Temporal variation of human encounters and the number of locations in which they occur: A longitudinal study of Hong Kong residents.

Kin On Kwok[a,b,c,d&] , Ben Cowling[d] , Vivian Wei[a,d] , Steven Riley[e] , Jonathan M. Read[f,g&]

[a]The Jockey Club School of Public Health and Primary Care, The Chinese University of Hong Kong, Hong Kong Special Administrative Region, China.

[b]Stanley Ho Centre for Emerging Infectious Diseases, The Chinese University of Hong Kong, Shatin, Hong Kong, Hong Kong Special Administrative Region, China.

[c]Shenzhen Research Institute of the Chinese University of Hong Kong, Shenzhen, China

[d]WHO Collaborating Centre for Infectious Disease Epidemiology and Control, School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong Special Administrative Region, China.

[e]MRC Centre for Outbreak Analysis and Modelling, Department for Infectious Disease Epidemiology, Imperial College London, UK.

[f]Centre for Health Informatics, Computation and Statistics, Lancaster Medical School, Faculty of Health and Medicine, Lancaster University, UK.

[g]Institute of Infection and Global Health, The Farr Institute@HeRC, University of Liverpool, UK.

[&] Correspondence to: kkokwok@cuhk.edu.hk & jonathan.read@lancs.ac.uk

Abstract

Patterns of social contact between individuals are important for the transmission of many pathogens and shaping patterns of immunity at the population scale. To refine our understanding of how human social behaviour may change over time, we conducted a longitudinal study of Hong Kong residents. We recorded the social contact patterns for 1,450 individuals, up to four times each between May 2012 and September 2013. We found individuals made contact with an average of 12.5 people within 2.9 geographical locations, and spent an average estimated total duration of 9.1 hours in contact with others during a day. Distributions of the number of contacts and locations in which contacts were made were not significantly different between study waves. Encounters were assortative by age, and the age mixing pattern was broadly consistent across study waves. Fitting regression models, we examined the association of contact rates (number of contacts, total duration of contact, number of locations) with covariates and calculated the inter- and intra-participant variation in contact rates. Participant age was significantly associated with the number of contacts made, the total duration of contact, and the number of locations in which contact occurred, with children and parental-age adults having the highest rates of contact. The number of contacts and contact duration increased with the number of contact locations. Intra-individual variation in contact rate was consistently greater than inter-individual variation. Despite substantial individual-level variation, remarkable consistency was observed in contact mixing at the population scale. This suggests that aggregate measures of mixing behaviour derived from cross-sectional information may be appropriate for population-scale modelling purposes, and that if more detailed models of social interactions are required for improved public health modelling, further studies are needed to understand the social processes driving intra-individual variation.

Introduction

The transmission of acute respiratory infections is thought to be driven by multiple factors, including the rate of social interactions and the duration of exposure [1]. In general, individuals who have high connectivity are considered to be at elevated risk of infection and of passing infection on [2], and control interventions which target those individuals are often efficient, particularly for sexually transmitted infections. It is an open question as to whether such an approach is feasible for respiratory infections, and the link between social connectivity and infection risk for respiratory infection has only recently received research attention [3, 4].

Representative studies which quantify patterns of social encounters are few, and are typically limited to the characterisation of social mixing behaviour of individuals over a single day [3]. A few smaller studies have measured encounter patterns of individuals over multiple days, but are generally limited to two day samples, and have focussed on quantifying differences between school-term and holidays for schoolchildren [5], and contrasting days of wellness and illness [6]. Longer studies of encounter patterns have so far been conducted in small and potentially non-representative sample [7-9]. There is, therefore, a need to understand how stable or consistent the mixing behaviour of individuals is over longer periods of time, both for determining the reliability of information gained from single day studies, but also for the ability to identify and target individuals at high risk of infection. The most appropriate measure of contact rate is also unclear. Both the number of different individuals encountered and the time spent with them are both important for transmission, but it is unclear how they may combine with exposure to infectious individuals to generate infection risk. Consequently, several studies of contact mixing patterns report both the total number of contacts and estimate the total contact duration [10, 11].

Between-individual variation in the rate at which contact occurs is known to have important implications for the transmission of infectious diseases and its control [12]. Daily differences in behaviour of an individual can also impact transmission, particularly if triggered by illness [13, 14]. For many acute infectious diseases that are spread through close contact, infectious individuals can pose a transmission risk for several consecutive days until the infection is cleared or treated. This may be particularly important for influenza, where individuals may be infectious prior to symptoms developing [15]. The set of people such individuals may encounter during this infectious period defines their effective neighbourhood of contacts – the totality of people they could potentially infect [16]. In other words, the speed and extent to which infection can transmit may be determined by how quickly contacts are made and how the number of people encountered may accumulate during the infectious period. The number of different people encountered by an individual may asymptote as the number of days considered increases [7]. This saturating relationship may reduce the final variation between individuals' effective neighbourhood size, such that variation in the number of secondary infections arising may not be as great as estimated by information from a single day, particularly for infections with long (multiple day) infectious periods. Currently, there is little evidence as to how individuals' contact rates may change over time [3, 7, 17]. Understanding how effective neighbourhood size may vary in different populations and for different infections has import implications for public health control, including the effort that should be invested in contact tracing during outbreaks.

Hong Kong is a densely populated city of more than seven million residents; it is an important city for international travel, with strong regional and international communication links. This connectivity is reflected in its significance for infectious diseases: SARS emerged in the region and spread to the rest of the world through Hong Kong [18]. Annual seasonal influenza is also thought to originate in the region [19]. A previous survey of social mixing behaviour was conducted in Hong Kong to examine how social connectivity related to incidence of influenza infection during the 2009 pandemic influenza [20, 21]. Here, we present an extension of that work: a longitudinal study of the social mixing behaviour of Hong Kong residents, where participants reported on the social encounters they made on up to four different days over 17 months. Using information collected by the study, we explore the patterns and variation in three key contact rates – the number of people encountered (number of contacts), the total duration of contact events, and the number of different locations in which contacts are reported.

Methods

Study overview

We followed an open cohort of individuals belonging to recruited households, over 17 months between May 2012 and September 2013. Four waves of telephone interviews were arranged to start in May 2012, November 2012, March 2013 and July 2013 respectively, with the duration of each recruitment period lasting between three and six months. The timing of study waves was as follows: wave one (R1) ran from May 2012 to October 2012; wave two (R2) from November 2012 to March 2013; wave three (R3) from March 2013 to May 2013; wave four (R4) from July 2013 to Sep 2013 (Figure 1). Questionnaires (contact diaries) – soliciting information on social encounters made during a randomly assigned day – were administered to participants in each wave via a telephone interview. Contact diary information was collected from each participant for up to four different days (one in each wave of the study). Contact information recorded the number of distinct individuals encountered, the duration of contact events with each, and the number of distinct locations in which contact occurred.

Recruitment

Households were the main recruitment unit for this study. In the early stage of the study (May 2012), a telephone recruitment company was commissioned to recruit all study households. We aimed to recruit approximately 1,000 households. Households participating in an existing cohort study [22] were invited to participate in this study; additional households were also recruited by random dialling digit using the sampling framework used to recruit households into the existing cohort. [22]. Both recruitment arms solicited participating households from the Hong Kong population, and all households were initially identified and approached via random-digit dialling and an initial telephone call to a fixed-line number. All individuals who typically slept in the household for at least 5 nights per week were eligible to enter the study; domestic helpers were ineligible for study participation due to concerns regarding coercion. The minimum age for participation was two years old; there was no upper age limit; all eligible members of the households were invited to have four telephone interviews. Additional households were recruited as required during each study wave to balance losses to follow-up.

Reporting contacts

Participating households received a study booklet at the start of their participation describing the purpose of the study, their involvement, the definition of contact and examples of the types of contact the study will ask them to report. Contact was defined as a social encounter with an individual which included a face-to-face conversation or touch (such as handshake, a kiss, games and sports or similar events involving body touch). For each study wave, participating households were assigned a date for which their contact behaviour was to be reported (hereon referred to as the '*reporting day'*). They would be interviewed about this reporting day within four days after the reporting day (referred to as the '*interview date*'). All individuals within a recruited household were assigned the same interview date and reporting day within each wave. The reporting day was allocated sequentially within the study wave period. The household was contacted and informed of their reporting day and interview date, with both dates being reallocated later in the wave and the process repeated if the participants communicated that they were unavailable for interview on the first interview date. Following the reporting day, households were contacted by the study team on the interview date, and the team administered a questionnaire (also called a contact dairy) on each eligible participant within that household to collect recalled information on their contact behaviour during the reporting day.

Participants were asked to recall all contact events – defined as encounters with distinct individual or group of individuals in a particular geographical location – during a reporting day [10]. The number of individuals associated with a contact event could be reported and recorded as either single individuals, or as groups of individuals sharing the same attributes within the same contact event [10, 11]. For a participant's reporting day, interviewers recorded all contact events reported by the participant, a name or description of the contact or group associated with that event, and a name or description of the geographical location in which it occurred, to distinguish between different locations during the interview. Additionally, for each contact event they also recorded the number of individuals within a group contact, the duration, age, social setting (home, school, work, other), whether the encounter included touch, and the typical frequency the participant would encounter that contact. The number of unique people with whom a participant reported contact during an interview (hereon referred to as number of contacts) was defined as the number of unique contact descriptors associated with each contact event multiplied by the number of individuals represented by the contact event. Contacts descriptors were anonymised, and did not identify people in such a way as to identify a repeat encounter with a contact by the same participant (across study waves) or encounters with the same person by two or more participants. Additional information on the recording of contacts and locations by the study is provided in the Supplementary Material (Appendix A).

Interviewers reviewed the contact event information and confirmed the information with the participant where multiple contacts had the same or similar names or descriptions. In turn, eligible participants in the household were interviewed. The above procedure was repeated in each wave of recruitment. Participants who agreed to complete the questionnaire were compensated with HKD20 of supermarket vouchers for each interview in which they participated. Individuals were permitted to participate in subsequent waves even if they missed one or two waves.

Age mixing matrices

To describe the pattern of social mixing and quantify the tendency of people to mix with others of similar ages or different ages over time, we calculated age based mixing matrices of participants in four waves of recruitment with four groups of participant ages (5-19,20-39,40-64,65+) and five groups of contact ages (0-5,6-19,20-39,40-64,65+) based on the ratio of the measured probability of a contact between individuals under an assumption of proportionate mixing [10]. We excluded information from participants in the 2-4 age group due to small sample sizes. Proportionate mixing was calculated using the age distribution from the 2011 Hong Kong census [23]. Ratio values above one in the matrix indicate more contact than expected at random between the pair of age groups, and values below one indicate less contact than expected. Confidence intervals were calculated by 1,000 bootstrap resampling of participants.

Estimation of total contact duration

While the number of individuals a participant may encounter can be a useful measure of their social connectivity, from the perspective of infection by respiratory pathogens, the duration for which they may be exposed to pathogens via social interactions may be just as important. Following established methodology [24, 25], we estimated the total duration each participant was in contact with other people during a reporting day. Firstly, we fitted an exponential model to the observed distribution of categorised durations recorded for all contacts using an adaptation of the expectation–maximization algorithm. Secondly, drawing randomly from this model, we assigned durations (minutes) to each contact event reported. Thirdly, we repeated this process 200 times, to permit estimation of uncertainty in derived duration metrics. Finally, the total contact duration was found by summing the estimated contact durations for each contact event, for each participant for each day they reported. We assumed interaction with groups (more than one person) to contribute towards the total contact duration as would a contact event with a single individual.

Statistical analysis

To derive overall averages of number of contacts, duration of contact and number of locations, we calculated the mean of participants' means, to account for repeated observations per participant. We explored the variation of the accumulation of contacts over multiple days using only participants who reported contact information from all four reporting days.

We applied multivariate mixed-effect regression models to the data using total number of contacts, total duration of contact events and number of locations in which contacts were encountered as response variables. Specifically, $\log(1 + K_{ij})$, $\log(1 + D_{ij})$ and $\log(1 + L_{ij})$ are defined as the response variables with Gaussian distributions, where K_{ij} is the total number of contacts reported by participant i during survey wave j, D_{ij} and L_{ij} are the equivalent variables for total duration of contact events and number of locations. Model fitting was performed using information from participants with two or more observations and implemented within the R statistical language [26] using the *gamm4* package [27]. Random effects were modelled as participant-specific intercepts. Explanatory variables included age at interview date and sex, study wave (R1 to R4, to test for temporal effects), and the day of reporting (to understand the effect of different days of the week). For models with number of contacts and total duration as response variables, we also included the number of contact locations reported (categorized as 0, 1, 2, 3, 4, 5, and 6 or more) as an additional explanatory variable, to understand the contribution of multiple locations to contact rates. We fitted penalized thin-plate splines to explore the potential for nonlinear relationships of the explanatory variables with age at interview date in decimal years (*i.e.*, measuring age as number of months). Percentage contributions for each of the covariates were calculated by predicting the relevant contact rate as a percentage of the predicted modelled rate for a comparator set of covariate values; we used a 50-year-old male, reporting contacts on a Monday in the first study wave, with a household size of one and with a single contact location as the comparator. Additional supporting regression models were fitted for alternative response variables and exploratory variables: these models are described within the Supplementary Material (Appendices F and H).

Results

Sample size and demography

Overall, 1450 individuals from 857 households were recruited, of whom 401 took part in all four waves of recruitment, 402 took part in exactly three waves, and 327 and 320 individuals took part in exactly two and one wave respectively. Across 4 study waves, 3784 interviews were conducted, 98.5% of which were successfully made within four days of the reporting day. 30% of the participants taking part in the current study wave did not participate in the subsequent wave: 321 subjects out of 1066 participating in wave 1 did not take part in wave 2; 320 out of 995 subjects in wave 2 did not participate in wave 3; 262 out of 887 participating in wave 3 did not participate in wave 4. Recruitment of additional participants and repeated follow-up of previously participating individuals helped to maintain a large number of subjects across four waves of recruitment (Figure 1).

Twenty-six participants did not provide complete personal demographic information (such as age) or contact information: these subjects were excluded from all analyses requiring the missing information. We found no difference in the age distribution and sex of participants between study waves, though there was a difference between study waves in the days of the week for which contacts were reported (Table 1). There was no statistical difference between the distribution of participants in terms of age or sex across the four waves of recruitment, though there was difference between waves in the distribution of weekdays recorded by participants (Table 1). Children were underrepresented in our sample, while adults and females were overrepresented (Figure S1).

Distribution of contact rates and number of locations where contact occurred

We found a remarkable consistency in the overall distribution of number of contacts reported by participants between waves (Figure 2A), with each wave having comparable mean values (Table S2) and showing a similar long-tailed degree distribution of contacts (Figure 2). The pattern of this distribution, particularly the long right-hand tail, was similar to the distribution observed in similar studies in China [10] and the UK [11, 28] – studies which also were designed to enable participants to report large numbers of contacts easily by reporting groups of similar contacts. We also found distribution consistency between waves for both total contact duration and number of locations (Figure 2B, 2C). Chi-squared tests showed no significant difference in the distributions between waves of the number of contacts or the number of locations reported; however, distributions of duration were different between waves ($p<0.001$). Stratified by study wave, the number of contacts made by age groups of participants also showed a similar pattern (Figure S3). Across all four waves of study, the mean average daily number of contacts reported was 12.5, recorded in an average of 2.9 different locations, while the mean duration of contact events was 9.1 hours per day.

While the aggregate distribution of number of contacts was very similar between waves, we found considerable variation at the individual level (Table S3 and Appendix G). The distribution of the difference between number of contacts made by a participant in any two waves were similar (Figure S2A), and there was a slight negative correlation between the number of contacts reported in any two waves (ranging from -0.034 to -0.005, but not significant), though a positive correction between waves 3 and 4 (Table S3). Among all possible pairs of wave comparison, only duration of contacts between wave 2 and 3 was found to be significantly correlated though the strength of the correlation was weakly positively (Table S3). Individual-level variation between waves was also observed for total contact duration (Figure S2B). For the number of locations in which contact occurred, again there was variation at the individual level (Table S3,

Figure S2C). We found a weak positive correlation between individual participant's coefficients of variation for number of contacts, contact duration and number of locations (Figure S12).

Patterns of mixing between age groups

The manner in which age groups interact with their own and other age groups is important for the spread of infection within a population [29]. We found broadly similar patterns of mixing between age categories across the four study waves (Figure 3; Table S4; Figure S3; Figure S4), though there are some differences which may be important when considering the potential spread of infections. All age groups, except 20-39 and 40-64 groups in wave 4, were significantly more likely to have a greater number of contacts with a member of their own age group than would be expected if mixing were at random across all four waves: this is indicative of age-assortative mixing. The strongest assortative mixing rates were made by younger (5-19 years old) and older (65+) participants: these individuals were respectively at least 3.4 and 1.9 times as likely to have contact with individuals of their own age than would be expected by proportionate mixing, in study waves 1 to 3 (Figure 3). In comparison, wave 4 showed reduced assortative mixing of the younger age group (5-19 years old). This may be due to more sampled days within this wave coinciding with the summer school holidays than for other waves. This explanation is supported by an observed reduction in the average number of contacts made in school by this age group in wave 4 (Table S5). From an infectious disease perspective, wave-to-wave differences in assortative mixing do translate into differences in epidemic growth rates, with wave 2 having the fastest growth (Figure S5, Appendix E). Similar average aggregate age-mixing patterns were observed for skin-on-skin touch contacts (Table S6), which may be a more appropriate representation of a transmission opportunity for particular diseases [7, 30].

Association of contact rates with demographic variables, study wave and weekday

To assess the variation in contact behaviour at the individual-level, while adjusting for factors thought *a priori* to be associated with contact rate, we fitted mixed effect regression models to the contact metrics. We modelled the effect of participant age and sex, day of the week, number of locations in which contact was reported (if included), and study wave on the total number of contacts reported by participants, estimated total contact duration and the number of locations visited where contact occurred as independently fitted models. All models accounted for repeated observations from participants.

We found a significant nonlinear association between the number of contacts reported and age of participant, with the greatest number of contacts associated with 10-20 year olds and 40-50 year olds, and a sharp decline in contact rate above the age of 60 (Figure 4A). We found no significant association of number of contacts and the sex of participants (Table S7). A greater number of contacts were associated with midweek days (Monday through Thursday than with weekend days (Figure 4B, Table S7). The number of locations in which contacts were reported was associated with an increasing number of contacts (Figure 4B, Table S7). Study wave 2 (R2) was associated with a significantly greater number of contacts than the other waves (Table S7).

We repeated the model fitting with number of contacts stratified by the social setting in which they were made (home, school or work, other) as independent models, to investigate the association between covariates and contact rates in different settings. We found the number of home contacts to be greatest in children and 40-45 year olds, and to increase with increasing household size (Table S7, Figure S7). We found no association of home contacts with week day or study wave. The greatest number of school or work contacts were associated with school- and working-age individuals, and contact number was greater for males than females and midweek days than weekends (Table S7, Figure S7. We found no association of number school/work contacts with contact locations greater than 1 or study wave. The number of contact made in other settings, which included leisure and shopping activities, was associated with age: the number of contacts in these settings decreased with age up to 30 years, and then increased with increasing age (Table S7, Figure S7). Females made more contacts in these settings than males, and more of these contacts were made at weekends; there was no effect of study wave. There was a very large effect of number of contact locations with these types of contact, suggesting that this type of contact may be responsible for the relationship with total number of contacts as the response variable. The number and proportion of contacts made in different social settings varied by study wave (Table S5).

Total contact duration was also significantly associated non-linearly with participant age, with a general reduction in duration observed with increasing age (Figure 4c). There was no significant effect of the sex of participants, but there was a significant effect of day of the week, with contact duration being longer on weekend days than Wednesdays Thursdays, and Fridays (Figure 4D, Table S7). Contact duration increased with the number of locations reported (Figure 4D, Table S7). Study wave 2 was associated with shorter contact duration than the other waves (Figure 4D, Table S6). We found our model findings to be insensitive to the uncertainty in the estimation of contact duration (Figure S8).

We found no significant association of number of locations visited with participant sex, study wave, but there was a significant non-linear association with participant age, where 45 to 50-year olds were associated with the greatest number of locations visited, and more locations were reported on Fridays than other days of the week.

The questionnaire survey also asked participants whether the day for which they were reporting contact events could be considered as a ‘typical’ day or not. 73.4% of observations were reported to be typical days, 26.3% were reported as non-typical days, and 0.03% (n=13) of interviews participants could not be sure (they responded “Don’t know”). Restricting our regression analysis to only observations categorised as ‘typical’ by participants gave similar associations with covariates as reported above (Figure S6).

Variation in contact rates

The longitudinal nature of our study and the random effect structure of our regression models allowed us to consider the proportion of variance in contact rate attributable to intra- and inter-individual variation (Table S8). When we considered the number of people encountered (the number of contacts), we found between individual variation (33.7%) to be less than the variation observed within individuals (66.3%). A similar distribution of variance was observed for total contact duration (28.6% between and 71.4% within individuals) and number of locations (25.9% between and 74.1 within individuals). When we limited our study observations to only those where participants reported ‘typical’ days, we found between variation to increase slightly, but still less than within individuals (Table S8). Similar patterns were found for models of number of contacts in different social settings (Table S8). Finally, to further explore wave-to-wave variation in individuals’ contact rates, we considered how likely individuals were to report a consistent number of contacts across study waves, by calculating the percentage of participants remaining in the same contact quantile as the number of quantiles increased (Figure S10). Only between 30% to 40% of participants had consistent contact rates in the range of quantiles we explored, though we consistently found a greater proportion of participants’ observations remained in their quantile than for a null model which excluded within-participant dependencies.

Variation between individuals and neighbourhood saturation

As the number of observations per participant increases, reflecting a corresponding increase in infectious period, we may expect the variation in cumulative contacts between individuals to decrease. Subsequently, we hypothesise that infections with different infectious periods may inhabit potentially different dynamics networks of transmission opportunities, even in the same host population. We explored changes in the variation of cumulative contact rates over multiple study waves (Figure S11). While nearly all measures of between-participant variation decreased with increasing number of study waves, in many cases we found the variation to be greater than that expected by a null model which excludes within-participant dependencies. Thus, there was evidence that contact rates saturate to some extent (particularly for contact duration and number of locations), though within-individual variation is still sizable. We found evidence of weak positive correlation between an individual participants' coefficients of variation for number of contacts, contact duration and number of locations (Figure S12).

Individual and group contacts

To understand how our finding related to how participants reported contacts, we considered the number of contacts reported as individuals and as groups independently (Appendix H). Participants tended to report less than one group per diary on average, with an average group size of between 9 and 11 people (Table S9). The distribution of contact number reported as individuals was consistent across waves, though there was more variability between waves for contacts reported as groups (Figure S13). We also fitted independent regression models following the method of that described for the total number of contacts, with two different response variables: the number of contacts reported as individuals or groups, respectively (Figure S14). These models present similar results to the combined contact model, though associations deviate for several of the variables, most notably the relationship with number of locations and study wave. Whether a contact is reported as an individual or part of a group is the choice of the participant, and participants tended to use groups for reporting large number of contacts. These deviations from the principle model likely reflect differences between participants, and encounters between waves, and may also reflect participants tending to reporting groups more often for as they grew accustomed to participating in the study.

Discussion

Social encounter patterns are an important driver of the spread of infectious diseases requiring close contact for transmission, particularly for respiratory viral pathogens [3]. Quantifying such behaviours enables improved modelling of epidemics for a variety of purposes, and helps identify effective interventions aimed at reducing transmissions. Here we present the results of a large longitudinal study of Hong Kong residents, a population inhabiting one of the highest density locations in the world and one which played an important role in the transmission of SARS [31].

We conducted a large cohort study where participants were asked to provide information on their social contacts and mixing behaviour at up to four different time points during two calendar years. At the aggregate level, we found remarkable consistency in the contact patterns made by participants across study waves, in terms of both the distribution of number and duration of contacts, as well as the distribution of number of distinct locations in which contacts were made and the pattern of mixing within and between different age groups. These aggregate contact patterns were similar to those observed in other studies based on European and Chinese populations [10, 11, 28, 29, 32], and a recent study of Hong Kong residents [33]. However, large representative studies of contact patterns have so far been limited to cross-sectional observations of behaviour over a single day per participant. Our findings suggest that this is an appropriate methodology when the objective of the study is to provide aggregate measures of contact patterns (*e.g*., contact rate per age group or age assortativity patterns) for modelling purposes. If the objective is to parameterise individual-based models or more finely resolved group-structured models, however, our findings suggest it may be important to incorporate individual-level variation in contact rate and social mixing behaviour.

We found a significant association of study wave on the number and duration of contact, and also found the pattern of mixing between age groups was subtly different between waves. We hypothesise that the effect of wave we observe may reflect seasonal patterns in contact behaviour, including Chinese New Year and summer school holidays. Differences in contact rate by wave may also be explained by the different 'mix' by wave of encounters made in different social settings. We found day of the week to be significantly associated with the number and total duration of contact, with weekends associated with fewer contacts but increased contact duration than weekdays.

The number of locations, the number of contacts and the time spent in contact with them appeared to be strongly linked. Individuals who visit more locations accrued a greater number of contacts, while contact duration quickly asymptotes with location number. Stratification by social settings in which encounters are made suggests that the effect of number of locations visits on contact number is driven by contacts made outside of the home, school and workplace environments. Individuals may visit many locations and make a correspondingly large number of contacts, but overall tend to spend less time per contact. This suggests an intriguing interplay between the spatial roaming of individuals and their network of social encounters in environments not easily represented by demographic and occupational derived models, presenting a complex challenge in representing social or transmission networks within geographical space, and in developing realistic models of infectious disease transmission .

We found considerable intra-individual variation in contact rates reported by individuals, even after accounting for potential confounders (day of week, number of locations in which contact occurred and study wave): intra-individual variance was greater than inter-individual variation for the number and duration of contact and the number of locations in which contact occurred. While our study suggests that the number of contacts made by individuals is variable on a day-to-day basis, we find that variation in contact rate reduces little when we consider the accumulation of contacts over multiple days. Our analysis is limited to only four observations per participant and our methodology does not permit unique contacts to be identified between waves. None-the-less, these results suggest that inter-individual variation in the number or people encountered, the time spent in contact with them, and the variety of locations they are encountered in, may not saturate as quickly as expected over longer infectious periods.

Our analysis does not explore the potential source of the intra- and inter-individual variation, and a deeper exploration of the setting and reasons for the contacts reported by participants may prove illuminating from both sociological modelling and public health perspectives.

Many respiratory pathogens of public health interest have infectious periods of longer than a single day, and the contacts made by infectious individuals during their period of infectivity will define the speed and extent of spread within the network. The re-wiring of an implicit contact network that we have measured may ensure that local saturation effects, where infectious individuals have opportunity to infect all susceptible individuals within their neighbourhood, are rare outside the household for pathogens with short infectious periods. The relationship of infectious period with the temporal dynamics and geographical patterns of social encounters we have observed is likely to drive the higher-order spread of infectious disease, and may provide important insights for public health interventions, such as contact tracing.

There are some limitations to this study. While our study is generally representative of Hong Kong households and population, we recruited very few participants under the age of 5 years old. We were also reliant on the recall of contact events by participants, and this might introduce bias in the number, duration, and location of reported contacts. A further limitation is that as our study was conducted across several waves spanning several months, we do not have contact behaviour information from participants from consecutive days. A consequence of the telephone study team not working on weekends, and the random assignment of contact reporting days to participants meant that lower numbers of contact days were recorded for Saturdays and Sundays. The bias in sampling of different days of the week is, therefore, a consequence of our study design; the principle aim through sampling was to recruit a representative sample of households and individuals therein, and representativeness for day of the week was secondary in our sampling aims. Weather conditions may have a confounding effect on the contact patterns we have observed [34] and we did not adjust for these in our analysis. Finally, due to the design of data collection, we cannot identify repeated contact made between a participant and the same individual (their contact), which limits our ability to fully identify any neighbourhood saturation effect.

In conjunction with information from other studies, our study provides important information for the parameterisation of realistic models of social encounters made in Hong Kong, with application to public health modelling. This study also provides support for the use of cross-sectional information for parameterising epidemic models which focus on describing the risk of infection for average individuals. However, our study also highlights the complexity of social encounters, particularly when considering their spatial context, and the need for improved understanding of the social processes driving population-scale mixing patterns.

Author contributions

KOK, JMR and SR designed the study. KOK and JMR analyzed data, interpreted data and drafted the manuscript. BJC, VWWI and SR edited and contributed to the manuscript.

Funding statement

This study was funded by the Research Fund for the Control of Infectious Diseases (grant number: 11100642). Health and Medical Research Fund of the Health, Welfare and Food Bureau of the Hong Kong SAR Government (grant nos. 11100642 and 13120732), the General Research Fund of the University Grants Committee (grant no. 776810), the Harvard Center for Communicable Disease Dynamics from the National Institute of General Medical Sciences (grant no. U54 GM088558), a commissioned grant from the Health and Medical Research Fund of the Health, Welfare and Food Bureau of the Hong Kong SAR Government, the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. T11-705/14N) , the Medical Research Council (UK, Project MR/J008761/1), the Wellcome Trust (UK, Project 093488/Z/10/Z; Investigator 200861/Z/16/Z; Collaborator 200187/Z/15/Z), National Institute for General Medical Sciences (US, MIDAS U01 GM110721-01), National Institute for Health Research (UK, for Health Protection Research Unit funding), Fogarty International Centre with the Science & Technology Directorate, Department of Homeland Security (USA, RAPIDD program), and Fogarty International Centre (USA, R01 TW008246-01), Economic and Social Research Council (ES/K004255/1, RES-355-25-0019) and Engineering and Physical Sciences Research Council (EP/N014499/1).

Acknowledgements

Li Ka Shing Institute of Health Sciences is acknowledged for providing technical support in the research.

Data accessibility

The dataset is available as supplementary information.

Ethics

This study has been approved by the Institutional Review Board of the University of Hong Kong/Hospital Authority Hong Kong West Cluster (Ref: UW11-367).

Legends and tables

Figure 1. Timeline of the study, showing the four waves of study participation. The

duration of study was 17 months.

Figure 2. Normalised distributions of (**A**) the number of contacts and (**B**) the total duration of contact events made, and (**C**) the number of locations at which contact events occurred for each of four waves of sampling. Waves are represented by unique colours and symbols as shown in **A**. Durations were binned into log-distributed periods prior to plotting. Inset plots show the corresponding inverse cumulative probability distributions for each wave, colour coded as for the main plots.

Figure 3. Age mixing matrices, stratified by subsequent study waves (R1, R2, R3, R4, A-D respectively). Bluer colours indicate less mixing between age groups than expected by random mixing, and yellower colours indicate more mixing. 95% confidence intervals are shown in the parenthesis, derived from 1,000 re-samples of participant contact diaries.

Figure 4. Estimates of percentage contribution in the predicted number of contacts (A and B), the total duration of contact events (C and D), and the number of locations in which contact occurred (E and F) from the regression models for different characteristics of the participants relative to the individual whose was a 50-year-old male taking part in the study in wave 1, making his contact in 1 location on Monday. A, C and F show the splines fitted to age for the two models, while B, D and E show the percentage contribution for sex, day of the week, categorized number of locations in which contact occurred (L0 to L6+) if included, and study wave (R1 to R4).

References

1. Johnstone-Robertson, S.P., et al., *Social mixing patterns within a South African township community: implications for respiratory disease transmission and control.* Am J Epidemiol, 2011. **174**(11): p. 1246-55.
2. Keeling, M.J. and P. Rohani, *Modeling infectious diseases in humans and animals*. 2008, Princeton: Princeton University Press. xi, 366 p.
3. Read, J.M., et al., *Close encounters of the infectious kind: methods to measure social mixing behaviour.* Epidemiol Infect, 2012. **140**(12): p. 2117-30.
4. Cauchemez, S., et al., *Role of social networks in shaping disease transmission during a community outbreak of 2009 H1N1 pandemic influenza.* Proc Natl Acad Sci U S A, 2011. **108**(7): p. 2825-30.
5. Eames, K.T., N.L. Tilston, and W.J. Edmunds, *The impact of school holidays on the social mixing patterns of school children.* Epidemics, 2011. **3**(2): p. 103-8.
6. Van Kerckhove, K., et al., *The impact of illness on social networks: implications for transmission and control of influenza.* Am J Epidemiol, 2013. **178**(11): p. 1655-62.
7. Read, J.M., K.T. Eames, and W.J. Edmunds, *Dynamic social networks and the implications for the spread of infectious disease.* J R Soc Interface, 2008. **5**(26): p. 1001-7.
8. Fu, Y.C., D.W. Wang, and J.H. Chuang, *Representative contact diaries for modeling the spread of infectious diseases in Taiwan.* PLoS One, 2012. **7**(10): p. e45113.

9. Beutels, P., et al., *Social mixing patterns for transmission models of close contact infections: exploring self-evaluation and diary-based data collection through a web-based interface.* Epidemiol Infect, 2006. **134**(6): p. 1158-66.
10. Read, J.M., et al., *Social mixing patterns in rural and urban areas of southern China.* Proc Biol Sci, 2014. **281**(1785): p. 20140268.
11. Danon, L., et al., *Social encounter networks: characterizing Great Britain.* Proc Biol Sci, 2013. **280**(1765): p. 20131037.
12. Lloyd-Smith, J.O., et al., *Superspreading and the effect of individual variation on disease emergence.* Nature, 2005. **438**(7066): p. 355-9.
13. Eames, K., et al., *Six challenges in measuring contact networks for use in modelling.* Epidemics, 2015. **10**: p. 72-7.
14. Bansal, S., et al., *The dynamic nature of contact networks in infectious disease epidemiology.* Journal of Biological Dynamics, 2010. **4**(5): p. 478-489.
15. Fraser, C., et al., *Factors that make an infectious disease outbreak controllable.* Proc Natl Acad Sci U S A, 2004. **101**(16): p. 6146-51.
16. Riley, S., *Large-scale spatial-transmission models of infectious disease.* Science, 2007. **316**(5829): p. 1298-301.
17. Ferrari, M.J., et al., *Network frailty and the geometry of herd immunity.* Proc Biol Sci, 2006. **273**(1602): p. 2743-8.
18. Peiris, J.S., et al., *The severe acute respiratory syndrome.* N Engl J Med, 2003. **349**(25): p. 2431-41.
19. Russell, C.A., et al., *The global circulation of seasonal influenza A (H3N2) viruses.* Science, 2008. **320**(5874): p. 340-6.
20. Kwok, K.O., et al., *Social contacts and the locations in which they occur as risk factors for influenza infection.* Proc Biol Sci, 2014. **281**(1789): p. 20140709.
21. Kucharski, A.J., et al., *The contribution of social behaviour to the transmission of influenza A in a human population.* PLoS Pathog, 2014. **10**(6): p. e1004206.
22. Riley, S., et al., *Epidemiological characteristics of 2009 (H1N1) pandemic influenza based on paired sera from a longitudinal community cohort study.* PLoS Med, 2011. **8**(6): p. e1000442.
23. Department, C.a.S., *2011 Population Census*. 2011.
24. Danon, L., et al., *Social encounter networks: Characterizing great Britain.* Proceedings of the Royal Society B: Biological Sciences, 2013. **280**(1765).
25. Read, J.M., et al., *Social mixing patterns in rural and urban areas of southern China.* Proceedings. Biological sciences / The Royal Society, 2014. **281**(1785): p. 20140268.
26. foundation, T.R. *The R Project for Statistical Computing*. 2016; Available from: https://www.r-project.org/.
27. Simon , W. *gamm4: Generalized additive mixed models using mgcv and lme4*. 2015; Available from: http://cran.r-project.org/web/packages/gamm4/index.html.
28. Danon, L., et al., *Social encounter networks: collective properties and disease transmission.* J R Soc Interface, 2012. **9**(76): p. 2826-33.
29. Mossong, J., et al., *Social contacts and mixing patterns relevant to the spread of infectious diseases.* PLoS Med, 2008. **5**(3): p. e74.
30. Melegaro, A., et al., *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): p. 143-51.
31. Donnelly, C.A., et al., *Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong.* Lancet, 2003. **361**(9371): p. 1761-6.
32. Kwok, K.O., et al., *[An international collaborative study on influenza viruses antibody titers and contact patterns of individuals in rural and urban household of Guangzhou].* Zhonghua Liu Xing Bing Xue Za Zhi, 2014. **35**(4): p. 433-6.

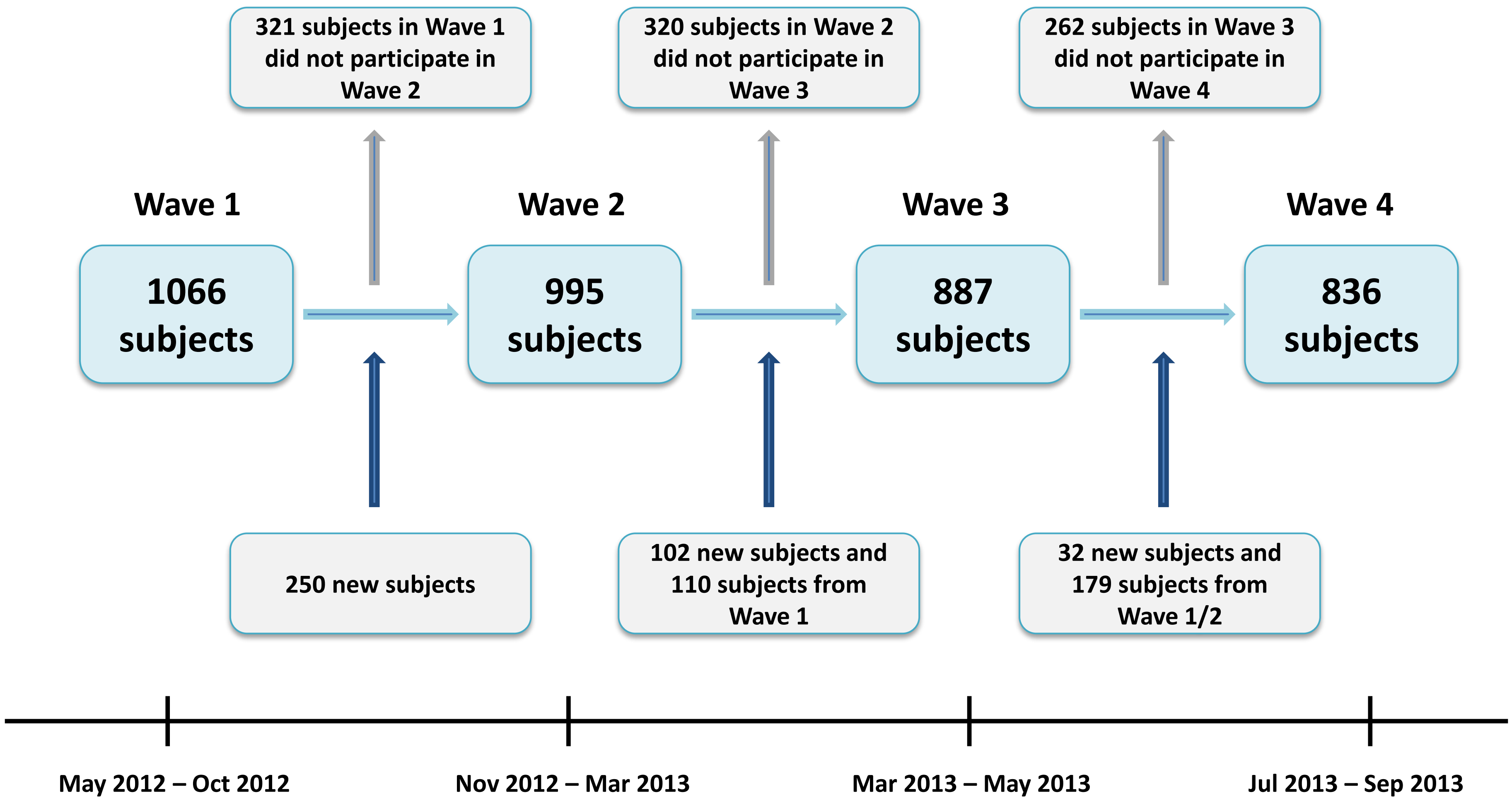
33. Leung, K., et al., *Social contact patterns relevant to the spread of respiratory infectious diseases in Hong Kong.* Sci Rep, 2017. **7**(1): p. 7974.
34. Chan, T.C., Y.C. Fu, and J.S. Hwang, *Changing social contact patterns under tropical weather conditions relevant for the spread of infectious diseases.* Epidemiol Infect, 2015. **143**(2): p. 440-51.

Table 1. Characteristics of the study subjects

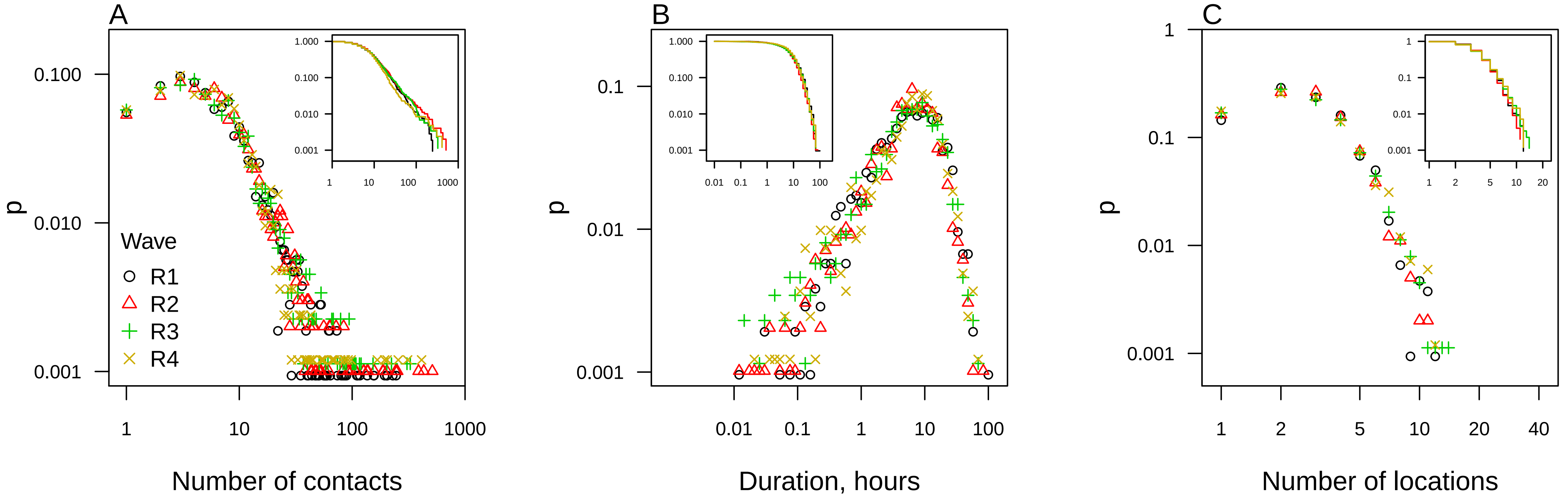
	Number of participants (%)				
	Wave 1 N=1066	Wave 2 N=995	Wave 3 N=887	Wave 4 N=836	p-value[b]
Age					
2-4	1 (0.1)	1 (0.1)	0 (0.0)	1 (0.1)	0.571
5-19	80 (7.5)	61 (6.1)	56 (6.3)	43 (5.1)	
20-39	203 (19.0)	189 (19.0)	182 (20.5)	166 (19.9)	
40-64	629 (59.0)	586 (58.9)	501 (56.5)	480 (57.4)	
65+	144 (13.5)	145 (14.6)	138 (15.6)	141 (16.9)	
Not recorded	9 (0.8)	13 (1.3)	10 (1.1)	5 (0.6)	
Sex					
Male	416 (38.0)	393 (39.5)	347 (39.1)	332 (39.7)	0.990
Female	650 (61.0)	602 (60.5)	540 (60.9)	504 (60.3)	
Not recorded	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	
Weekday					
Sunday	154 (14.4)	123 (12.4)	133 (15.0)	160 (19.1)	<0.001
Monday	196 (18.4)	116 (11.7)	195 (22.0)	132 (15.9)	
Tuesday	163 (15.3)	241 (24.2)	98 (11.0)	97 (11.6)	
Wednesday	125 (11.7)	133 (13.3)	90 (10.1)	83 (9.9)	
Thursday	140 (13.1)	131 (13.2)	119 (13.4)	97 (11.6)	
Friday	159 (14.9)	87 (8.7)	114 (13.0)	109 (13.0)	
Saturday	129 (12.1)	164 (16.4)	137 (15.4)	157 (18.8)	
Not recorded	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	
Response rate (%)[a]	74.9	69.9	62.3	58.7	

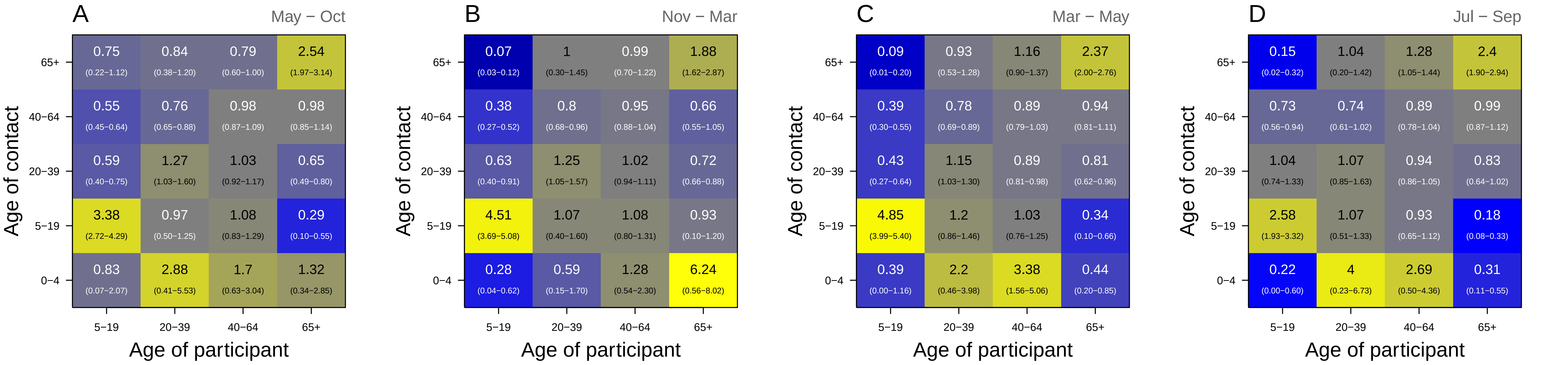
[a] Calculated based on 1424 participants who had ever participated in any waves of the recruitment with full information of age, sex and weekday

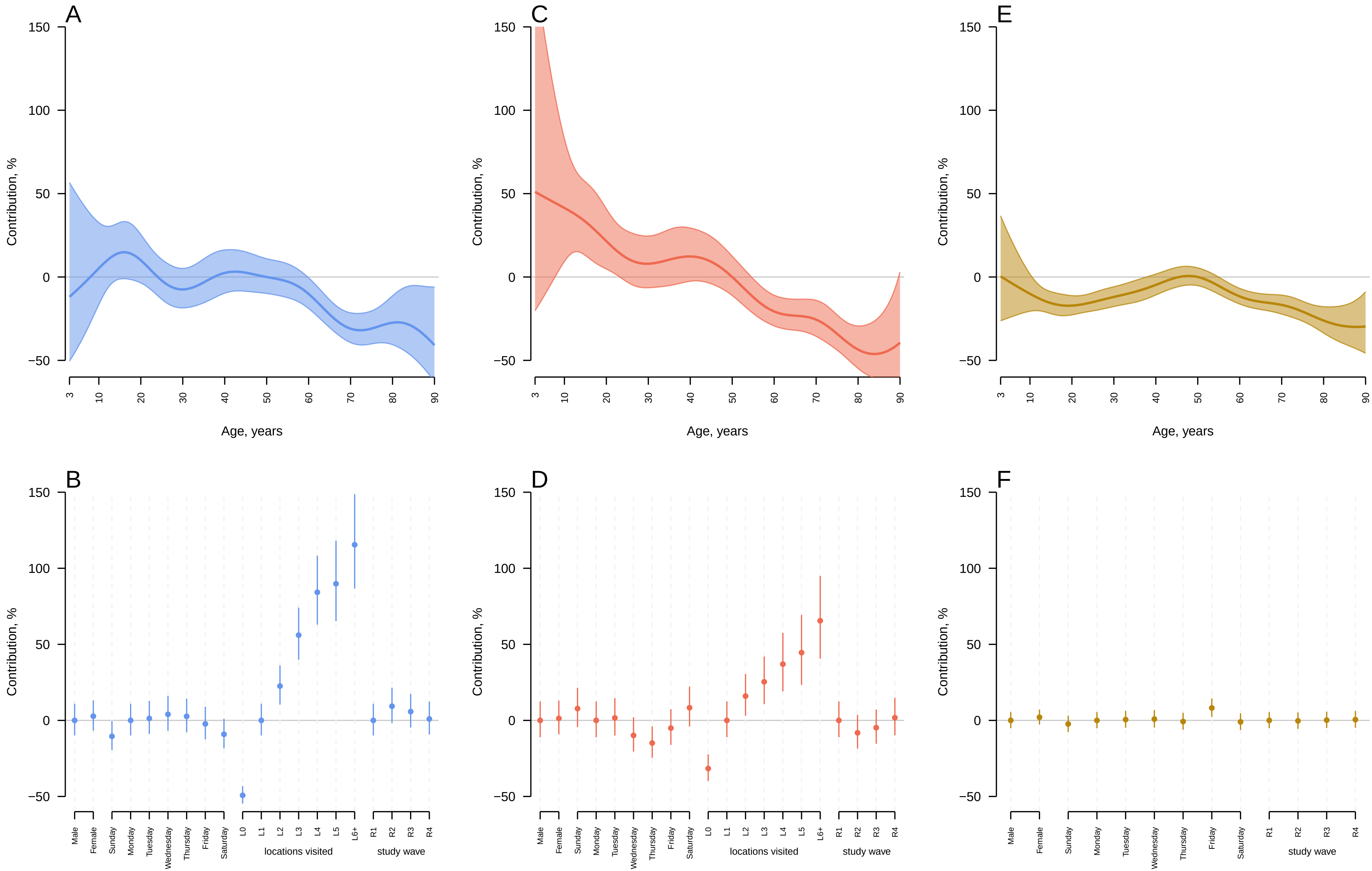
[b] Chi-square test for independence



321 subjects in Wave 1 did not participate in Wave 2
320 subjects in Wave 2 did not participate in Wave 3
262 subjects in Wave 3 did not participate in Wave 4
Wave 1
1066 subjects
Wave 2
995 subjects
Wave 3
887 subjects
Wave 4
836 subjects
250 new subjects
102 new subjects and 110 subjects from Wave 1
32 new subjects and 179 subjects from Wave 1/2
May 2012 – Oct 2012
Nov 2012 – Mar 2013
Mar 2013 – May 2013
Jul 2013 – Sep 2013


A
B
C
p
Number of contacts
Duration, hours
Number of locations
Wave
R1
R2
R3
R4
0.100
0.010
0.001
1
10
100
1000
0.1
0.01
0.01
0.1
1
10
100
1
2
5
10
20
40
1.000
0.100
0.010
0.001


A
May − Oct
B
Nov − Mar
C
Mar − May
D
Jul − Sep
Age of contact
Age of participant
65+
40−64
20−39
5−19
0−4
5−19
20−39
40−64
65+


A
B
C
D
E
F
Contribution, %
Age, years
150
100
50
0
−50
3
10
20
30
40
50
60
70
80
90
Male
Female
Sunday
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
L0
L1
L2
L3
L4
L5
L6+
locations visited
R1
R2
R3
R4
study wave

Supplementary Material

List of Tables and Figures

Appendix A. Recording of contacts and locations.

A participant's contacts and characteristics of those contacts were recorded by study researchers through interviews with the participant, and the information provided was recorded by researchers in two tables shown in Table S1. Participants were interviewed and prompted to recall the people they encountered as well as *contact events*, where an event was defined as an encounter made with one or more other people in a particular location within a discrete time period, and an encounter was defined as a face-to-face conversation between a participant and another person where they are within 1 meter of each other and/or where a participant touched someone's skin with their skin (examples provided to participants included shaking or holding hands, or a kiss). Each participant's contacts were assigned a unique name or description by the interviewer. Thus, a participant could record encounters made in the same location with possibly the same individuals but at very different times of the day (for example, meeting the same group of commuters on the way to work and on the way back from work).

Interviews proceeded as follows. First, subjects were asked to recall the different individuals or groups of individuals they encountered on the pre-assigned recording day, to populate the Person table (Table S1). Second, participants were asked to answer some basic information about each contact individual or group. Participants were asked to report: the age range of the contact(s) (0-4, 5-19, 20-39, 40-64, 65 or older); the typical frequency of encountering the contact(s) (Regular contact: 4 or more days a week, 2-3 days a week, once a week, Non-regular contact: less than once a week, met for the first time that day). For groups, participants were instructed to report the characteristics that would apply to the majority of individuals present within the group. Third, participants were asked to recall, for each individual/group encountered, the different locations in which they encountered that particular individual/group. Responses were used to populate the Contact Event Table (Table S1). Characteristics of the contact event were recorded at this time: whether the encounter involved skin-on-skin touch; the social context in which the contact event occurred (the participant's home, work or school, travel, shopping, meet or others), an estimate of the duration of the contact event (<10 minutes, 10-29 minutes, 30-59 minutes, 1 hour, 1 hour to 2 hours, 2 hours to 4 hours and 4 hours or more).

Table S1. Contact diary recording tables.

Table name	Items recorded
Person Table	• A short description of each individual or group of individuals encountered by the participant during their recording day (for interviewing purposes) • Number of people encountered (if reporting a group) • Age category of individual/group • Typical frequency with which the participant encounters the individual/group
Contact Event Table	• Reference of individual/group encountered during contact event (link to entry in Person Table) • Start time at which the participant was at that location (for reference purposes during interview) • A short description of the location (for reference purposes during interview) • Setting or setting of encounter event (home, work, school, shopping, restaurant, travel, leisure, other) • Duration of the contact event • Whether the contact event involved skin-on-skin touch

Appendix B. Study sample demography.

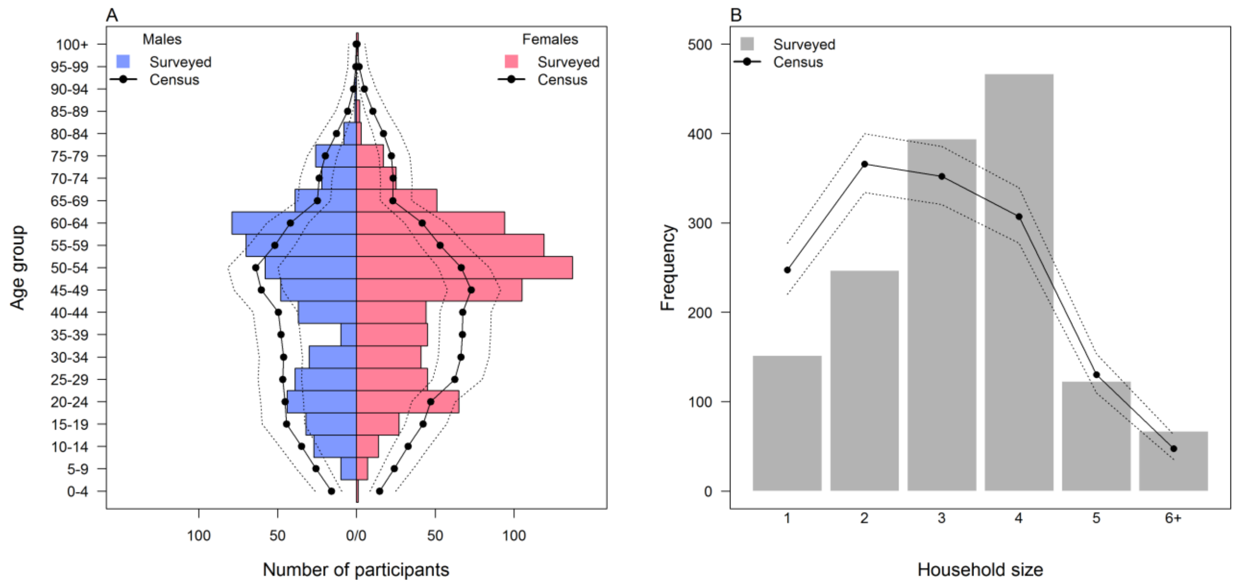


Figure S1. Study population demography. (A) Age and sex of participants at recruitment. (B) Household size of participant (including participant). Black dots denote expected distributions based on distributions derived from Census 2011 data provided by Hong Kong Census and Statistics Department. Dotted lines denote 95% confidence bound given final recruited participant across all waves.

Appendix C. Comparison of contact patterns between study waves.

Table S2. Mean average of contact rates stratified by study wave. Confidence intervals were derived from 1,000 bootstrap resamples of the data.

Contact metric	Study wave	Mean (95% confidence interval)
Number of contacts	R1	11.91 (10.75-13.11)
	R2	14.10 (12.21-16.08)
	R3	12.67 (11.23-14.32)
	R4	11.25 (9.71-13.08)
Total contact duration	R1	9.48 (8.88-10.08)
	R2	8.34 (7.80-8.84)
	R3	9.00 (8.42-9.62)
	R4	9.48 (8.82-10.12)
Number of locations	R1	2.98 (2.88-3.10)
	R2	2.96 (2.87-3.06)
	R3	3.01 (2.89-3.14)
	R4	3.00 (2.88-3.12)

Table S3. Comparison of individuals' contact metrics between different study waves. Confidence intervals were derived from 1,000 bootstrap resamples of the data.

Contact metric	**Paired study waves**	**Mean difference (95% confidence interval)**	**Spearman's correlation coefficient**
Number of contacts	R1-R2	-2.807 (-5.544, -0.208)	-0.022, p=0.556
	R1-R3	-0.742 (-3.185, 1.858)	-0.005, p=0.906
	R1-R4	0.325 (-2.298, 2.662)	-0.028, p=0.494
	R2-R3	1.693 (-1.483, 5.315)	-0.034, p=0.379
	R2-R4	2.187 (-0.706, 5.346)	-0.027, p=0.492
	R3-R4	1.800 (-1.389, 4.797)	0.060, p=0.132
Total contact duration	R1-R2	1.603 (0.570, 2.553)	-0.019, p=0.602
	R1-R3	0.460 (-0.783, 1.629)	-0.084, p=0.040
	R1-R4	0.235 (-0.985, 1.488)	0.014, p=0.727
	R2-R3	-0.996 (-1.926, -0.087)	0.080, p=0.042
	R2-R4	-1.069 (-1.977, -0.139)	0.032, p=0.422
	R3-R4	-0.227 (-1.280, 0.829)	-0.028, p=0.490
Number of locations	R1-R2	-0.048 (-0.209, 0.136)	0.035, p=0.336
	R1-R3	-0.073 (-0.272, 0.135)	-0.011, p=0.785
	R1-R4	-0.037 (-0.262, 0.168)	-0.026, p=0.525
	R2-R3	-0.062 (-0.243, 0.119)	-0.020, p=0.600
	R2-R4	-0.032 (-0.220, 0.178)	-0.008, p=0.829
	R3-R4	0.005 (-0.197, 0.216)	-0.002, p=0.966

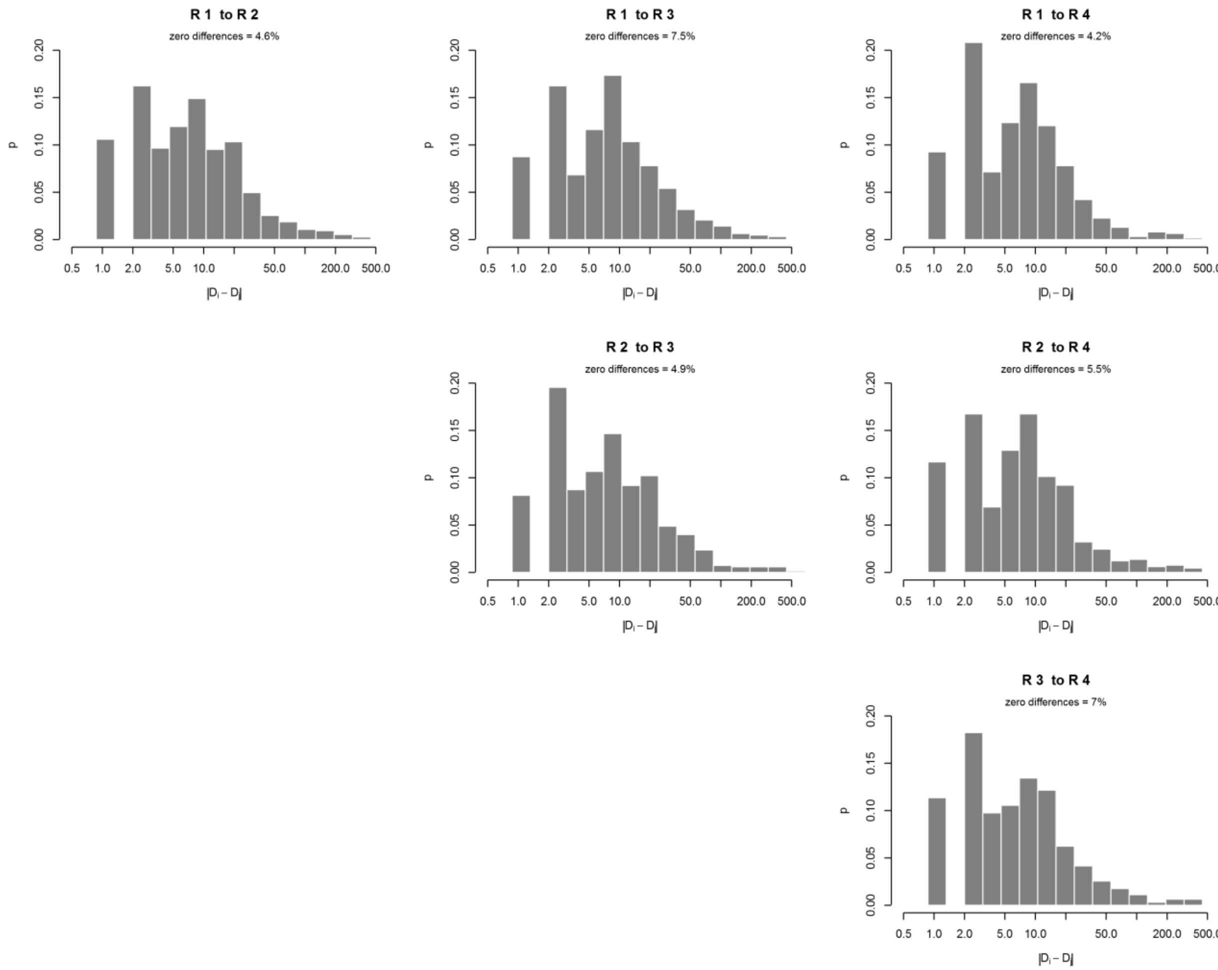


Figure S2a. Individual-level distribution of the different in number of contacts reported by participant between pairs of study waves. Note, the absolute difference is binned by logarithmically spaced breaks.

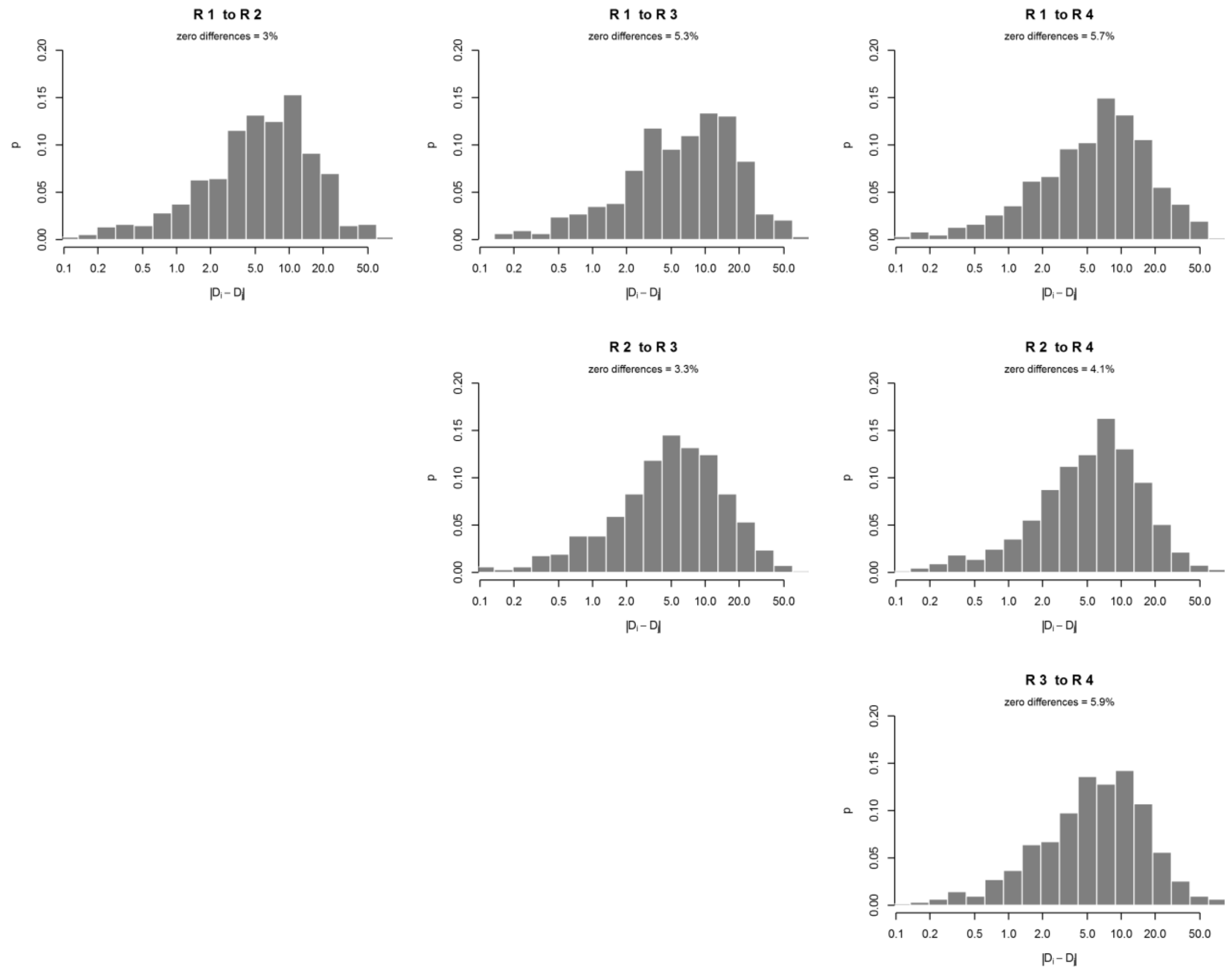


Figure S2b. Individual-level distribution of the different in duration of contacts reported by participant between pairs of study waves. Note, the absolute difference is binned by logarithmically spaced breaks.

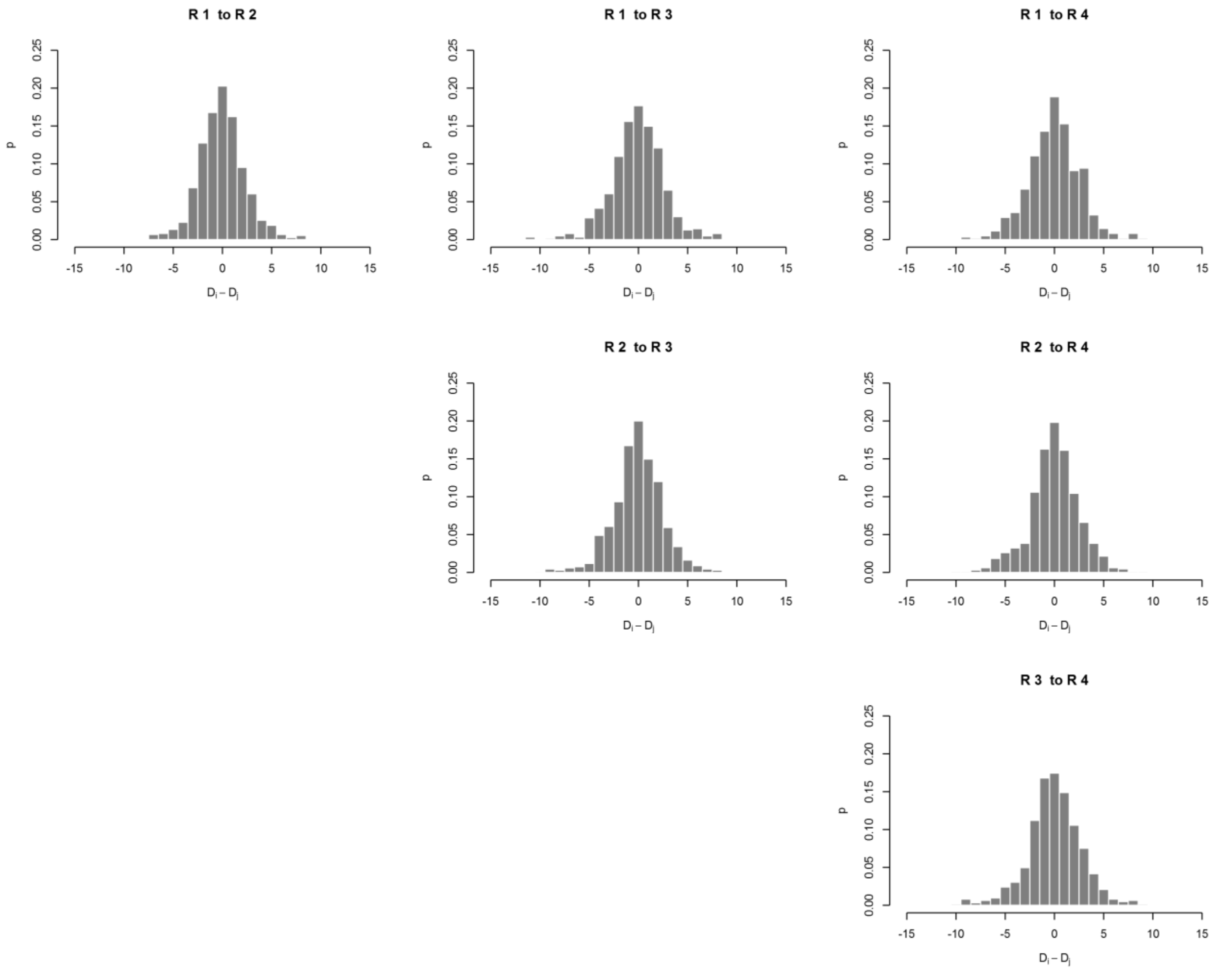


Figure S2c. Individual-level distribution of the different in number of locations reported by participant between pairs of study waves.

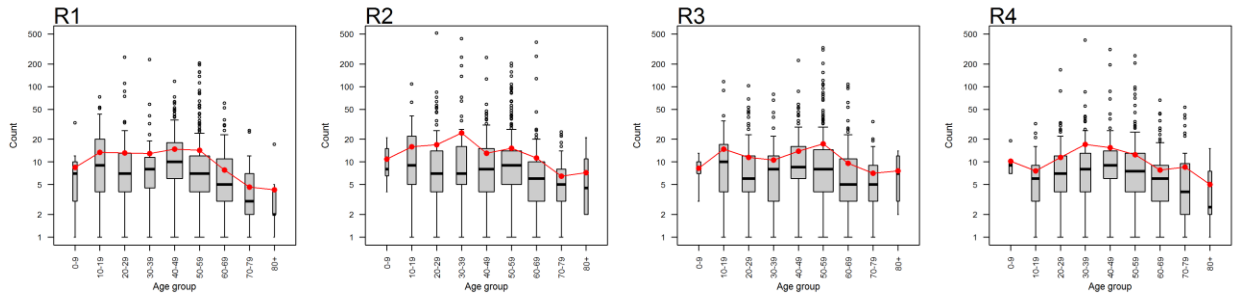


Figure S3. Boxplots of total number of contacts made by age groups of participants stratified by study wave. Box widths are indicative of the sample size. Red dots denote the mean number of contacts for each group. Note, the y-axis is plotted as a log-scale, and so participants making zero contacts are not represented. The number of zero contact observations in each study wave are as follows: R1, 21; R2 11; R3, 16; R4 21.

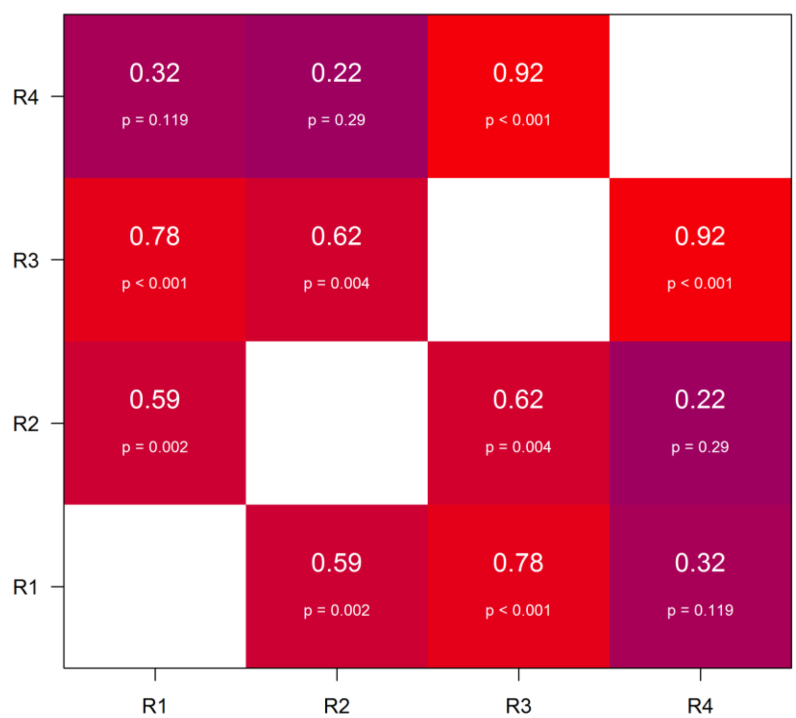


Figure S4. Correlations between age-mixing matrices from each study wave. This matrix shows the spearman correlation coefficients (and associated p-values) between pairs of wave-specific age-mixing matrices shown in Figure 4. R1 to R4 are each of the four study waves.

Appendix D. Average number of total and touch contacts stratified by participant-contact age groups, setting, and study wave

Table S4. Mean number of contacts, stratified by study wave, and age group of participant and contact.

		Age of participant[1]				
Study wave		**0-4**	**5-19**	**20-39**	**40-64**	**65+**
R1	*n*	*1*	*80*	*203*	*629*	*144*
	0-4	1.00	0.31	1.45	0.71	0.25
	5-19	1.00	8.53	3.27	3.05	0.31
Age of contact	**20-39**	4.00	3.15	8.99	6.12	1.48
	40-64	4.00	3.86	7.10	7.71	2.99
	65+	0.00	1.77	2.66	2.10	2.53
R2	*n*	*1*	*61*	*189*	*586*	*145*
	0-4	20.00	0.10	0.48	0.58	2.75
	5-19	0.00	10.69	5.83	3.30	2.73
Age of contact	**20-39**	0.00	2.54	14.53	6.47	4.56
	40-64	1.00	2.43	12.18	8.04	5.48
	65+	0.00	0.15	5.14	2.83	5.25
R3	*n*	*0*	*56*	*182*	*501*	*138*
	0-4	-	0.12	0.89	1.93	0.12
	5-19	-	10.30	3.24	3.91	0.49
Age of contact	**20-39**		1.89	6.52	7.11	2.46
	40-64	-	2.30	5.86	9.49	3.82
	65+	-	0.16	2.39	4.16	3.20
R4	*n*	*1*	*43*	*166*	*480*	*141*
	0-4	0.00	0.05	2.55	1.38	0.07
	5-19	6.00	3.72	4.55	3.20	0.23
Age of contact	**20-39**	0.00	3.16	9.60	6.82	2.13
	40-64	6.00	2.93	8.80	8.52	3.36
	65+	0.00	0.21	4.20	4.13	2.77

1. No adjustment for the age distribution of the population or participants has been made to mean contact numbers.
Diagonal (same age mixing) is marked for convenience

Table S5. Mean number of contacts, stratified by study wave, age group of participant, and social setting.

Study wave		All participants	Age of participant[1] 0-4	5-19	20-39	40-64	65+
R1	*n*	*1066*	*1*	*80*	*203*	*629*	*144*
Setting	**home**	2.56	4.00	3.77	2.54	2.61	1.74
	school	0.60	6.00	5.05	0.35	0.19	0.01
	work	4.82	0.00	1.75	7.48	5.42	0.33
	other	4.54	1.00	3.21	3.05	5.35	4.01
R2	*n*	*995*	*1*	*61*	*189*	*586*	*145*
Setting	**home**	2.53	1.00	3.74	2.69	2.41	2.28
	school	0.87	20.00	8.33	0.96	0.21	0.06
	work	6.33	0.00	0.07	13.41	6.25	0.52
	other	4.78	0.00	3.90	3.19	4.95	6.79
R3	*n*	*887*	*0*	*56*	*182*	*501*	*138*
Setting	**home**	2.17	-	2.73	2.35	2.21	1.60
	school	0.67	-	7.27	0.42	0.05	0.51
	work	5.37	-	0.00	5.65	7.16	0.86
	other	4.89	-	4.21	3.04	5.74	4.48
R4	*n*	*836*	*1*	*43*	*166*	*480*	*141*
Setting	**home**	2.16	2.00	2.42	2.67	2.20	1.40
	school	0.23	0.00	1.53	0.29	0.16	0.00
	work	4.27	0.00	0.35	8.76	4.13	0.67
	other	5.00	7.00	4.21	2.57	5.94	5.01

1. No adjustment for the age distribution of the population or participants has been made to mean contact numbers.

Table S6. Mean number of contacts involving touch, stratified by study wave, and age group of participant and contact.

		Age of participant[1]				
Study wave		**0-4**	**5-19**	**20-39**	**40-64**	**65+**
R1	*n*	*1*	*80*	*203*	*629*	*144*
	0-4	1.00	0.06	0.25	0.22	0.09
	5-19	1.00	3.23	0.49	1.31	0.06
Age of contact	**20-39**	4.00	1.27	1.75	1.29	0.28
	40-64	3.00	1.69	1.50	1.68	0.83
	65+	0.00	0.16	0.38	0.41	0.61
R2	*n*	*1*	*61*	*189*	*586*	*145*
	0-4	0.00	0.03	0.30	0.27	0.16
	5-19	0.00	6.03	0.71	1.12	0.24
Age of contact	**20-39**	0.00	0.77	1.83	0.96	2.06
	40-64	1.00	1.00	1.14	1.12	0.97
	65+	0.00	0.07	0.35	0.43	0.72
R3	*n*	*0*	*56*	*182*	*501*	*138*
	0-4	-	0.04	0.14	0.46	0.10
	5-19	-	5.50	0.60	1.01	0.09
Age of contact	**20-39**	-	0.64	1.40	0.68	0.26
	40-64	-	1.27	1.06	1.27	0.54
	65+	-	0.07	0.46	0.61	0.34
R4	*n*	*1*	*43*	*166*	*480*	*141*
	0-4	0.00	0.05	0.13	0.15	0.07
	5-19	2.00	1.86	0.70	0.48	0.10
Age of contact	**20-39**	0.00	1.02	2.08	1.15	0.26
	40-64	5.00	1.44	1.23	1.89	0.33
	65+	0.00	0.07	0.22	0.59	0.64

1. No adjustment for the age distribution of the population or participants has been made to mean contact numbers.
Diagonal (same age mixing) is marked for convenience.

Appendix E: Force of infection based on wave-specific age-mixing.

Here we estimate the Force of infection (λ) for each study wave using the observed age-mixing patterns as shown in Figure 3 of the main text. We assume that the population is entirely susceptible, except for initial infecteds, and that the transmission rate per contact is constant and independent of participant or contact age. The population size in each age class is N_i, based on Hong Kong Census information. We make two assumptions regarding the number of initial infectious individuals in each age class I_i: (1) the number in each age category within the population is a fixed proportion, $I_i = pN_i$; (2) the number in each age category is the same across all age categories, $I_i = \frac{N_i}{\sum N_i} \times \sum pN_i$, such that the total number of infectious individuals under each assumption is the same.

We define the force of infection for each age class as

$$\lambda_i = \sum_j \beta Q_{ij} I_j$$

where Q_{ij} is the observed average number of contacts made by age class i with age class j per day, and β is the transmission rate per day given contact.

We present the age-class specific force of infection by study wave in figures **S5A** and **S5B** below and the total Force of Infection ($\sum \lambda_i$) by wave for each assumption regarding initial infected in figure **S5C**. We used $\beta = 1e^{-4}$, and $p = 1e^{-4}$, which corresponds to 697 initial infecteds in a population of 6,971,882. We excluded the 0 to 4 age class from our force of infection calculations (though did include this age group as potential infectors) due to the low number of observations in our study for these ages.

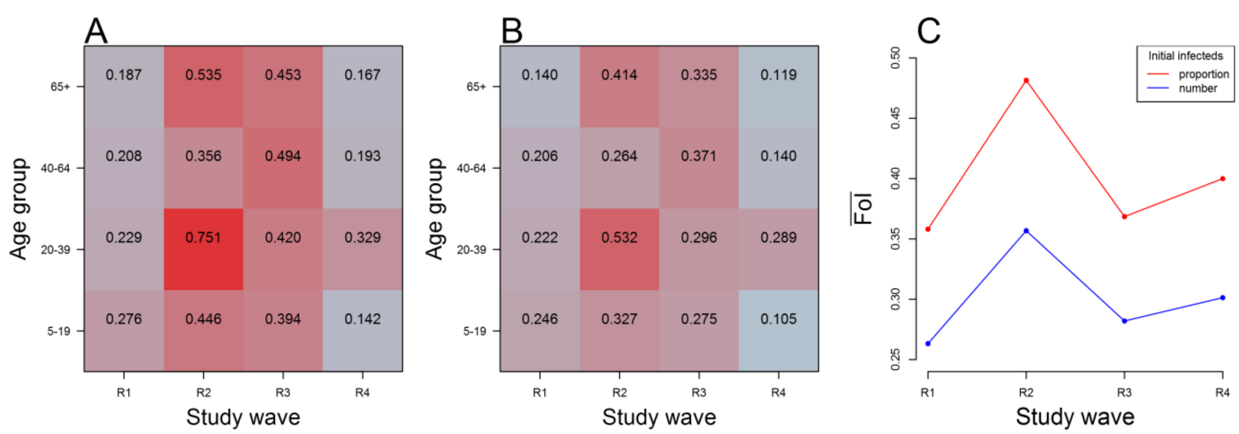


Figure S5. Force of infection estimates based on observed age-specific mixing patterns for each study wave. **A**ge-pair specific force of infection estimates assuming the same proportion (A) or number (B) of infecteds in each age group. (C) Average force of infection across age-groups, weighted by census-derived population size of each age group, over each study wave.

Appendix F. Regression models: fixed and random effects, and sensitivity analyses.

Table S7. Estimated fixed effects of the regression models shown in Figure 4 of the main text, excluding spline terms.

		Model outcome								
		Number of contacts			**Duration of contact**			**Number of locations**		
Variable		Estimate	Standard Error	P value	Estimate	Standard Error	P value	Estimate	Standard Error	P value
Intercept		1.219	0.049	<0.001	1.621	0.056	<0.001	1.293	0.023	<0.001
Sex	Male	0	-	-	0	-	-	0	-	-
	Female	0.027	0.032	0.393	0.013	0.035	0.711	0.020	0.017	0.233
Day of week	Saturday	0	-	-	0	-	-	0	-	-
	Sunday	-0.015	0.041	0.719	-0.005	0.048	0.912	-0.014	0.024	0.555
	Monday	0.096	0.041	0.019	-0.080	0.047	0.089	0.010	0.023	0.669
	Tuesday	0.108	0.042	0.009	-0.064	0.048	0.182	0.015	0.024	0.518
	Wednesday	0.135	0.044	0.002	-0.184	0.051	<0.001	0.019	0.025	0.459
	Thursday	0.122	0.043	0.004	-0.241	0.050	<0.001	0.003	0.024	0.915
	Friday	0.073	0.044	0.096	-0.132	0.050	0.009	0.089	0.025	<0.001
Number of locations	0	0	-	-	0	-	-	-	-	-
	1	0.679	0.036	<0.001	0.380	0.042	<0.001	-	-	-
	2	0.882	0.037	<0.001	0.528	0.043	<0.001	-	-	-
	3	1.123	0.042	<0.001	0.607	0.048	<0.001	-	-	-
	4	1.290	0.051	<0.001	0.695	0.059	<0.001	-	-	-
	5	1.320	0.062	<0.001	0.749	0.071	<0.001	-	-	-
	6 or more	1.446	0.063	<0.001	0.884	0.072	<0.001	-	-	-
Study wave	R1	0	-	-	0	-	-	0	-	-
	R2	0.089	0.029	0.002	-0.085	0.034	0.012	-0.003	0.017	0.857
	R3	0.056	0.030	0.062	-0.049	0.035	0.162	0.002	0.017	0.917
	R4	0.009	0.030	0.755	0.018	0.036	0.617	0.005	0.018	0.778

(Table S7 continued on following page)

Table S7 continued.

		Model outcome								
		Number of contacts Home			Number of contacts School or work			Number of contacts Other		
Variable		Estimate	Standard Error	P value	Estimate	Standard Error	P value	Estimate	Standard Error	P value
Intercept		0.851	0.044	<0.001	0.013	0.078	0.864	0.220	0.053	<0.001
Sex	Male	0	-	-	0	-	-	0	-	-
	Female	-0.006	0.022	0.804	-0.162	0.049	0.001	0.129	0.030	<0.001
Day of week	Saturday	0	-	-	0	-	-	0	-	-
	Sunday	-0.054	0.032	0.089	-0.084	0.065	0.199	0.148	0.045	0.001
	Monday	-0.083	0.031	0.008	0.486	0.065	<0.001	-0.158	0.045	<0.001
	Tuesday	-0.095	0.032	0.003	0.492	0.066	<0.001	-0.133	0.046	0.004
	Wednesday	-0.109	0.034	0.001	0.564	0.070	<0.001	-0.164	0.049	0.001
	Thursday	-0.139	0.033	<0.001	0.613	0.068	<0.001	-0.177	0.047	<0.001
	Friday	-0.096	0.033	0.004	0.533	0.069	<0.001	-0.159	0.048	0.001
Household size	1	0	-	-	-	-	-	-	-	-
	2	0.095	0.040	0.018	-	-	-	-	-	-
	3	0.295	0.038	<0.001	-	-	-	-	-	-
	4	0.445	0.039	<0.001	-	-	-	-	-	-
	5	0.541	0.049	<0.001	-	-	-	-	-	-
	6+	0.585	0.064	<0.001	-	-	-	-	-	-
Number of locations	0	-	-	-	0	-	-	0	-	-
	1	-	-	-	0.501	0.057	<0.001	0.586	0.040	<0.001
	2	-	-	-	0.500	0.059	<0.001	1.125	0.041	<0.001
	3	-	-	-	0.603	0.066	<0.001	1.506	0.045	<0.001
	4	-	-	-	0.566	0.082	<0.001	1.865	0.056	<0.001
	5	-	-	-	0.565	0.098	<0.001	1.887	0.068	<0.001
	6 or more	-	-	-	0.449	0.099	<0.001	2.169	0.069	<0.001
Study wave	R1	0	-	-	0	-	-	0	-	-
	R2	0.009	0.023	0.682	0.120	0.046	0.009	0.020	0.033	0.542
	R3	-0.064	0.023	0.006	0.073	0.048	0.127	0.060	0.034	0.079
	R4	-0.052	0.024	0.030	-0.007	0.049	0.889	0.059	0.035	0.090

Table S8. Variance within the random effects of the regression models. This table contains the variance associated with the random effect terms of the models presented in Figure 4 of the main text (Section A, using all observations), and those from additional regression models which restricted the observations to those from participants reporting their contact day was a 'typical' day (section B) or where the outcome was the number of contacts reported in a specific setting (section C).

Data	Outcome variable	Individual variation	Variance	% variance
A All observations (no. obs. = 3,383; n. parts. = 1,123)	Number of contacts	inter	0.14622	33.7
		intra	0.28756	66.3
	Duration of contacts	inter	0.16054	28.6
		intra	0.40113	71.4
	Number of locations	inter	0.03523	25.9
		intra	0.10069	74.1
B Only days reported by participants as 'typical' (n. obs.= 2,016; n. parts. = 740)	Number of contacts	inter	0.18080	42.3
		intra	0.24676	57.7
	Duration of contacts	inter	0.17078	30.7
		intra	0.38556	69.3
	Number of locations	inter	0.03969	34.0
		intra	0.07711	66.0
C All observations (no. obs. = 3,382; n. parts. = 1,123)	Number of contacts made in home settings	inter	0.05309	22.0
		intra	0.18820	78.0
	Number of contacts made in school or work settings	inter	0.31160	29.3
		intra	0.75138	70.7
	Number of contacts not made in home, school or work settings	inter	0.07372	14.9
		intra	0.42127	85.1

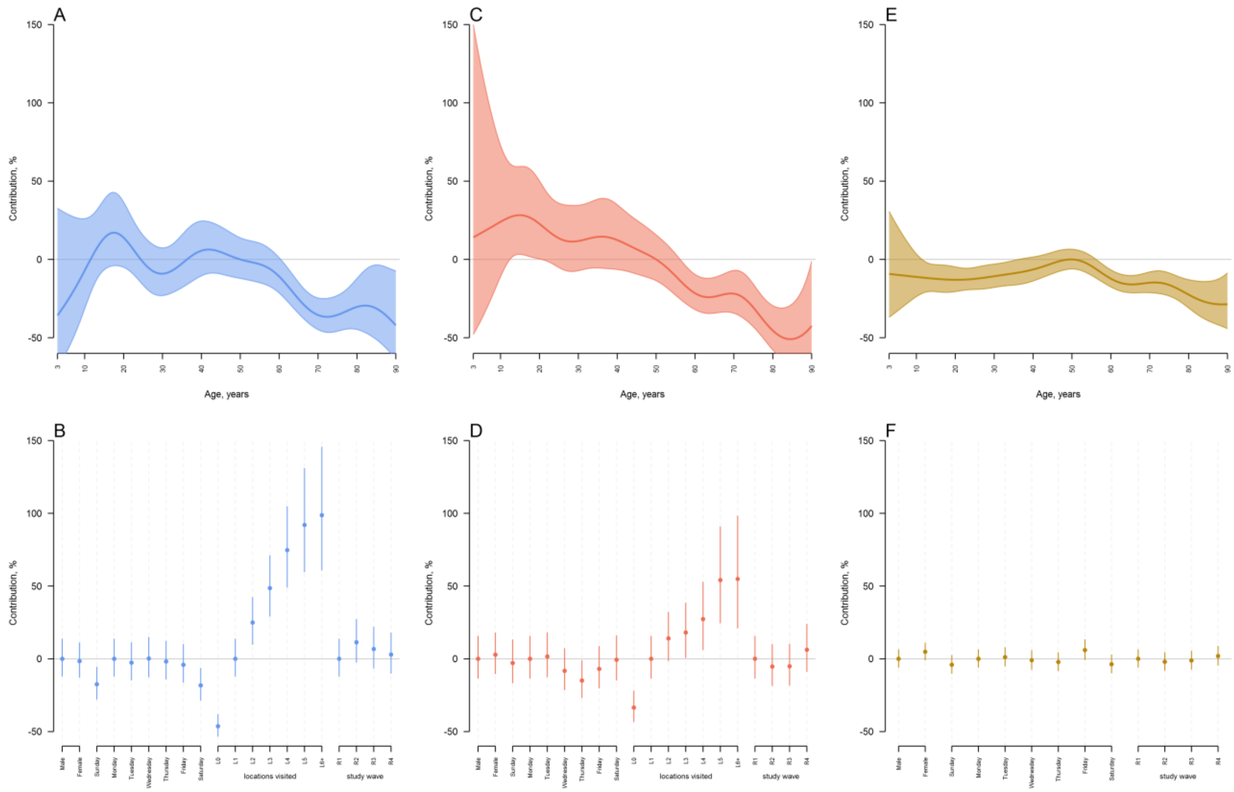


Figure S6. Modelled contact rates for days reported by participants as being 'typical'. Here we show the percentage contribution to contact rate (number of contact, duration of contact, number of locations) by the various covariates included in each model, relative to the contact rate predicted for a male 50-year-old from a household of size 1, on a Monday, with one contact location and during study wave 1. Models were fitted to data restricted to observations where participants reported their reporting day to be 'typical' and participants for whom there were at least two observations. Outcome variables were number of contacts (A and B), the total duration of contact events (C and D), and the number of locations in which contact occurred (E and F).

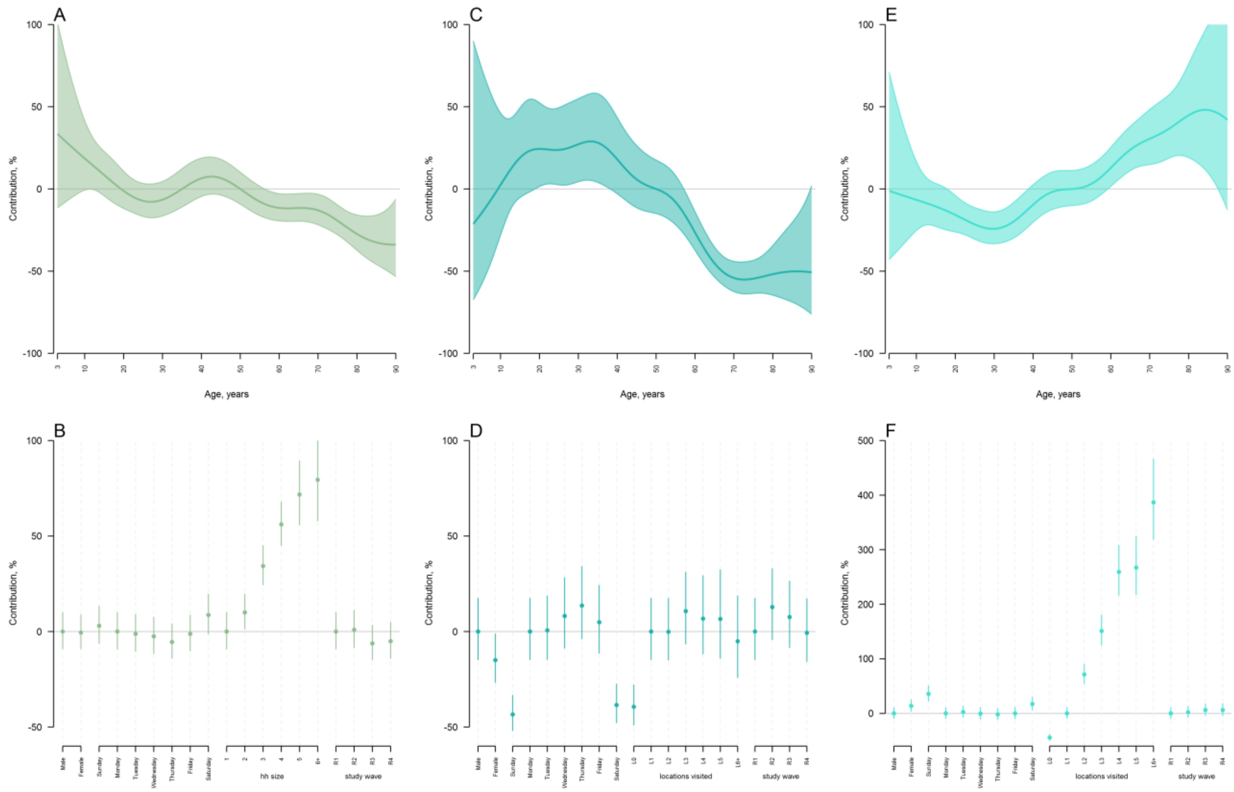


Figure S7. Modelled number of contact made in (A,B) home, (C,D) work or school, and (E,F) other social settings. Here, we show the percentage contribution to the number of contacts by covariates included in each model, relative to the contact rate predicted for a male 50-year-old from a household of size 1, on a Monday, with one contact location and during study wave 1. Models were fitted to data restricted to observations where participants for whom there were at least two observations. Note, for the number of contacts at home model (A,B), we excluded number of locations as an explanatory variable and instead included household size as a variable.

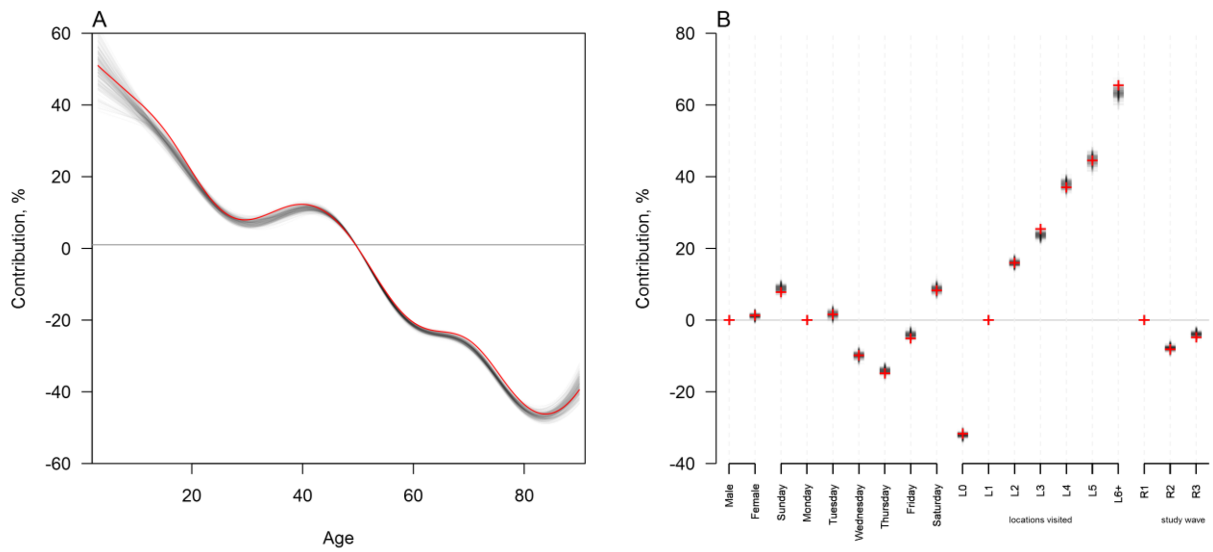


Figure S8. Sensitivity of fitted regression model to estimated contact durations. 200 estimates of contact duration were calculated for each contact event reported (with a valid contact duration category), and consequently there are 200 estimates of the total contact duration for each observation (participant-wave). Here, we explore the sensitivity of the regression model fit presented in Figure 4b/4d and Table S3 by fitting the same regression model to each set of 200 observations independently. (A) Predicted age contribution curves for all 200 models (grey lines) and the model reported in the paper (red line). (B) Predicted contribution for the other fixed effects of the 200 models (black crosses) and the model reported in the paper (red crosses).

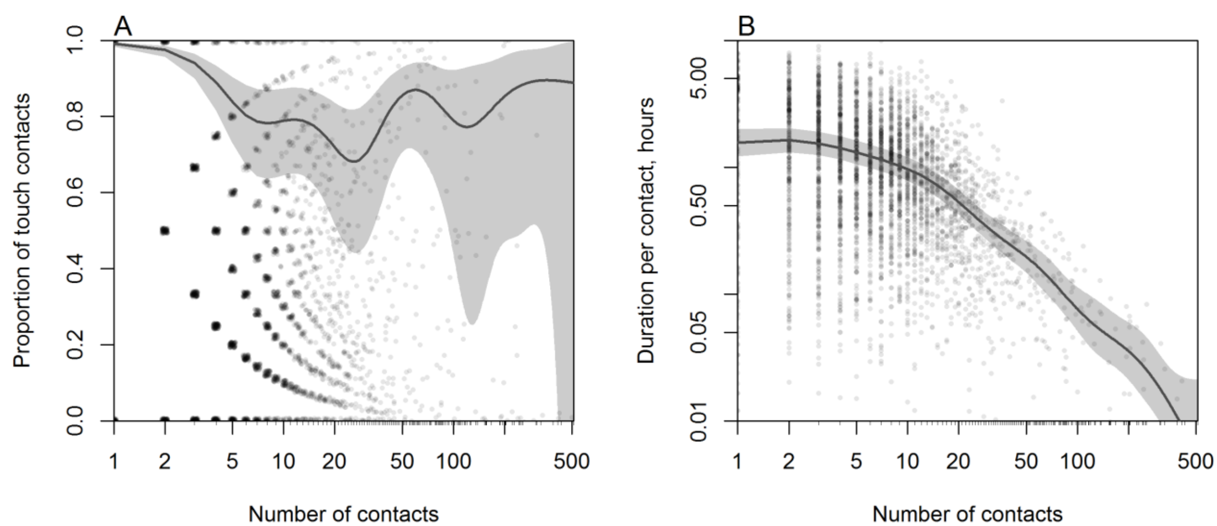


Figure S9. Regression models exploring the relationship between the proportion of contacts involving touch and the average duration per contact and the total number of contacts reported. The plots show the splines fitted to logged number of contacts as an explanatory variable, with (**A**) proportion of touch contacts and (**B**) the log average duration per contact as response variables. Models adjusted for age (spline), sex, day of the week, household size, and study. Raw observations are shown as points in both plots, though they are jittered in **A** for clarity.

Appendix G. Individual-level variation in contact rates.

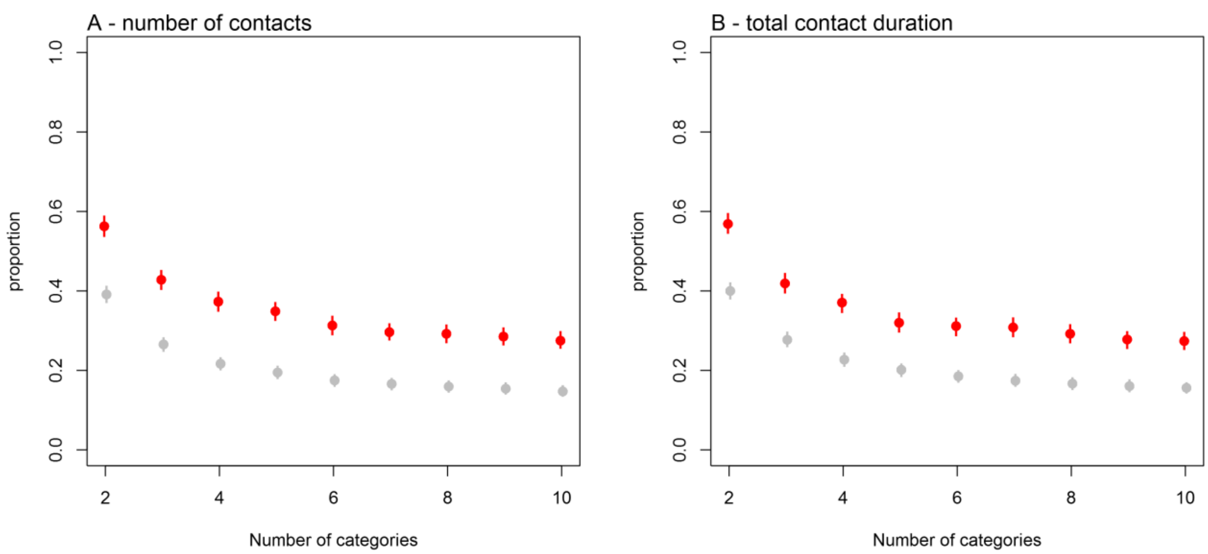


Figure S10. Intra-participant variation in contact rates. Proportion of individuals (with two or more observations) who remain within a single contact rate quantile category across all waves, against the number of quantiles used, for (A) number of contacts and (B) contact duration. Bootstrap estimates for both observed data (red) and null model synthetic data (grey) are shown. Null 'synthetic' data was generated from our observed data, where the individual-level contact metrics for study wave are resampled without replacement from the observations – essentially breaking the within-individual dependencies of our observed contact rates, while preserving the distribution of rates within each wave. Here, we assign each participant's wave-specific contact rate into a quantile category. Category breaks were defined by finding the required number of quantiles from all observed contact rates for individuals participating for their first time. We excluded individuals for which there was only a single (wave) observation. Lines represent bootstrapped 95% confidence intervals, which were generated through 500 resamples.

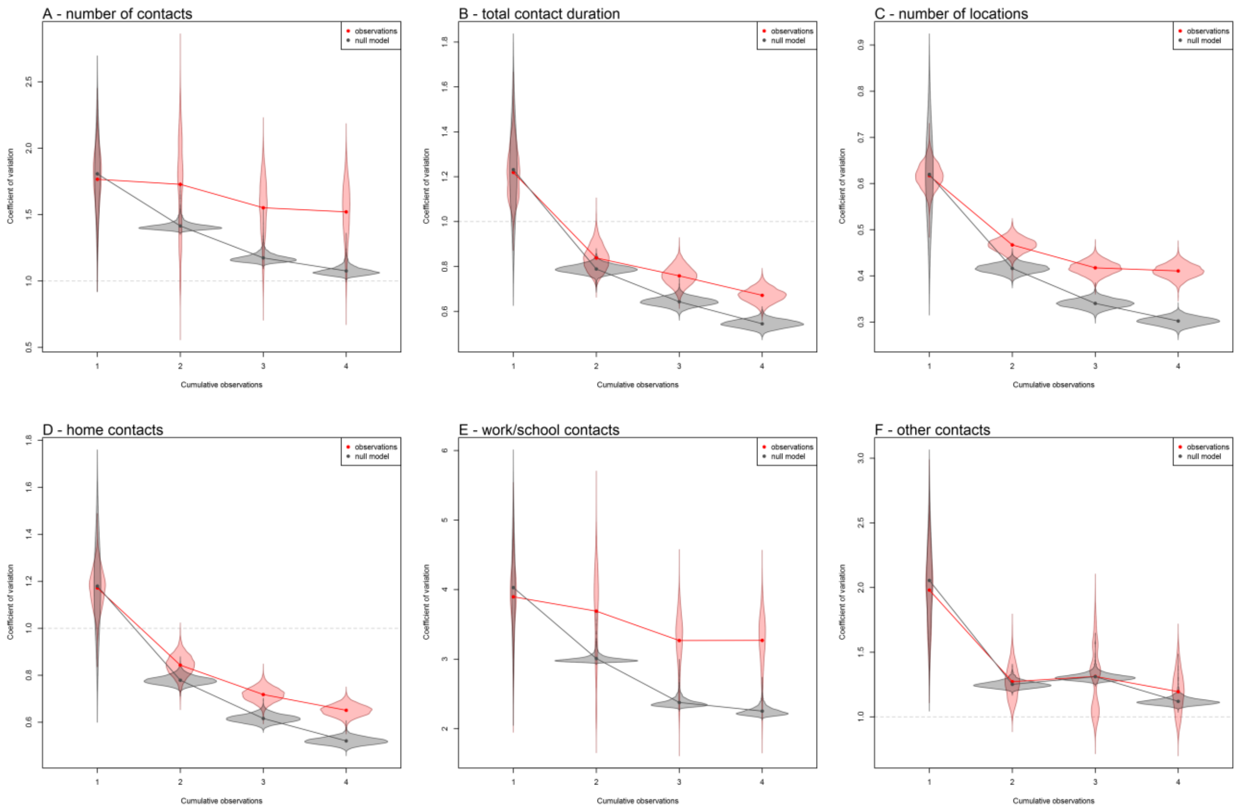


Figure S11. Plots examining how the coefficient of variation of different contact metrics changes as observations accumulate with each study wave. Contact metrics are (A) number of contacts, (B) total contact duration, (C) number of locations in which contact was made, (D) number of contacts made in home setting, (E) number of contacts made in school or work settings, (F) number of contacts made in other settings. Only individuals who participated in all four waves were considered (*n*=401). For each level of cumulative observations and for each individual, we calculate the cumulative contact metric reported (total number of contacts, total duration or total number of locations). We then calculate the coefficient of variation for that population of individuals. Pale red regions represent the distributions of CoV derived from 5,000 bootstrap resamples of the 401 participants. Grey regions show the equivalent CoV distributions for observation based synthetic data, where the individual-level contact metrics for study wave are resampled without replacement from the observations – essentially breaking the within-individual dependencies of our observed contact rates, while preserving the distribution of rates within each wave – and cumulative metrics derived.

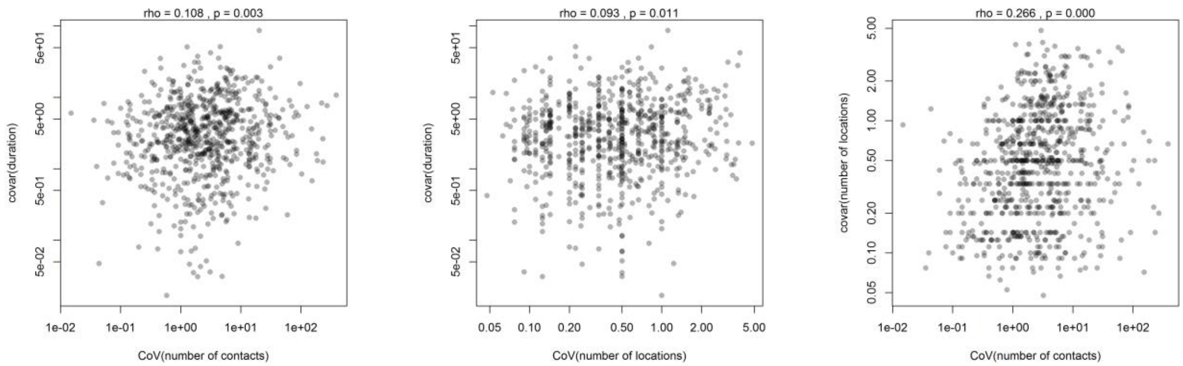


Figure S12. Relationship between individual-level coefficient of variation (CoV) for number of contacts, contact duration and number of locations. Spearman correlation estimates and associated p-values are shown above their corresponding plot. Only participants with 3 or more observations are included (number of participants=803).

Appendix H. Contacts reported by participants as individual or group contacts.

Table S9. Mean and standard deviation of reported individual and group contacts, for each study wave.

	Study wave[1]			
	R1	**R2**	**R3**	**R4**
Combined contacts	11.91 (20.34)	14.10 (32.76)	12.67 (23.57)	11.25 (24.62)
Individual contacts	5.25 (4.47)	4.34 (3.39)	4.62 (3.81)	4.50 (3.58)
Group contacts	6.66 (20.18)	9.76 (32.90)	8.05 (23.52)	6.74 (24.51)
Number of groups	0.61 (1.08)	0.93 (1.32)	0.82 (1.28)	0.83 (1.20)
Group size	11.34 (18.67)	10.22 (23.03)	9.46 (14.88)	8.72 (22.70)
Number of locations	2.98 (1.71)	2.96 (1.64)	3.01 (1.86)	3.00 (1.86)

1. mean (standard deviation)

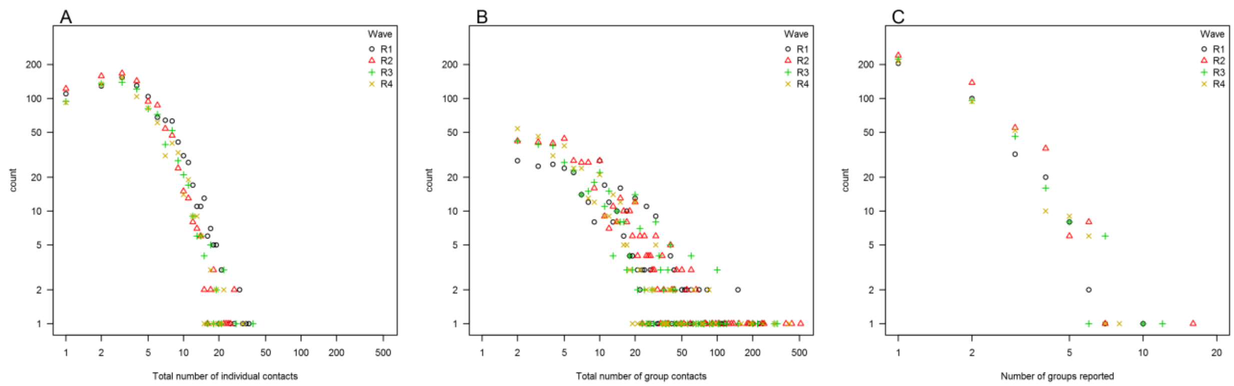


Figure S13. Distribution of number of contacts reported as individuals and group contacts, stratified by study wave. Distributions for (A) individual contacts, (B) group contacts, and (C) Distribution of the number of groups reported by participants, stratified by study wave.

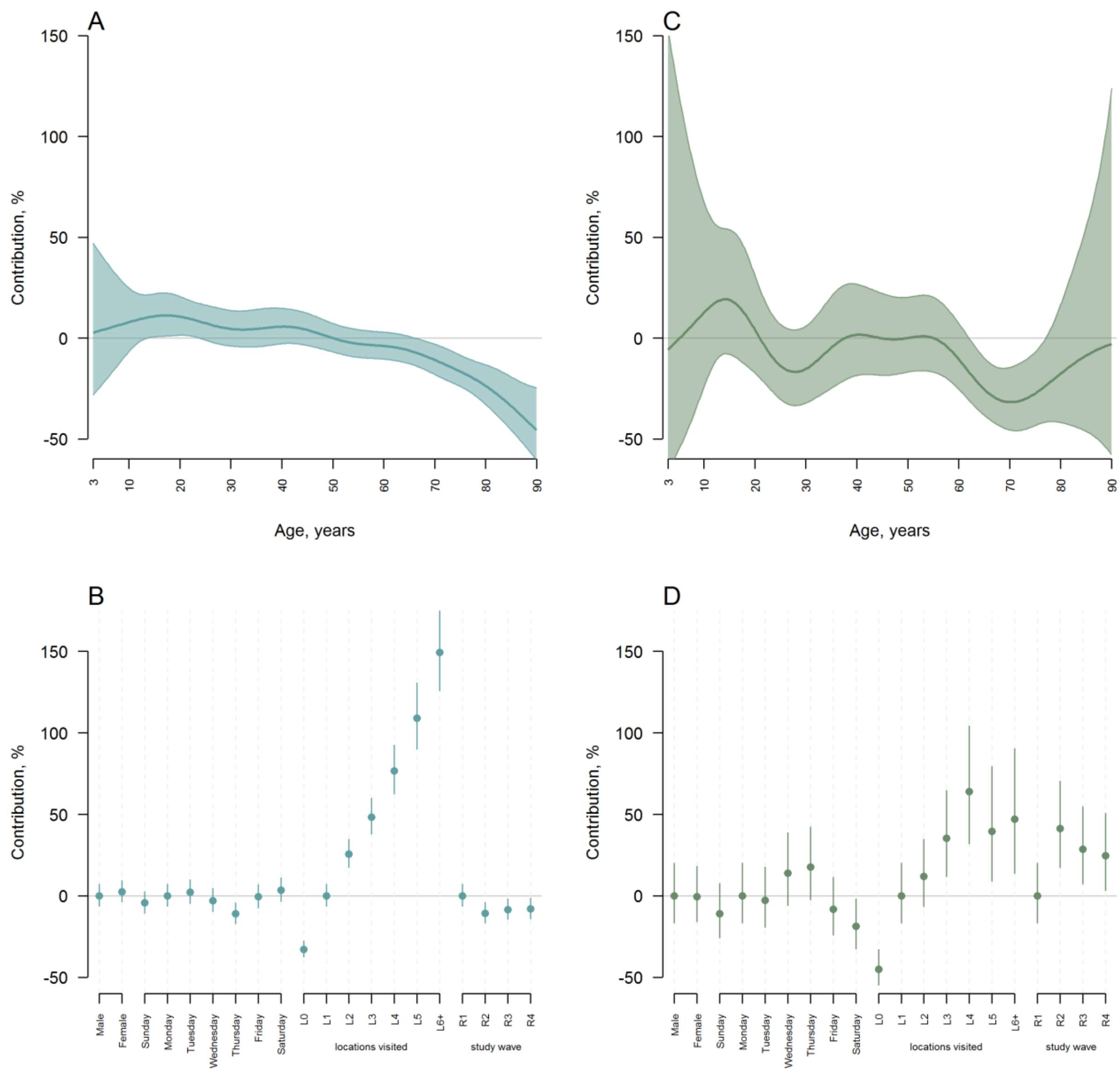


Figure S14. Modelled number of contacts reported as (A,B) individuals and (C,D) groups. Here, we show the percentage contribution to the number of contacts by covariates included in each model, relative to the contact rate predicted for a male 50-year-old from a household of size 1, on a Monday, with one contact location and during study wave 1. Regression analysis performed as for the main text, apart from the new outcome variables.

Appendix I: Data Release

Two data files are released with this manuscript: **hk_contact_number.csv** and **hk_contact_duration.csv**.

- **hk_contact_number.csv** – Information for each participant observation (a single recording day within a wave) including participant information, total number of contacts and number of locations. Rows = 3784, columns = 26. A data dictionary is provided in Table S9.
- **hk_contact_duration.csv** – 200 estimates of the total duration (minutes) of contact events, corresponding to the observations in **hk_contact_number.csv**. Rows = 3784, columns = 200. Column 1 was used as the outcome variable for the regression models presented in the main text and ESM. The value of theta used in the exponential model to estimate the contact durations was 0.01355328, and was fitted using an adaptation of the expectation–maximization algorithm as described in Read, J.M., et al., Social mixing patterns in rural and urban areas of southern China. *Proc Biol Sci*, 2014. 281(1785): p. 20140268.

Table S10. Data dictionary for released data.

Variable name	Description	Type
pid	Participant ID code.	Integer
hid	Household ID code.	Integer
age	Age (years) of the participant on the day of the observation. Participants older than 85 are assigned an age of 85 to preserve anonymity.	Integer
sex	Sex of the participant.	Categorical
n.samples	Number of observations (in total) for this individual	Integer
wave	Study wave, corresponding to R1 – R4.	Integer
reporting.day	Day of the week for which contact was reported.	Categorical
typical.day	Was this a typical day? (yes, no)	Categorical
n.contact.total	Total number of contacts reported. Specifically, the total number of unique individuals as identified through unique person/group descriptors and number of individuals reported (if a group).	Integer
n.locations	Number of unique contact locations reported	Integer
n.loc.group	Number of unique contact locations reported: 0, 1, 2, 3, 4, 5, 6+.	Categorical
n.contact.0-4	Number of contacts reported where contacts were between 0 and 4 years old.	Integer
n.contact.5-19	Number of contacts reported where contacts were between 5 and 19 years old.	Integer
n.contact.20-39	Number of contacts reported where contacts were between 20 and 39 years old.	Integer
n.contact.40-64	Number of contacts reported where contacts were between 40 and 64 years old.	Integer
n.contact.65+	Number of contacts reported where contacts were 65 years old or older.	Integer
n.contact.touch	Number of contacts involving touch	integer
n.contact.0-4.touch	Number of contacts involving touch where contacts were between 0 and 4 years old.	Integer
n.contact.5-19 touch	Number of contacts involving touch where contacts were between 5 and 19 years old.	Integer
n.contact.20-39 touch	Number of contacts involving touch where contacts were between 20 and 39 years old.	Integer
n.contact.40-64 touch	Number of contacts involving touch where contacts were between 40 and 64 years old.	Integer
n.contact.65+ touch	Number of contacts involving touch where contacts were 65 years old or older.	Integer
n.contact.home	Number of contacts made within a home setting.	Integer
n.contact.school	Number of contacts made within a school setting.	Integer
n.contact.work	Number of contacts made within a workplace setting.	Integer
n.contact.other	Number of contacts made within any other setting.	Integer