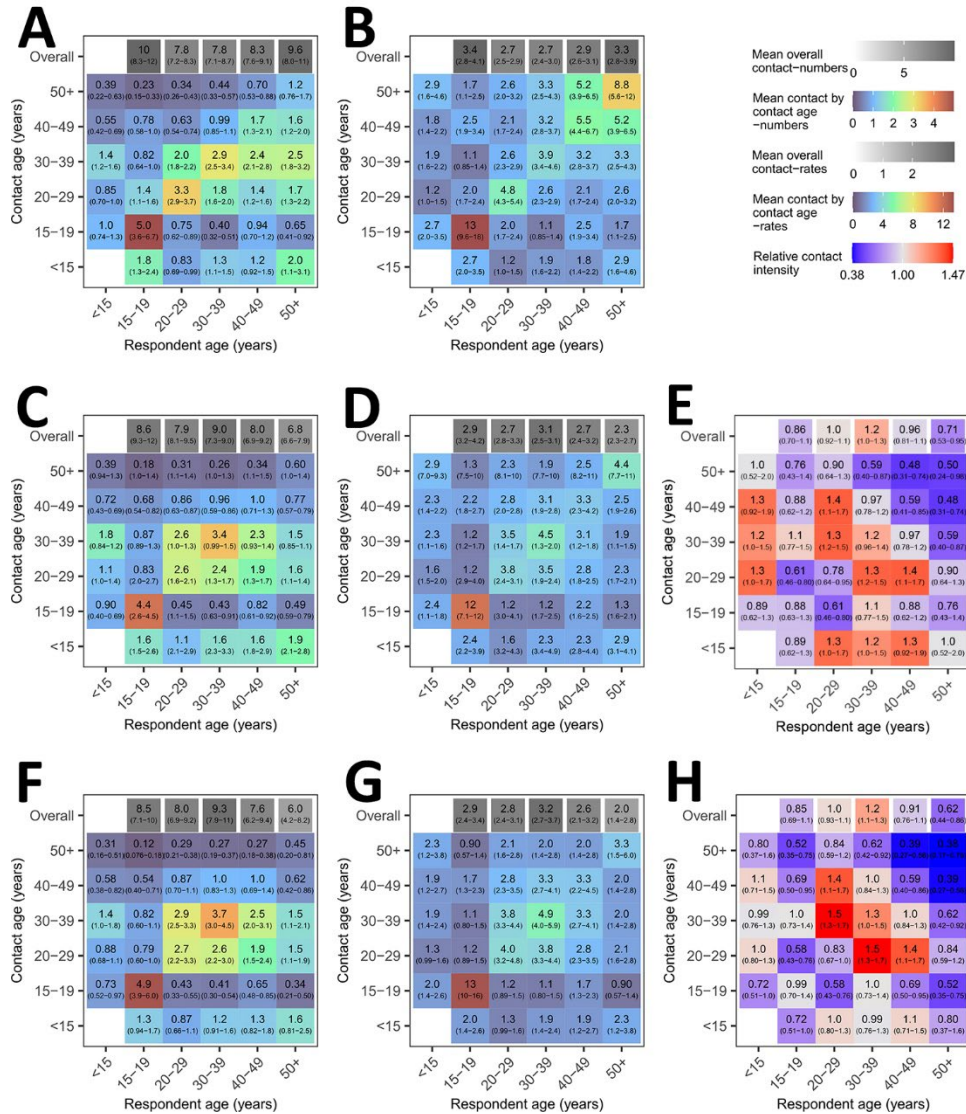
Characteristic		Overall		Household members		Non-household members	
		Mean (95% CI)	p-value	Mean (95% CI)	p-value	Mean (95% CI)	p-value
Employment	R5000–R10000	204 (170–245)	0.9	97 (83–113)	<0.001	108 (76–152)	0.05
	>R10000	222 (175–282)		91 (72–113)		132 (88–197)	
	Unknown/missing	204 (137–303)		69 (54–87)		135 (71–256)	
	Full-time	188 (165–214)		87 (79–95)		101 (80–129)	
	Part-time	206 (148–286)		82 (66–101)		124 (72–213)	
Household size	None/missing	191 (181–201)	0.03	112 (108–117)	<0.001	78 (69–88)	0.4
	1	157 (124–198)		65 (56–77)		91 (61–136)	
	2–4	176 (152–204)		78 (71–87)		98 (76–127)	
	5–7	182 (164–201)		94 (89–100)		87 (71–108)	
	8–10	204 (187–224)		119 (111–128)		85 (68–106)	
Day reported	11+	207 (191–223)	0.7	136 (126–146)	0.8	71 (58–87)	0.5
	Sunday	196 (170–225)		100 (91–110)		96 (72–127)	
	Monday	177 (157–200)		106 (96–118)		70 (52–95)	
	Tuesday	191 (165–220)		113 (101–126)		78 (57–107)	
	Wednesday	194 (171–221)		106 (96–117)		88 (66–118)	
	Thursday	178 (159–200)		106 (95–118)		72 (55–95)	
	Friday	199 (176–226)		103 (94–113)		96 (74–125)	
	Saturday	199 (176–226)		108 (98–118)		91 (70–120)	
Total	191 (182–200)	106 (102–110)	85 (76–94)				

**Appendix Table 6.** Mean overall, household, and non-household casual contact time (hours) in Western Cape. P-values were calculated using a Wald test.

Characteristic		Overall		Household members		Non-household members	
		Mean (95% CI)	p-value	Mean (95% CI)	p-value	Mean (95% CI)	p-value
Sex	Male	142 (125–162)	0.2	38 (34–42)	<0.001	104 (88–124)	0.8
	Female	161 (146–179)		56 (52–62)		105 (90–122)	
Age	15–19	177 (154–204)	0.1	56 (50–64)	0.01	121 (98–148)	0.2
	20–29	141 (123–162)		44 (39–49)		98 (81–119)	
	30–39	162 (140–188)		44 (39–50)		119 (97–144)	
	40–49	141 (115–173)		51 (42–61)		90 (67–122)	
	50+	123 (83–182)		57 (44–74)		66 (34–131)	
Residence	Peri-Urban	151 (140–164)	0.2	47 (44–50)	0.001	105 (93–117)	0.2
	Urban	NA		NA		NA	
	Rural	NA		NA		NA	
Monthly household income	<R1000	134 (105–171)	0.001	53 (39–73)	0.001	81 (55–118)	0.2
	R1000–R2500	160 (137–187)		56 (49–64)		105 (83–131)	
	R2500–R5000	166 (144–192)		44 (39–49)		122 (101–148)	
	R5000–R10000	131 (103–166)		36 (32–42)		95 (68–132)	
	>R10000	133 (103–173)		46 (37–58)		87 (59–128)	
Employment	Unknown/missing	141 (115–172)	0.001	46 (39–55)	<0.001	95 (70–129)	<0.001
	Full-time	159 (138–182)		37 (33–42)		122 (102–145)	
	Part-time	183 (152–221)		43 (35–52)		140 (111–178)	
Household size	None/missing	128 (116–141)	<0.001	59 (54–64)	<0.001	69 (59–82)	0.4
	1	95 (73–123)		12 (8–17)		83 (62–112)	
	2–4	154 (139–170)		42 (39–44)		113 (98–129)	
	5–7	186 (159–217)		88 (79–98)		98 (73–131)	
	8–10	235 (144–383)		133 (109–162)		102 (37–279)	
Day reported	11+	364 (183–722)	0.9	192 (114–325)	0.9	171 (35–829)	0.9
	Sunday	150 (123–51)		49 (42–49)		100 (75–32)	
	Monday	156 (130–55)		47 (41–59)		110 (85–36)	
	Tuesday	148 (122–50)		43 (37–55)		105 (80–34)	
	Wednesday	146 (120–50)		45 (39–52)		100 (75–31)	
	Thursday	137 (108–42)		46 (37–33)		90 (65–27)	
	Friday	149 (113–36)		48 (37–32)		101 (69–24)	
	Saturday	175 (141–47)		49 (39–34)		126 (93–32)	
Total	151 (140–164)	47 (44–50)	105 (93–117)				

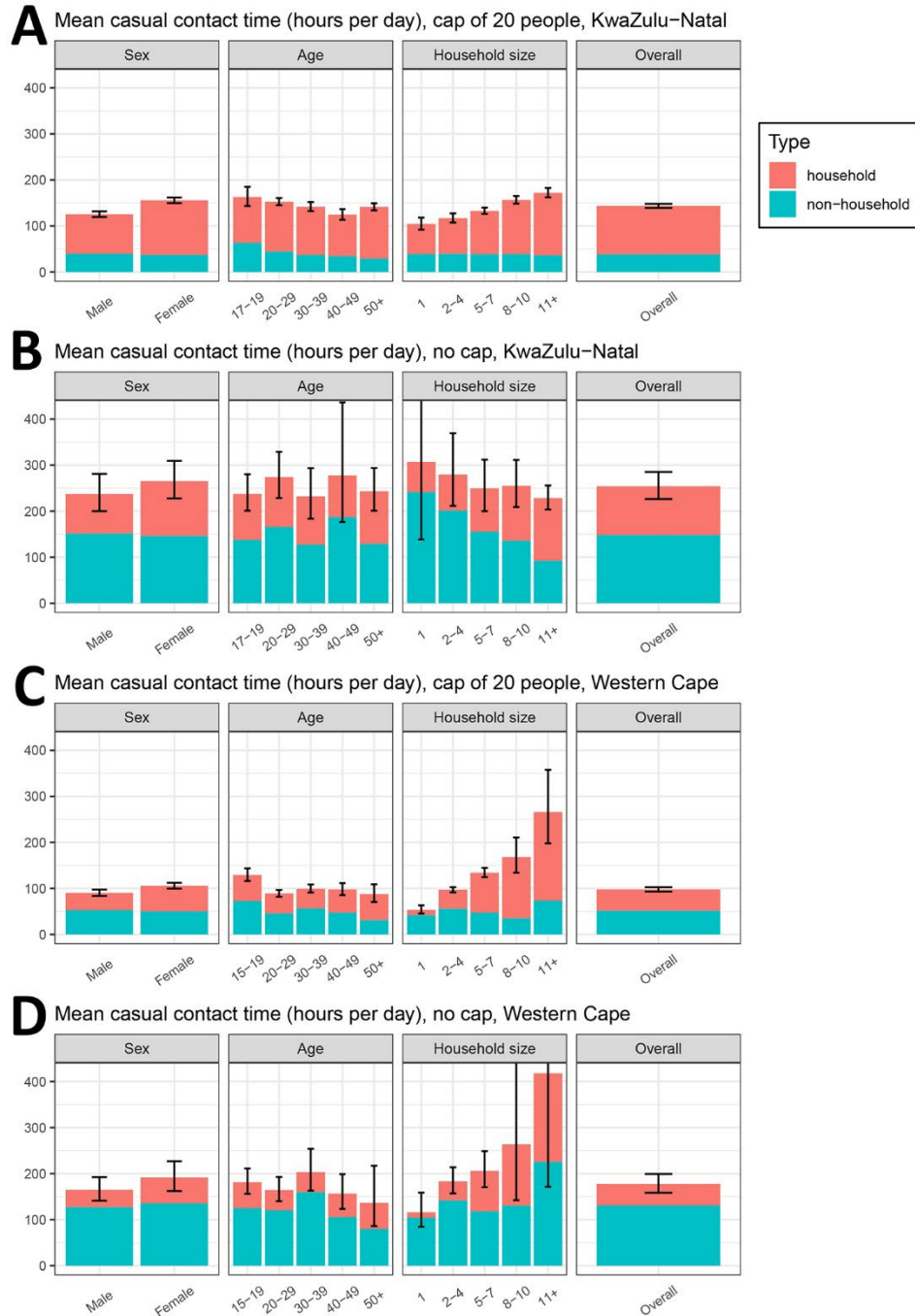


**Appendix Figure 1.** Age mixing matrices relevant for droplet transmission, non-saturating airborne transmission, and Mycobacterium tuberculosis transmission, for KwaZulu-Natal. Graphs a, c, and f show absolute contact intensities between respondents and contacts in each age group. Graphs b, d, and g show intensities of contact between each member of each age group. Graphs e and h show intensities for airborne infections and Mtb compared to intensities for droplet infections respectively. Numbers shown in graph a are the mean number of contacts respondents in each age group have with contacts in each age group per day. Numbers shown in graph b are the rate of contact between each individual in the population per day, expressed as rates  $\times 105$ . 'Numbers' and 'rates' in graphs c, d, f, and g are standardized so that the mean overall contact intensity by reported by adult respondents is the same as the mean number of overall close contacts reported by adult respondents (graph a). Contact numbers between child 'respondents' and contacts in each age group were estimated from data on contact between adult respondents and child contacts. Ranges shown are bootstrapped 95% plausible ranges.

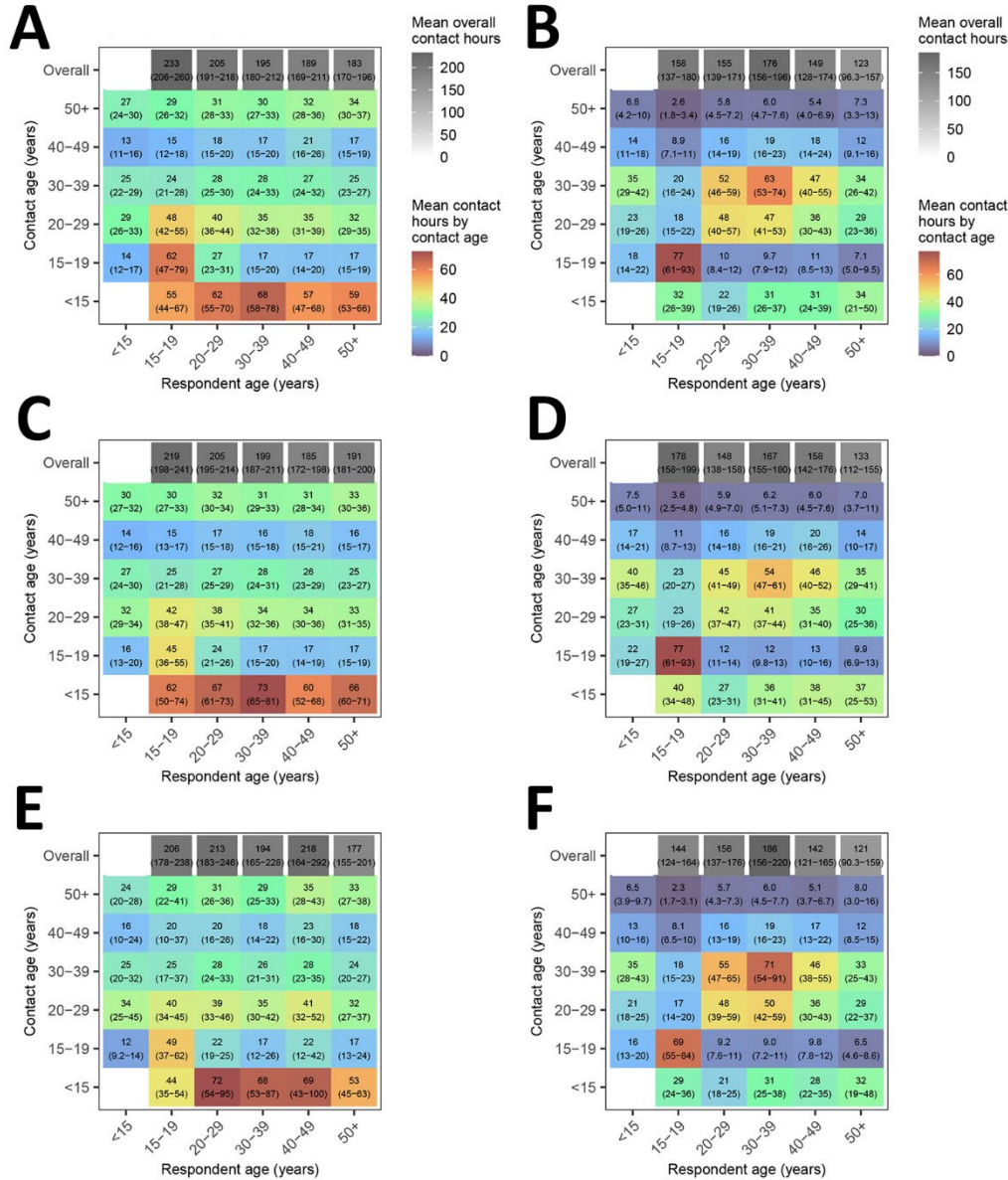


**Appendix Figure 2.** Age mixing matrices relevant for droplet transmission, non-saturating airborne transmission, and Mycobacterium tuberculosis transmission, for Western Cape. Graphs a, c, and f show absolute contact intensities between respondents and contacts in each age group. Graphs b, d, and g show intensities for each member of each age group. Graphs e and h show intensities for airborne infections and Mtb compared to intensities for droplet infections respectively. Numbers shown in graph a are the mean number of contacts respondents in each age group have with contacts in each age group per day. Numbers shown in graph b are the rate of contact between each individual in the population per day, expressed as rates  $\times 105$ . ‘Numbers’ and ‘rates’ in graphs c, d, f, and g are standardized so that the mean overall contact intensity by reported by adult respondents is the same as the mean number of overall close contacts reported by adult respondents (graph a). Contact numbers between child ‘respondents’ and contacts in each age group were estimated from data on contact between adult respondents and child contacts. Ranges shown are bootstrapped 95% plausible ranges.

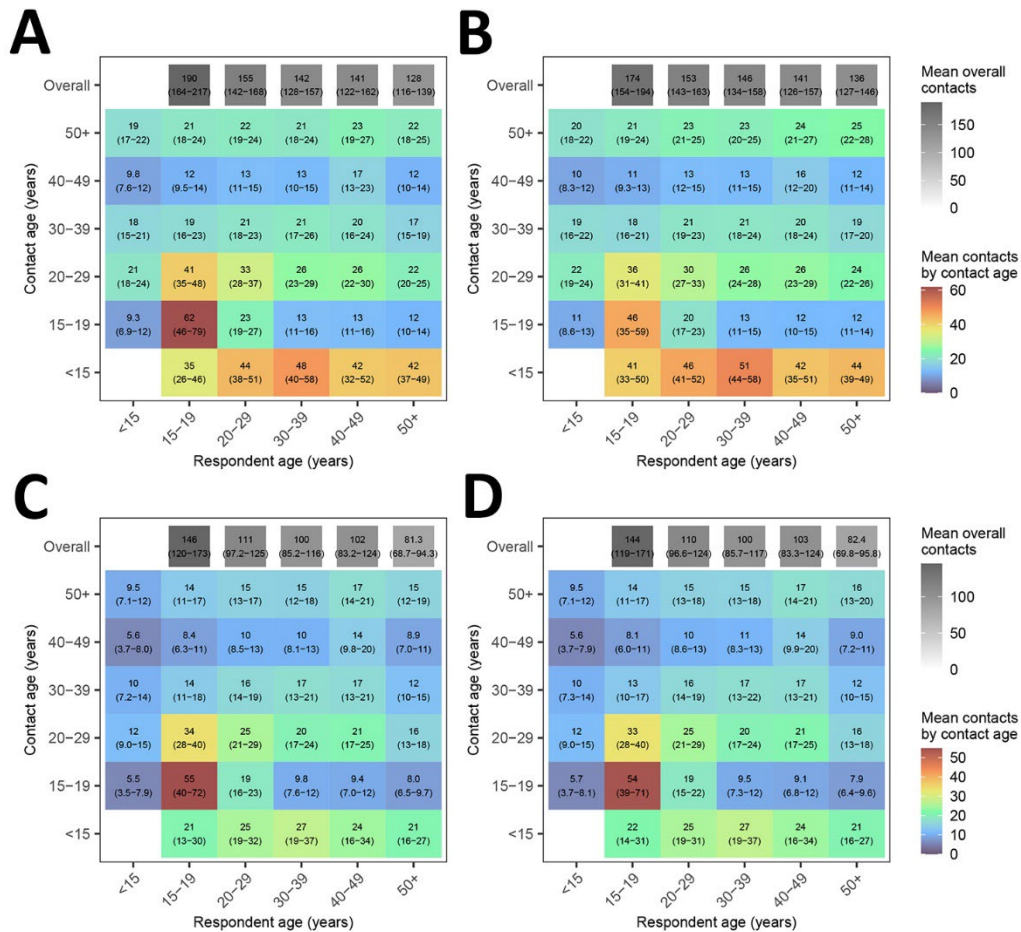




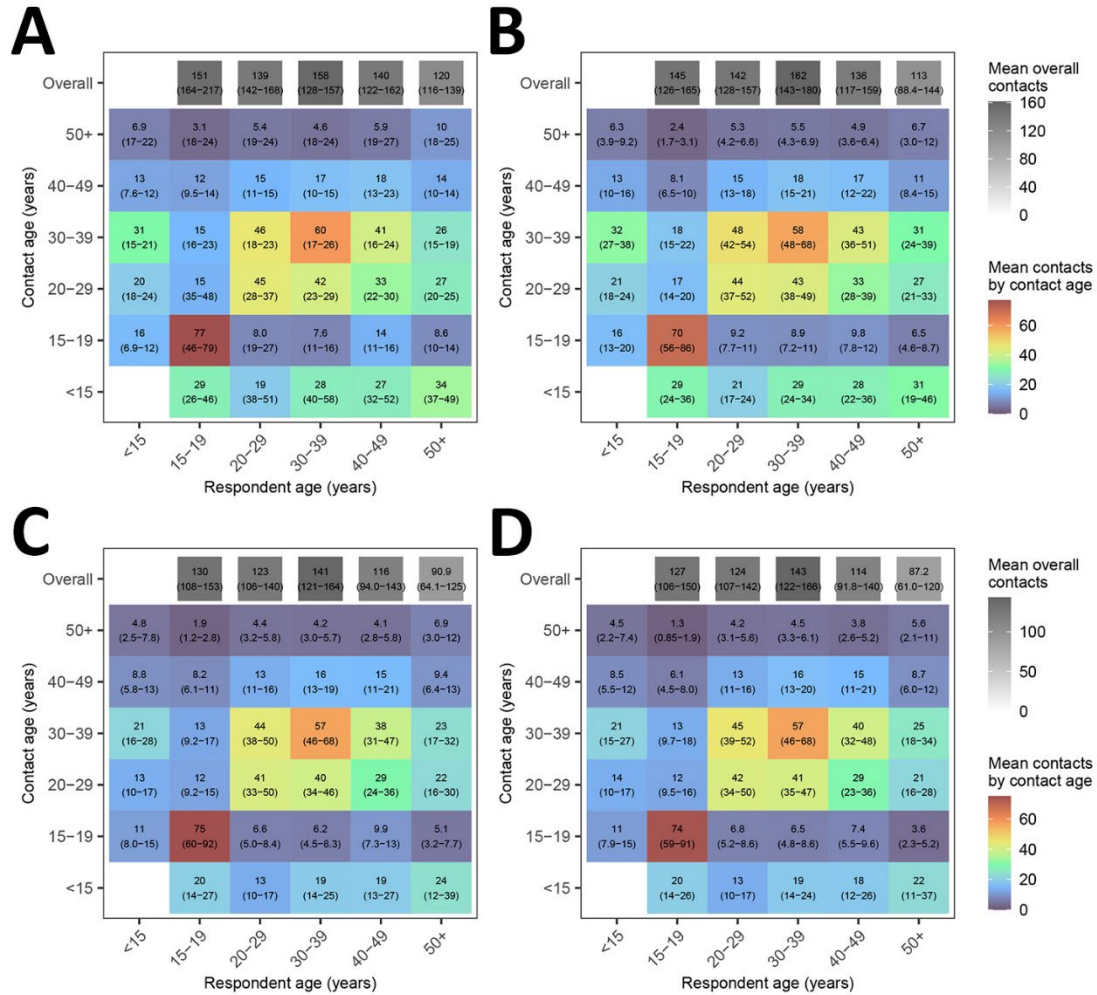
**Appendix Figure 3.** Household and non-household close contact numbers, close contact time, and casual contact time in KwaZulu-Natal and Western Cape, with a cap of 20 people at risk per location, and with no cap on the number at risk. Error bars show 95% confidence intervals for total contact numbers or time.



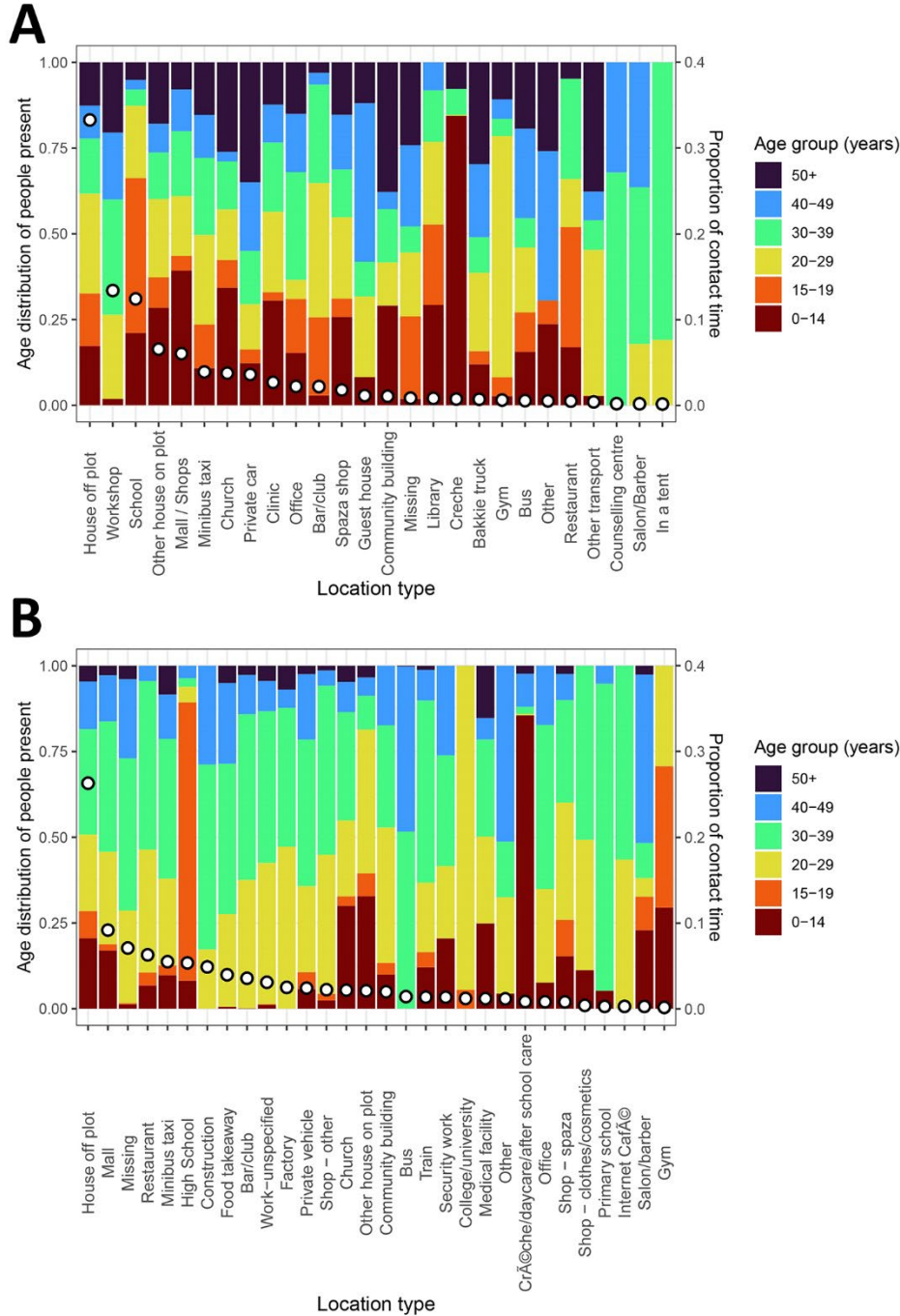
**Appendix Figure 4.** Casual contact time age mixing patterns for KwaZulu-Natal and Western Cape with a cap of 100 people at risk per location (baseline scenario), a cap of 20 people (sensitivity analysis), and no cap (sensitivity analysis). Panels a and b show estimated mean contact hours per day between respondents and contacts in each age group. Panels c-f show estimated mean contact hours, scaled to give the same overall mean adult contact hours per adult as panels a (c and e) and b (d and f).



**Appendix Figure 5.** Age mixing patterns relevant for non-saturating airborne transmission and the transmission of *Mycobacterium tuberculosis* in KwaZulu-Natal, with household age mixing estimated using close (baseline) and casual (sensitivity analysis) contact data. Panels show mean contact hours per day between respondents and contacts in each age group. Numbers in b and d are scaled to give the same overall mean adult contact hours per adult as panels a and c respectively.

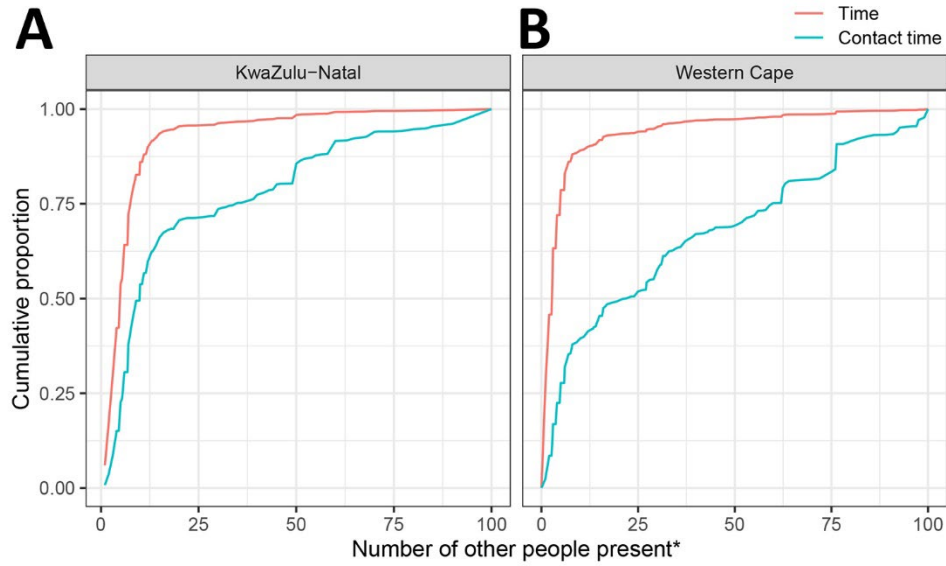


**Appendix Figure 6.** Age mixing patterns relevant for non-saturating airborne transmission and the transmission of *Mycobacterium tuberculosis* in Western Cape, with household age mixing estimated using close (baseline) and casual (sensitivity analysis) contact data. Panels show mean contact hours per day between respondents and contacts in each age group. Numbers in b and d are scaled to give the same overall mean adult contact hours per adult as panels a and c respectively of locations visited and transport used



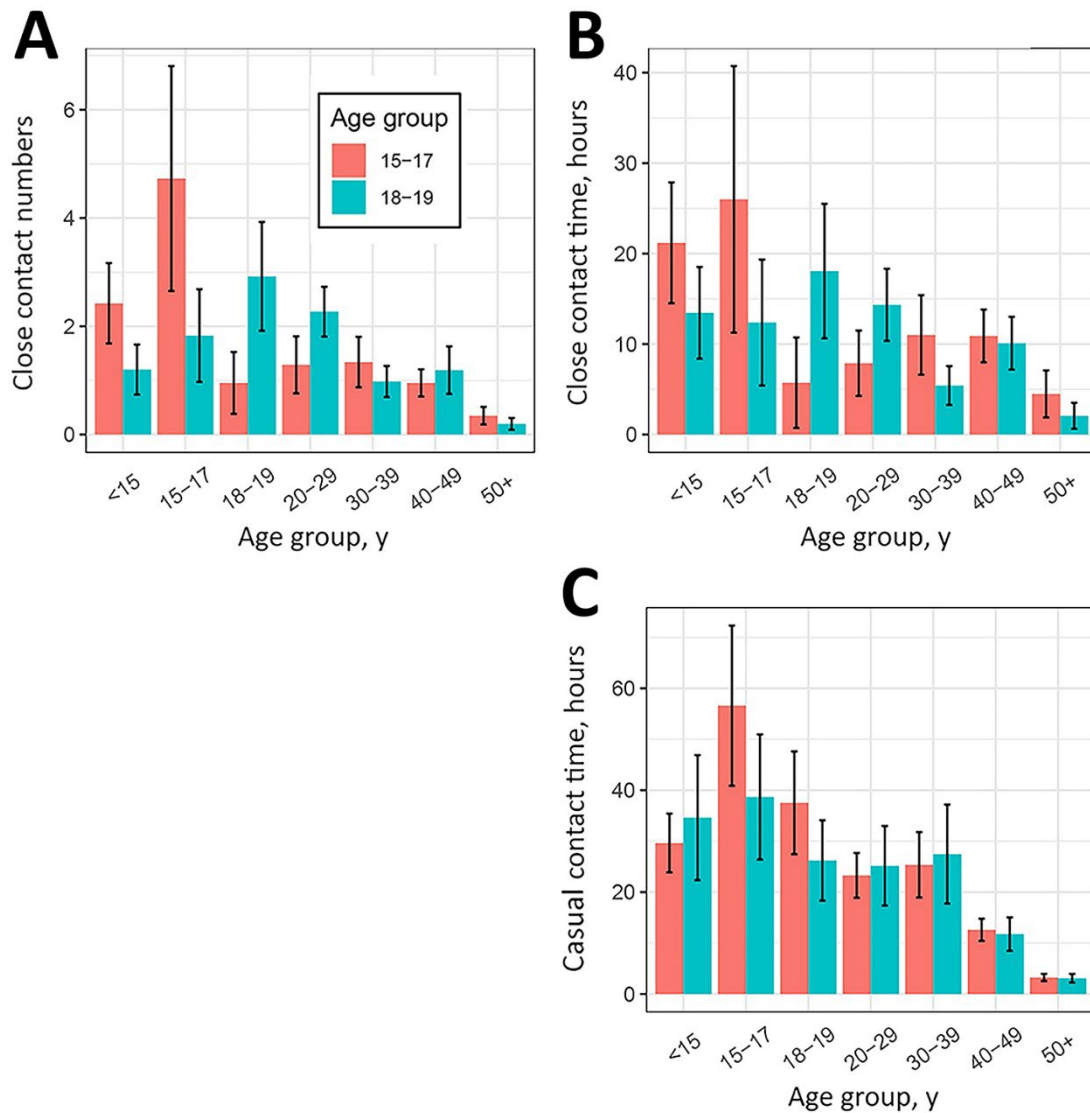
**Appendix Figure 7.** Age distribution of people present and proportion of contact time by location type, in KwaZulu-Natal and Western Cape. Bars (left axis) show the estimated proportion of people present in a location who are in each age group. White circles (right axis) show the estimated proportion of all contact time by adults in the community (excluding people’s own homes) that occurs in each type of location.



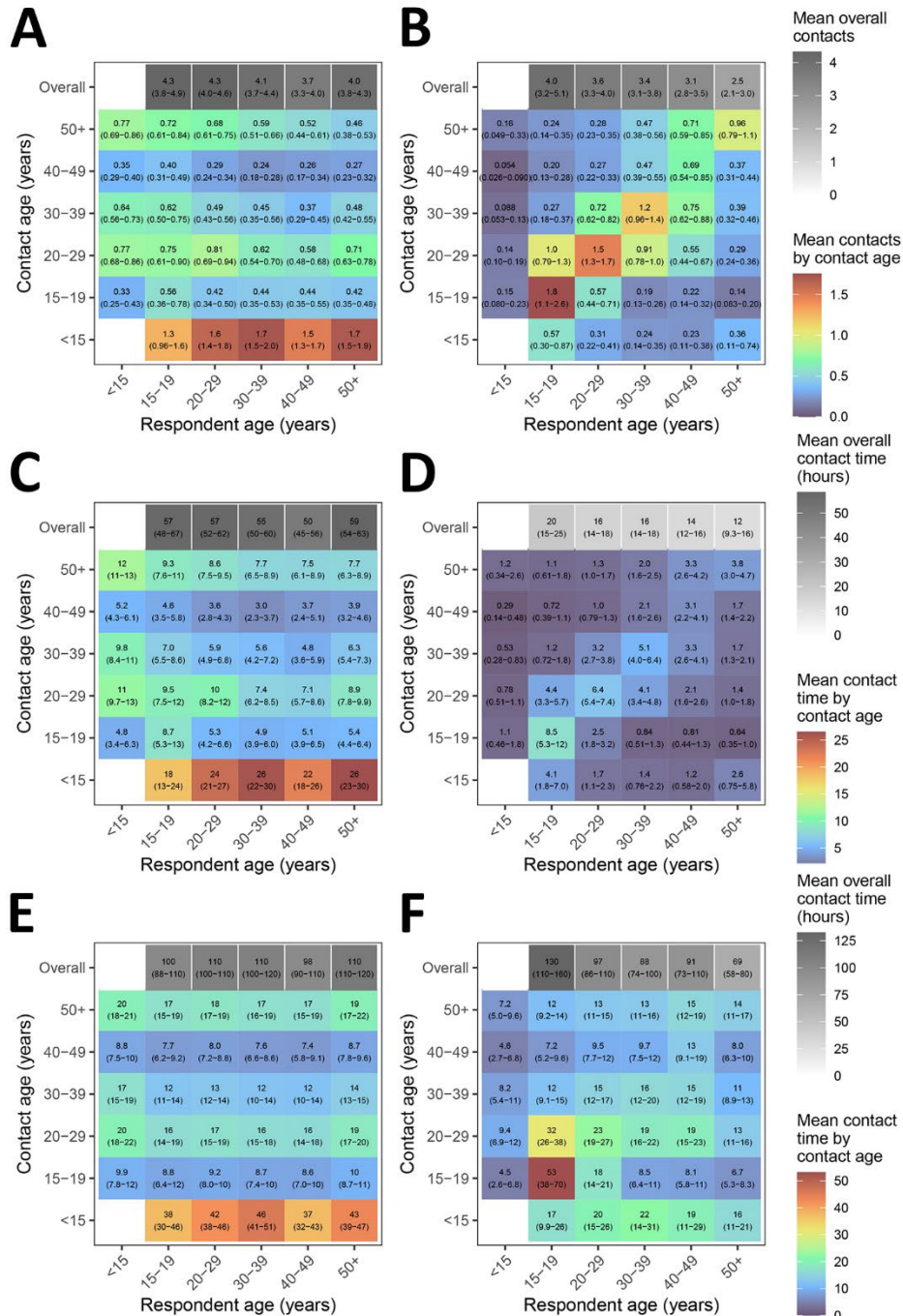


**Appendix Figure 8.** Cumulative proportion of time and casual contact time occurring in locations, by number of other people present, in KwaZulu-Natal and Western Cape. \*Number of other people present is capped at 100

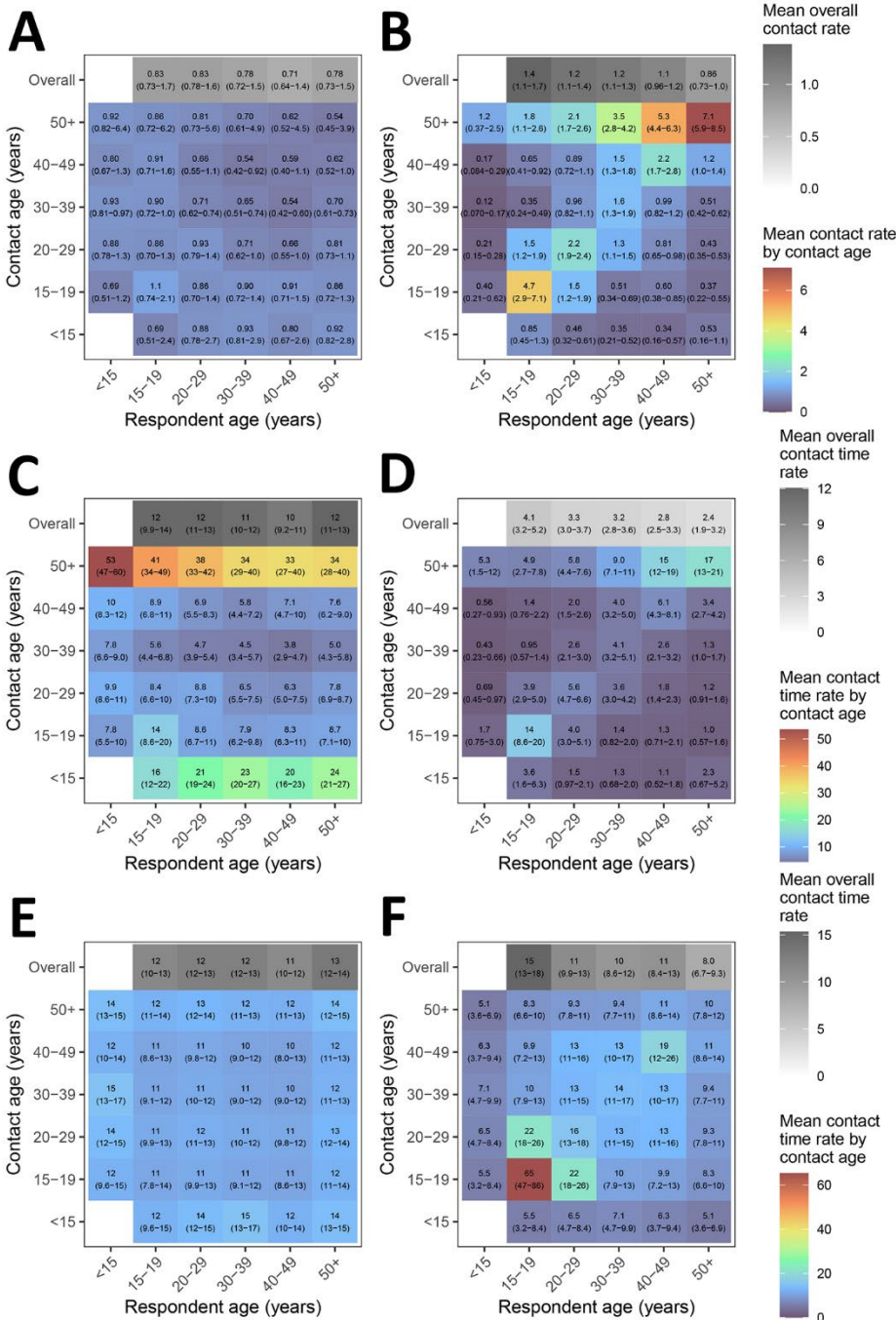




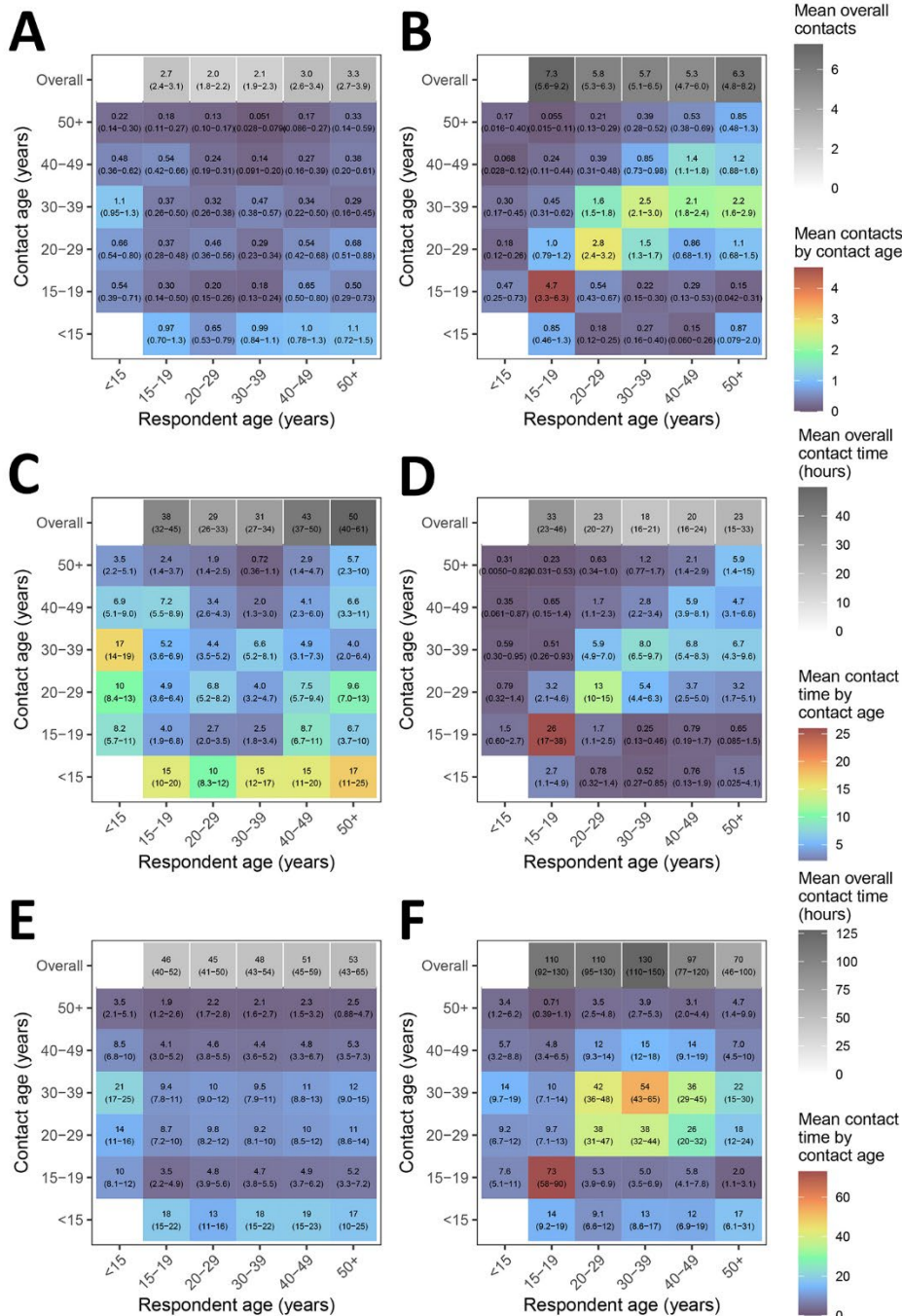
**Appendix Figure 9.** Adolescent contact patterns in Western Cape. Mean daily close contact numbers, close contact time, and casual contact time by contact age group for respondents aged 15–17 years and 18–19 years in Western Cape. Error bars show 95% confidence intervals.



**Appendix Figure 10.** Household and non-household close contact numbers, close contact time, and casual contact time age-mixing matrices for KwaZulu-Natal. Numbers shown in graphs a and b are the mean number of contacts respondents in each age group have with contacts in each age group per day. Numbers shown in graphs c-f are the mean number of contact hours respondents in each age group have with contacts in each age group per day. Contact numbers between child ‘respondents’ and contacts in each age group were estimated from data on contact between adult respondents and child contacts. Ranges shown are bootstrapped 95% plausible ranges.

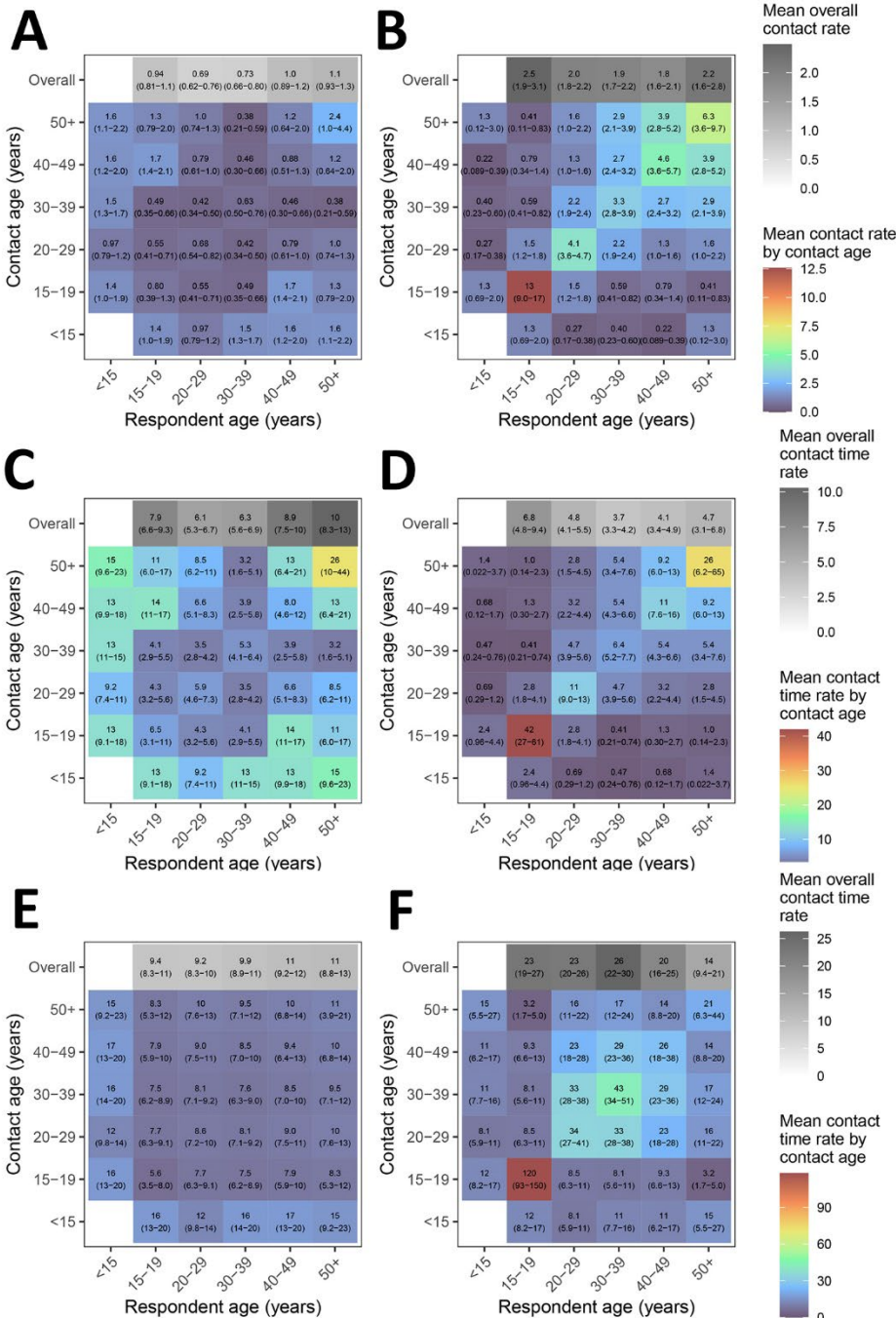


**Appendix Figure 11.** Household and non-household close contact numbers, close contact time, and casual contact time age-mixing matrices for KwaZulu-Natal. Numbers shown in graphs a and b are the rate of contact between each individual in the population per day, expressed as rates  $\times 10^5$ . Numbers shown in graphs c-f are the mean minutes of contact between each individual in the population per day, expressed as rates  $\times 10^2$ . Contact rates between child 'respondents' and contacts in each age group were estimated from data on contact between adult respondents and child contacts. Ranges shown are bootstrapped 95% plausible ranges.



**Appendix Figure 12.** Household and non-household close contact numbers, close contact time, and casual contact time age-mixing matrices for Western Cape. Numbers shown in graphs a and b are the mean number of contacts respondents in each age group have with contacts in each age group per day. Numbers shown in graphs c-f are the mean number of contact hours respondents in each age group have with contacts in each age group per day. Contact numbers between child ‘respondents’ and contacts in each age group were estimated from data on contact between adult respondents and child contacts. Ranges shown are bootstrapped 95% plausible ranges.





**Appendix Figure 13.** Household and non-household close contact numbers, close contact time, and casual contact time age-mixing matrices for Western Cape. Numbers shown in graphs a and b are the rate of contact between each individual in the population per day, expressed as rates  $\times 10^5$ . Numbers shown in graphs c-f are the mean minutes of contact between each individual in the population per day, expressed as rates  $\times 10^2$ . Contact rates between child 'respondents' and contacts in each age group were estimated from data on contact between adult respondents and child contacts. Ranges shown are bootstrapped 95% plausible ranges.