Article:

Mousa, A., Winskill, P., Watson, O.J. et al. (2022) Social contact patterns and implications for infectious disease transmission: a systematic review and meta-analysis of contact surveys. eLife, 10. e70294. ISSN: 2050-084X

https://doi.org/10.7554/elife.70294

Social Contact Patterns and Implications for Infectious Disease Transmission: A Systematic Review and Meta-Analysis of Contact Surveys

Andria Mousa[1*], Peter Winskill[1], Oliver J Watson[1], Oliver Ratmann[2], Mélodie Monod[2], Marco Ajelli[3,4], Aldiouma Diallo[5], Peter J Dodd[6], Carlos G Grijalva[7], Moses Chapa Kiti[8], Anand Krishnan[9], Rakesh Kumar[9], Supriya Kumar[10], Kin On Kwok[11,12,13], Claudio F Lanata[14,15], Olivier Le Polain de Waroux[16], Kathy Leung[17,18], Wiriya Mahikul[19], Alessia Melegaro[20], Carl D Morrow[21,22], Joël Mossong[23], Eleanor FG Neal[24,25], David J Nokes[8,26], Wirichada Pan-ngum[27], Gail E Potter[28,29], Fiona M Russell[24,25], Siddhartha Saha[30], Jonathan D Sugimoto[31,32,33], Wan In Wei[11], Robin R Wood[21], Joseph T Wu[17,18], Juanjuan Zhang[34], Patrick GT Walker[1] & Charles Whittaker[1*]

*For correspondence: a.mousa17@imperial.ac.uk; charles.whittaker16@imperial.ac.uk

[1] MRC Centre for Global Infectious Disease Analysis; and the Abdul Latif Jameel Institute for Disease and Emergency Analytics (J-IDEA), School of Public Health, Imperial College London, London, UK.

[2] Department of Mathematics, Imperial College London, London, UK.

[3] Department of Epidemiology and Biostatistics, Indiana University School of Public Health, Bloomington, IN, USA.

[4] Laboratory for the Modeling of Biological and Socio-technical Systems, Northeastern University, Boston, MA.

[5] VITROME, Institut de Recherche pour le Developpement, Senegal.

[6] School of Health and Related Research, University of Sheffield, UK.

[7] Division of Pharmacoepidemiology, Department of Health Policy. Vanderbilt University Medical Center. Nashville, TN, USA.

[8] KEMRI-Wellcome Trust Research Programme, Kilifi, Kenya.

[9] Centre for Community Medicine, All India Institute of Medical Sciences, New Delhi, India.

[10] Bill & Melinda Gates Foundation, Seattle, USA.

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

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

[13] Shenzhen Research Institute of The Chinese University of Hong Kong, Shenzhen, China.

[14] Instituto de Investigación Nutricional, Lima, Peru.

[15] Department of Medicine, Vanderbilt University, Nashville, TN, USA.

[16] London School of Hygiene and Tropical Medicine, London, UK.

[17] WHO Collaborating Centre for Infectious Disease Epidemiology and Control, School of Public Health, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China.

[18] Laboratory of Data Discovery for Health (D24H), Hong Kong Science Park, New Territories, Hong Kong SAR, China.

[19] Faculty of Medicine and Public Health, HRH Princess Chulabhorn College of Medical Science, Chulabhorn Royal Academy, Bangkok 10210, Thailand.

[20] Dondena Centre for Research on Social Dynamics and Public Policy, Department of Social and Political Sciences, Bocconi University, Milan, Italy.

[21] Desmond Tutu HIV Centre, Department of Medicine, Faculty of Health Sciences, University of Cape Town, South Africa.

[22] Centre for Infectious Disease Epidemiology and Research (CIDER), School of Public Health and Family Medicine, Faculty of Health Sciences, University of Cape Town South Africa.

[23] Health Directorate, Luxembourg.

[24] Infection & Immunity, Murdoch Children's Research Institute, Parkville, Victoria, Australia

[25] Department of Paediatrics, University of Melbourne, Parkville, Victoria, Australia.

[26] School of Life Sciences, University of Warwick, Coventry UK.

[27] Department of Tropical Hygiene, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand

[28]National Institute for Allergies and Infectious Diseases, National Institutes of Health, Rockville MD, USA.

[29] The Emmes Company, Rockville MD, USA.

[30] Influenza Programme, US Centers for Disease Control and Prevention, India Office, US Embassy, New Delhi.

[31] Seattle Epidemiologic Research and Information Center, Cooperative Studies Program, Office of Research and Development, United States Department of Veterans Affairs, USA.

[32] Department of Epidemiology, University of Washington, USA.

[33] Fred Hutchinson Cancer Research Center, Seattle, WA, USA.

[34] School of Public Health, Fudan University, Key Laboratory of Public Health Safety, Ministry of Education, Shanghai, China.

Abstract

Background: Transmission of respiratory pathogens such as SARS-CoV-2 depends on patterns of contact and mixing across populations. Understanding this is crucial to predict pathogen spread and the effectiveness of control efforts. Most analyses of contact patterns to date have focussed on high-income settings.

Methods: Here, we conduct a systematic review and individual-participant meta-analysis of surveys carried out in low- and middle-income countries and compare patterns of contact in these settings to surveys previously carried out in high-income countries. Using individual-level data from 28,503 participants and 413,069 contacts across 27 surveys we explored how contact characteristics (number, location, duration and whether physical) vary across income settings.

Results: Contact rates declined with age in high- and upper-middle-income settings, but not in low-income settings, where adults aged 65+ made similar numbers of contacts as younger individuals and mixed with all age-groups. Across all settings, increasing household size was a key determinant of contact frequency and characteristics, with low-income settings characterised by the largest, most intergenerational households. A higher proportion of contacts were made at home in low-income settings, and work/school contacts were more frequent in high-income strata. We also observed contrasting effects of gender across income-strata on the frequency, duration and type of contacts individuals made.

Conclusions: These differences in contact patterns between settings have material consequences for both spread of respiratory pathogens, as well as the effectiveness of different non-pharmaceutical interventions.

Funding: This work is primarily being funded by joint Centre funding from the UK Medical Research Council and DFID (MR/R015600/1).

Introduction

Previous outbreaks of Ebola(Mbala-Kingebeni et al., 2019), influenza(Khan et al., 2009), and the ongoing COVID-19 pandemic have highlighted the importance of understanding the transmission dynamics and spread of infectious diseases, which depend fundamentally on the underlying patterns of social contact between individuals. Together, these patterns give rise to complex social networks that influence disease dynamics(Eubank et al., 2004; Ferrari et al., 2006; Firth et al., 2020; Zhang et al., 2020), including the capacity for emergent pathogens to become endemic(Ghani and Aral, 2005; Jacquez et al., 1988), the overdispersion of the offspring distribution underlying the reproduction number(Delamater et al., 2019) and the threshold at which herd-immunity is reached(Fontanet and Cauchemez, 2020; Mistry et al., 2021). They can similarly modulate the effectiveness of non-pharmaceutical interventions (NPIs), such as school closures and workplace restrictions, that are typically deployed to control and contain the spread of infectious diseases (Prem et al., 2020).

Social contact surveys provide insight into the features of these networks, which is typically achieved through incorporating survey results into mathematical models of infectious disease transmission frequently used to guide decision making in response to outbreaks(Chang et al., 2021; Davies et al., 2020). Such inputs are necessary for models to have sufficient realism to evaluate relevant policy questions. However, despite the known importance of contact patterns as determinants of the infectious disease dynamics, our understanding of how they vary globally remains far from complete. Reviews of contact patterns to date have focussed on High-Income countries (HICs)(Hoang et al., 2019). This is despite evidence that social contact patterns differ systematically across settings in ways that have material consequences for the dynamics of infectious disease transmission and the evolution of epidemic trajectories(Prem et al., 2017; Walker et al., 2020). Previous reviews has also primarily explored the total number of contacts made by individuals(Hoang et al., 2019) and/or how these contacts are distributed across different age/sex groups(Horton et al., 2020). Whilst these factors are a vital component underpinning disease spread, recent work has also underscored the importance of the characteristics of contacts (such as the location, duration and extent of physical contact) in determining transmission risk(Thompson et al., 2021).

Here, we carry out a systematic review of contact surveys (conducted prior to the emergence of COVID-19) in Lower-Income, Lower-Middle and Upper-Middle-Income countries (LICs, LMICs and UMICs, respectively). Alongside previously published data from HICs(Kwok et al., 2018, 2014; Leung et al., 2017; Mossong et al., 2008), we collate individual participant data (IPD) on social contacts from published work spanning 27 surveys from 22 countries and over 28,000 individuals. We use a Bayesian framework to explore drivers and determinants of contact patterns across a wider range of settings and at a more granular scale than has previously been possible. Specifically, we assess the influence of key factors such as age, gender and household structure on both the total number and characteristics (such as duration, location and type) of contact made by an individual, and explore how the comparative importance of different factors varies across different settings. We additionally evaluate the extent and degree of assortativity in contact patterns between different groups, and how this varies across settings.

Methods

Systematic Review

Data sources and search strategy: Two databases (Ovid MEDLINE and Embase) were searched on 26th May 2020 to identify studies reporting on contact patterns in LICs, LMICs and UMICs (Appendix 1-Table 1). Collated records underwent title and abstract screening for relevance, before full-text screening using pre-determined criteria. Studies were included if they reported on any type of face-to-face or close contact with humans and were carried out in LICs, LMICs or UMICs only. No restrictions on collection method (e.g. prospective diary-based surveys or retrospective surveys based on a face-to-face/phone interview or questionnaire) were applied. Studies were excluded if they did not report contacts relevant to air-borne diseases (e.g. sexual contacts), were conducted in HICs, were contact tracing studies of infected cases, or were conference abstracts. All studies were screened independently by two reviewers (AM and CW). Differences were resolved through consensus and discussion. The study protocol can be accessed through PROSPERO (registration number: CRD42020191197). Income group classification (LIC/LMIC, UMIC, or HIC) was based on 2019 World Bank data (fiscal year 2021)(World Bank Group, 2020).

Data extraction: Individual-level data were obtained from publication supplementary data, as well as online data repositories such as Zenodo, figshare and OSF. When not publicly available, study authors were contacted to request data. Extracted data included the participant's age, gender, employment, student status, household size and total number of contacts, as well as the day of the week for which contacts were reported. Some studies reported information at the level of individual contacts and included the age, gender, location and duration of the contact, as well whether it involved physical contact. Individual-level data from HICs, not systematically identified, were used for comparison, and included three studies from Hong Kong(Kwok et al., 2018, 2014; Leung et al., 2017) and the 8 European countries from the POLYMOD study(Mossong et al., 2008). Data were collated, cleaned and standardised using Stata version 14. Country-specific average household size were obtained from the United Nations Database on Household Size and Composition(United Nations Department of Economic and Social Affairs Population Division, 2019). Gross domestic product based on purchasing power parity (GDP PPP) was obtained from the World Data Bank database(World Bank International Comparison Programme, 2021). Findings are reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist of items specific to IPD meta-analyses (Appendix 1 Table 2). Risk of bias was assessed using the AXIS critical appraisal tool used to evaluate quality of cross-sectional studies(Downes et al., 2016), modified to this study's objectives (Appendix 1 Table 3). Each item was attributed a zero or a one, and a quality score was assigned to each study, ranging from 0% ("poor" quality) to 100% ("good" quality). The individual-level data across all studies and analysis code are available at https://github.com/mrc-ide/contact_patterns (see Appendix 1 - Table 4 for data assumptions and Appendix – Table 5 for data dictionary).

Statistical analysis

The mean, median and interquartile range of total daily unique contacts were calculated for subgroups including country income status, individual study, survey methodology (diary-based or questionnaire/interview-based), survey day (weekday/weekend), and respondent characteristics such as age, sex, employment/student status and household size. Detailed description of data assumptions for each study can be found in Appendix 1 - Table 4.

A negative binomial regression model was used to explore the association between the total number of daily contacts and the participant's age, sex, employment/student status and household size, as well as methodology and survey day. Incidence rate ratios from these regressions are referred to as "Contact Rate Ratios" (CRRs). A sensitivity analysis was carried out that excluded additional contacts (such as additional work contacts, group contacts, and number missed out, which were recorded separately and in less detail by participants compared to their other contacts (Ajelli and Litvinova, 2017; Kumar et al., 2018; Leung et al., 2017; Zhang et al., 2020)). Logistic regressions were used to explore determinants of contact duration (<1hr/1hr+) and type (physical/non-physical), using the same explanatory variables as in the total contacts analyses. There were differences in the contact duration categories defined by studies, and the threshold of 1 hour for longer durations was used to maximise sample size, by allowing inclusion of all available data. An additional sensitivity analysis, weighing all studies equally within an income stratum, explored the impact of study size on the estimated CRRs and ORs for all main outcomes (total contacts, duration and whether physical). The proportion of contacts made at each location (home, school, work and other) was explored descriptively and contacts made with the same individual in separate locations/instances were considered as separate contacts.

All analyses were done in a Bayesian framework using the probabilistic programming language Stan, using uninformative priors in all analyses and implemented in R via the package *brms*(Bürkner, 2018, 2017). All analyses were stratified by three income strata (LICs and LMICs were combined to preserve statistical power) and included random effects by study, to account for heterogeneity between studies. The only exceptions to this were any models adjusting for methodology which did not vary by study. The effect of each factor was explored in an age- and gender-adjusted model. All models exploring the effect of student status or employment status were restricted to children aged between 5 and 18 years and adults over 18, respectively. In the remaining models including all ages, age was adjusted as a categorical variable (<15, 15 to 65 and over 65 years). CRRs, Odds Ratios (ORs) and their associated 95% Credible Intervals are presented for all regression models. Here, we report estimates adjusted for age and gender (referred to as adjCRR or adjOR). Studies which collated contact-level data were used to assess assortativity of mixing by age and gender for different country-income strata by calculating the proportions of contacts made by participants that are male or female and those that belong to three broad age groups (children, adults, and older adults).

Results

Systematic Review and Individual-Participant-Data (IPD) Meta-analysis

A total of 3,409 titles and abstracts were retrieved from the databases, and 313 full-text articles were screened for eligibility (Appendix 1- Figure 1). This search identified 19 studies with suitable contact data from LIC, LMIC and UMIC settings– individual-level data were obtained from 16 of these studies, including one study from a LIC, six studies from a LMIC and nine studies from an UMIC. These were analysed alongside four HIC studies from Hong Kong and Europe. The majority of the studies collected data representative of the general population, through random sampling and included a combination of both rural and urban sites (see Appendix 1 for further details). Although most studies included respondents of all ages, one study restricted their participants to ages over 18 years (Dodd et al., 2015), one to ages over 15 years(Mahikul et al., 2020), one to ages over 6 months(Huang et al., 2020), one study only collected contact data on infants under 6 months(Oguz et al., 2018) and another on contacts of children under 6 years and their caregivers(Neal et al., 2020). The distribution of participant age groups in each study was also dependent on the sampling method. For instance, two studies focused on school and university students and their contacts, thereby oversampling older children and young adults (Ajelli and Litvinova, 2017; Stein et al., 2014). Details of the identified studies and a full description of the systematic review findings can be found in Appendix 1 and Appendix 1-Table 5 and Appendix 1-Table 6.

In total, this meta-analysis yielded 28,503 participants reporting on 413,069 contacts. All studies contained information on main demographic variables such as age and gender. Availability of other variables analysed here for each study are listed in Appendix 1- Table 7. All studies reported the number of contacts made in the past 24 hours of (or day preceding) the survey. The definitions of contacts were broadly similar across studies (Appendix 1- Table 6). Specifically, contacts were defined as skin-to-skin (physical) contact or a two-way conversation in the physical presence of another person. All studies scored above 65% of the items on the AXIS risk of bias tool, suggesting good or fair quality (Appendix 1 - Table 3). Among all participants 47.5% were male, 30.1% were aged under 15 years and 7.2% were aged over 65 years. The majority (83.4%) of participants were asked to report the number of contacts they made on a weekday. A large proportion (34.1%) of respondents lived in large households of 6 or more people but this was largely dependent on income setting (LIC/LMIC=63.2%, UMIC=35.9%, HIC=4.9%). Among school-aged children (5 to 18 years), 88.1% were students, and 59.1% of adults aged over 18 were employed.

Total number of contacts and contact location

The median number of contacts made per day across all the studies was 9 (IQR= 5-17), and was similar across income strata (LIC/LMIC=10[5-17], UMIC=8[5-16], HIC=9[5-17]; Table 1). There was a large variation in contact rates across different studies, with the median number of daily contacts ranging from 4 in a Zambian setting(Dodd et al., 2015) to 24 in an online Thai survey(Stein et al., 2014). When stratifying by study methodology, median daily contacts was higher in diary-based surveys compared to interview-/questionnaire- based surveys, which was true across all income strata (Table 1, Appendix 2- Figure 1).

Overall, children aged 5 to 15 had the highest number of daily contacts (Figure 1A-C), although there was substantial variation between studies and across income-strata in how the number of daily contacts varied with age (Figure 1A-C). Across UMICs and HICs, the number of daily contacts made by participants decreased with age, with this decrease most notable in the oldest age-groups (adjCRR for 65+ vs. <15 years [95%CrI]: UMIC=0.67[0.63-0.71] and HIC=0.57[0.54-0.60]). By contrast, there was no evidence of contact rates declining in the oldest age-groups in LICs/LMICs (adjCRR for 65+ vs. <15 years [95%CrI]=0.94[0.89-1.00]). We observed contrasting effects of gender on the number of daily contacts, with men making more daily contacts compared to women in LICs/LMICs after accounting for age (adjCRR=1.17, 95%CrI:1.15-1.20; Figure 1D), but no effect of gender on total daily contacts for other income strata (CRR[95%CrI]: UMIC=1.01[0.98-1.04], HIC=0.99[0.97-1.02]). There were also differences in the number of daily contacts made according to the methodology used and whether the survey was carried out on a weekday or over the weekend – in both instances, contrasting effects of these factors on the number of daily contacts according to income strata were observed (Figures 1D-1F).

We also examined the influence of factors that might influence both the total number and location (home, work, school and other) of the contacts individuals make. Across all income-strata, students (defined as those currently in education, attending school and aged between 5 and 18 years) made more daily contacts than non-students aged between 5 and 18 (adjCRR [95%CrI]: LIC/LMIC=1.26[1.16-1.37], UMIC=1.18[1.03-1.35] and HIC=1.54[1.42-1.66]; Figure 1D-F). Similarly, we observed strong and significant effects of employment in all income strata, with adults who were employed having a higher number of total daily contacts compared to those not in employment (adjCRR [95%CrI]: LIC/LMIC= 1.17[1.12-1.23], UMIC= 1.07[1.03-1.13], HIC= 1.60[1.54-1.65]; Figure 1D-F). The number of daily contacts made at home were proportional to the participant's household size (Appendix 2 – Figure 2). Total daily contacts increased with household size (Figure 2A, Appendix 2 – Figure 1) across all income-strata; individuals living in large households (6+ members) had 1.47 (95%CrI:1.32-1.64) (LIC/LMICs), 2.58 (95%CrI:2.37-2.80) (UMICs) and 1.51 (95%CrI:1.40-1.63) (HICs) times more daily contacts than those living alone, after accounting for age and gender (Figure 1E-F). Sensitivity analyses excluding additional contacts (as defined in Methods), showed little difference in effect sizes for total daily contacts, and were strongly correlated with the effect sizes shown in Figure 1D-F (Appendix 2- Figure 3).

Motivated by this suggestion of strong, location-related (school, work and household) effects on total daily contact rates, we further explored the locations in which contacts were made. Contact location was known for 314,235 contacts, 42.7% of which occurred at home (13.1% at work, 12.5% at school and 31.7% in other locations). Across income-strata, there was significant variation in the proportion of contacts made at home – being highest in LICs/LMICs (68.3%) and lowest in HICs (37.0%) (Figure 2B). Age differences were also observed in the number of contacts made at home, particularly for LICs/LMICs (Figure 2C-2D). Relatedly, a higher proportion of contacts occurred at work and school (14.6 % and 11.3%) in HICs compared to LICs/LMICs (3.9% and 5.2%, respectively; Appendix 2 – Figure 4). Strong, gender specific patterns of contact location were also observed. Across all income strata males made a higher proportion of their contacts at work compared to females, although this difference was largest for LICs/LMICs (Appendix 2 – Figure 4 and Appendix 2 –

Figure 5). Further, we found significant variation between income strata in median household size (7 in LICs/LMICs, 5 in UMICs and 3 in HICs). This trend of decreasing household size with increasing country income was consistent with global data (Figure 2E). The larger households observed for LIC/LMIC settings were also more likely to be intergenerational – in LICs/LMICs, 59.4% of participants aged over 65 lived in households of at least 6 members compared to 17.5% in UMICs and only 2.2% in HICs.

Type and duration of contact

Data on the type of contacts (physical and non-physical) were recorded for 20,910 participants. The mean percentage of physical contacts across participants was 56.0% and was the highest for LICs/LMICs (64.5%). At the study level, the highest mean percentage of physical contacts was observed for a survey of young children and their caregivers conducted in Fiji(Neal et al., 2020) (84.0%) and the lowest in a Hong Kong contact survey(Leung et al., 2017)(18.9%). Physical contact was significantly less common among adults compared to children under 15 years in all settings (ORs ranged between 0.22 to 0.48) (Figure 3A-F). Despite the proportion of physical contacts generally decreasing with age, there was a higher proportion observed for adults aged 80 or over (Figure 3A-C). Contacts made by male participants were more likely to be physical compared to female participants in UMICs (adjOR= 1.13, 95%CrI=1.10-1.16) and HICs (adjOR= 1.09, 95%CrI=1.07-1.12), but in LICs/LMICs men had a lower proportion of physical contacts than women (adjOR= 0.81, 95%CrI=0.79-0.83; Figure 3D-F). Most physical contacts made by women in LICs were made at home (73.5%), whilst for HICs this was just 41.4% - similar differences across income-strata were observed for men, although the proportions were always lower than observed for women (62.4% for LIC/LMICs and 36.4% for HICs). Increasing household size was generally associated with a higher proportion of contacts being physical (for households of 6+ members compared to 1 member: adjCRR[95%CrI]: LIC/LMIC=1.73[1.48-2.02], UMIC= 1.30[1.12-1.52], HIC= 1.57[1.48-1.67]; Figure 3D-F). Employment was associated with having a significantly lower proportion of physical contacts in

LICs/LMICs (adjOR=0.83, 95%CrI:0.79-0.87) and HICs (adjOR=0.71, 95%CrI:0.69-0.73), but not in UMICs (adjOR=1.11, 95%CrI:1.03-1.19). The proportion of physical contacts among all contacts was the highest for households (70.4%), followed by schools (58.5%), community (55.7%) and work (33.6%) (Appendix 2 – Figure 6).

Data on the duration of contact (<1 or ≥1hr) were available for 22,822 participants. The percentage of contacts lasting at least 1 hour was 63.2% and was highest for UMICs (76.0%) and lowest for LICs/LMICs (53.1%). Across both UMICs and HICs, duration of contacts was lower in individuals aged over 15 years compared to those aged 0-15, with the extent of this disparity most stark for HICs (for ages 65+ compared to <15 years: adjCRR [95%CrI]: LIC/LMIC= 0.61[0.57-0.64], UMIC= 0.61[0.58-0.65], HIC= 0.35[0.33-0.37]; Figure 4A-F). We observed contrasting effects of gender across income-strata: males made longer-lasting contacts than females in UMICs (adjOR=1.11, 95%CrI=1.08-1.14); Figure 4D-F), but not in LIC/LMICs (adjOR=0.92, 95%CrI=0.90-0.95) or HICs (adjOR=0.98, 95%CrI=0.97-1.00). Participants reported shorter contacts on weekends compared to weekdays in LICs/LMICs (adjOR=0.91, 95%CrI: 0.88-0.95), and HICs (adjOR=0.95, 95%CrI: 0.92-0.97), but not in UMICs (adjOR=1.12, 95%CrI=1.03-1.21). Contacts lasting over an hour as a proportion of all contacts was highest for households (72.7%), followed by schools (67.9%), community (47.0%) and work (44.0%). However, it was only in HICs that there was a significant effect of being a student (adjOR=1.18, 95%CrI: 1.09-1.27; Figure 4D-F) on the proportion of contacts lasting ≥1 hour. For all income strata, the proportion of contacts >1h increased with increasing household size (Figure 4D-F). The sensitivity analysis weighing all studies equally within an income group yielded similar results to those from the main analysis (range of Pearson's correlation coefficients between main analysis and sensitivity analysis effect sizes: 0.92-1.00; Appendix 2 - Figure 7 and Appendix 2 – Table 1), and any differences are discussed in Appendix 2.

Assortativity by age and gender

Twelve studies collected information on the gender of the contact and eight studies contained information on age allowing assignment of contacts to one of the three age-groups described in Methods (Appendix 1 – Table 7, Appendix 2). We found evidence to suggest that contacts were assortative by gender for all income strata, as participants were more likely to mix with their own gender (Appendix 2 – Table 2 and Appendix 2 – Table 3). Mixing was also assortative by age, with participants more likely to contact individuals who belonged to the same age group this degree of age-assortativity was lowest for LICs/LMICs, where only 29% of contacts made by adults were with individuals of the same age group. By contrast, in HICs we observed a higher degree of assortative mixing, with most contacts (51.4%) made by older adults occurring with individuals belonging to the same age group.

Discussion

Understanding patterns of contact across populations is vital to predicting the dynamics and spread of infectious diseases, as well understanding the control interventions likely to have the greatest impact. Here, using a systematic review and individual-participant data meta-analysis of contact surveys, we summarise research exploring these patterns across a range of populations spanning 28,503 individuals and 22 countries. Our findings highlight substantial differences in contact patterns between income settings. These differences are driven by setting-specific sociodemographic factors such as age, gender, household structure and patterns of employment, which all have material consequences for transmission and spread of respiratory pathogens.

Across the collated studies, the total number of contacts was highest for school-aged children. This is consistent with previous results from HICs(Béraud et al., 2015; Fu et al., 2012; Hoang et al., 2019; Ibuka et al., 2016; Lapidus et al., 2013) and shown here to be generally true for LICs/LMICs and UMICs also. Interestingly however, we observed differences in patterns of contact in adults across income strata. Whilst contact rates in HICs declined in older adults, this was not observed in

LICs/LMICs, where contact rates did not differ in the oldest age-group compared to younger ages. This is consistent with variation in household structure and size across settings, with nearly two thirds of participants aged 65+ in included LIC/LMIC surveys living in large, likely intergenerational, households (6+ members), compared to only 2% in HICs. HICs were also characterised by more assortative mixing between age-groups, with older adults in LICs/LMICs more likely to mix with individuals of younger ages, again consistent with the observed differences between household structures across the two settings. These results have important consequences for the viability and efficacy of protective policies centred around shielding of elderly individuals (i.e. those most at risk from COVID-19 or influenza. In these settings other strategies may be required to effectively shield vulnerable populations, as has been previously suggested (Dahab et al., 2020).Our results support the idea of households as a key site for transmission of respiratory pathogens(Thompson et al., 2021), with the majority of contacts made at home. Our analysis highlights that the number of contacts made at home is mainly driven by household size. However, the relative importance of households compared to other locations is likely to vary across settings. We observed significant differences across income settings in the distribution of contacts made at home, work and school. The proportion of contacts made at home was highest for LIC/LMICs, where larger average household sizes were associated with more contacts, more physical contacts, and longer lasting contacts. By contrast, participants in HICs tended to report more contacts occurring at work and school. The lower number of contacts at work in LIC/LMIC may be explained by the types of employment (e.g agriculture in rural surveys) and a selection bias (women at home/homemakers more likely to be surveyed in questionnaire-based surveys). Our analyses similarly highlighted significant variation in the duration and nature of contacts across settings. Contacts made by female participants in LICs/LMICs were more likely to be physical compared to men, whilst the opposite effect was observed for HICs and UMICs, potentially reflecting context-specific gender roles. In all settings, we observed a general decline of physical contacts with age, except in the very old(Mossong et al., 2008), potentially reflecting higher levels of dependency and the need for physical care.

Altogether, these results suggest differences between settings in the comparative importance of different locations (such as the household or the workplace) to transmission of SARS-CoV-2, a finding which would likely modulate the impact of different NPIs (such as workplace or school closures, stay at home orders etc). Moreover, it suggests that previous estimates of NPI effectiveness (primarily derived from European data and settings (Brauner et al., 2021) may be of limited generalisability to non-European settings characterised by different structures and patterns of social contact. However, beyond highlighting heterogeneity in where and how transmission is likely to occur, it remains challenging to disentangle exactly how these differences in contact patterns would shape patterns of transmission. Whilst the collated data provide a cross-sectional snapshot into the networks of social contact underpinning transmission, they remain insufficient to completely resolve this network or its temporal dynamics. Our results therefore do not consider key features relevant to population-level spread and transmission (such as overall network structure or the extent of repeated contacts, which would be most likely to occur with household members) which previous work has demonstrated can have a significant impact on infectious disease dynamics, both in general terms (Bansal et al., 2010; Keeling and Eames, 2005) as well as with COVID-19 (Rader et al., 2020). It is in this context that recent results generating complete social networks (including both the frequency and identity of an individual's contacts) from high-resolution GPS data represent promising developments in understanding social contact networks and how they shape transmission (Firth et al., 2020).

There are important caveats to these findings. Data constraints limited the numbers of factors we were able to explore – for example, despite evidence(Kiti et al., 2014) suggesting that contact patterns differ across rural and urban settings, only 3 studies(Kiti et al., 2014; O. le Polain de Waroux et al., 2018; Neal et al., 2020) contained information from both rural and urban sites, allowing classification. Similarly, we were unable to examine the impact of socioeconomic factors such as household wealth, despite experiences with COVID-19 having highlighted strong socio-economic disparities in both transmission and burden of disease(De Negri et al., 2021; Routledge et al., 2021; Ward et al., 2021; Winskill et al., 2020) and previous work suggesting that poorer individuals are less likely to be employed in occupations amenable to remote working(Loayza, 2020). A lack of suitably detailed information in the studies conducted precludes analysis of these factors but highlights the importance of incorporating economic questions into future contact surveys, such as household wealth and house square footage. Other factors also not controlled for here, but that may similarly shape contact patterns include school holidays or seasonal variations in population movement and composition that we are unable to capture given the cross-sectional nature of these studies.

Another important limitation to these results is that we are only able to consider a limited set of contact characteristics (the location and duration of the contact and whether it was physical). Previous work has highlighted the importance of these factors in determining the risk of respiratory pathogen transmission(Chang et al., 2021; Dunne et al., 2018; Olivier le Polain de Waroux et al., 2018; Neal et al., 2020; Thompson et al., 2021), but only a limited number of studies reported whether a contact was “close” or “casual”(Kwok et al., 2018, 2014; O. le Polain de Waroux et al., 2018) and whether the contact was made indoors or outdoors(Wood et al., 2012); both factors likely to influence transmission risk(Bulfone et al., 2021; Chu et al., 2020). More generally, the relevance and comparative importance of different contacts to transmission likely varies according to the specific pathogen and its predominant transmission modality (e.g. aerosol, droplet, fomite etc). It is therefore important to note that these results do not provide a direct indication of explicit transmission risk, but rather an indicator of factors likely to be relevant to transmission.

Relatedly, it is also important to note that the studies collated here were conducted over a wide time-period (2005-2018). In conjunction with the cross-sectional nature of the included studies, this precludes us from being able to examine for potential time-related trends in contact patterns. Additionally, the collated surveys were all carried out prior to the onset of the SARS-CoV-2 pandemic. Previous work has documented significant alterations to patterns of social contact in response to individual-level behaviour changes or government implemented NPIs aimed at controlling SARS-CoV-2 spread, and that these changes are dynamic and time-varying (Gimma et al., 2021; McCreesh et al., 2021). A detailed understanding of the impact of changing contact patterns on disease spread necessarily requires both an understanding of baseline contact patterns (as detailed in the studies collated here), and what changes have occurred as a result of control measures – however this latter data remains sparse and is available for only a limited number of settings(Jarvis et al., 2021, 2020; Liu et al., 2021). Description of contact location was also coarse and precluded more granular analyses of specific settings, such as markets, which have previously been shown to be important locations for transmission in rural areas(Grijalva et al., 2015).

Heterogeneity between studies was larger for LICs/LMICs and UMICs, which we partly accounted for, through fitting random study effects. These study differences may be attributed to the way individual contact surveys were conducted, making comparisons of contact patterns among surveys more difficult (e.g. prospective/retrospective diary surveys, online/paper questionnaires, face-to-face/phone interviews, and different contact definitions). For instance, there is evidence suggesting that prospective reporting, which is less affected by recall bias, can often lead to a higher number of contacts being reported(Mikolajczyk and Kretzschmar, 2008) and a lower probability of casual or short-lasting contacts being missed. The relatively high contact rates observed in HICs may be explained by the fact that all but two HIC surveys used diary methods. Our study highlights that a unified definition of "contact" and standard practice in data collection could help increase the quality of collected data, leading to more robust and reliable conclusions about contact patterns. Whilst we aggregate results by income strata due to the limited availability of data (particularly in lower- and middle-income countries), it is important to note that the outcomes considered here are likely to be shaped by several different factors other than country-level income. Whilst some of these factors will be correlated with a country's income status (e.g. household size(Walker et al., 2020)), many others will be unique to a particular setting or geographical area or correlate only weakly with country-level data. Examples include patterns of employment, the role of women, and other contextual factors. These analyses are therefore intended primarily to provide indications of prevailing patterns, rather than a definitive description of contact patterns in a specific context and highlight the significant need for further studies to by carried out in a diversity of different locations.

Despite these limitations however, our results highlight significant differences in the structure and nature of contact patterns across settings. These differences suggest that the comparative importance of different locations and age-groups to transmission will likely vary across settings and have critical consequences for the efficacy and suitability of strategies aimed at controlling the spread of respiratory pathogens such as SARS-CoV-2. Most importantly, our study highlights the limited amount of work that has been undertaken to date to better understand and quantify patterns of contact across a range of settings, particularly in lower- and middle-income countries, which is vital in informing control strategies reducing the spread of such pathogens.

Ethics statement

All original studies included were approved by an institutional ethics review committee. Ethics approval was not required for the present study.

Acknowledgements

We would like to acknowledge the Fiji Ministry of Health and Medical Services for their contribution to the study set in Fiji, M. Elizabeth Halloran for sharing the Senegal data, and Nickson Murunga for processing the data request for the Kenyan survey.

Competing interests

M.A. has received research funding from Seqirus outside the submitted work. G.E.P. was employed by the Emmes Company while analyzing the Niakhar Senegal social contact network data included in this study. The Emmes Company was contracted to perform data cleaning and data analysis of the Niakhar, Senegal clinical trial data (but not the social contact network data) for this study before G.E.P. joined the Emmes Company (in October 2015). After G.E.P. joined the Emmes Company, the sole support from Emmes for this manuscript was in the form of salary support for G.E.P. All other authors declare no conflicts of interest. Outside of the submitted work C.G.G. has received grants, contracts, or consulting fees from the following bodies: CDC, AHRQ, FDA, NIH, Campbell Alliance/Syneos Health, Sanofi, Pfizer and Merck.

Financial disclosures

A.M., P.W., P.G.T.W. and C.W acknowledge joint Centre funding from the UK Medical Research Council and DFID (MR/R015600/1). O.J.W. acknowledges funding from the UK Foreign Commonwealth and Development Office. K.O.K acknowledges support by CUHK Direct grant for research (2019.020), Health and Medical Research Fund (reference number: INF-CUHK-1, 17160302, 18170312), General Research Fund (reference number: 14112818), Early Career Scheme (reference number: 24104920) and Wellcome Trust (UK, 200861/Z/16/Z). P.J.D. was supported by a fellowship from the UK Medical Research Council (MR/P022081/1); this UK-funded award is part of the European and Developing Countries Clinical Trials Partnership 2 (EDCTP2) programme supported by the EU. E.F.G.N. holds an Australian Government Research Training Program Scholarship. F.M.R. receives funding from the Australian National Health and Medical Research Council, WHO, the Bill & Melinda Gates Foundation; Wellcome Trust, DFAT. M.M. acknowledges funding from the EPSRC through the EPSRC Centre for Doctoral Training in Modern Statistics and Statistical Machine Learning. J.D.S received funding for this work from the University of Washington and a grant from US

National Institutes of Health, NIAID. C.G.G. declares funding from NIH (K24AI148459). G.E.P. was supported previously by General Medical Sciences / National Institute of Health U01-GM070749. G.E.P was employed by the Emmes Company while analyzing the Niakhar Senegal social contact network data included in this study. The Emmes Company was contracted to perform data cleaning and data analysis of the Niakhar, Senegal clinical trial data (but not the social contact network data) for this study before G.E.P. joined the Emmes Company (in October 2015). After G.E.P. joined the Emmes Company, the sole support from Emmes for this manuscript was in the form of salary support for G.E.P. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

Ajelli M, Litvinova M. 2017. Estimating contact patterns relevant to the spread of infectious diseases in Russia. *J Theor Biol* **419**:1–7. doi:10.1016/j.jtbi.2017.01.041

Bansal S, Read J, Pourbohloul B, Meyers LA. 2010. The dynamic nature of contact networks in infectious disease epidemiology. *http://mc.manuscriptcentral.com/tjbd* **4**:478–489. doi:10.1080/17513758.2010.503376

Béraud G, Kazmercziak S, Beutels P, Levy-Bruhl D, Lenne X, Mielcarek N, Yazdanpanah Y, Boëlle P-Y, Hens N, Dervaux B. 2015. The French Connection: The First Large Population-Based Contact Survey in France Relevant for the Spread of Infectious Diseases. *PLoS One* **10**:e0133203. doi:10.1371/journal.pone.0133203

Brauner JM, Mindermann S, Sharma M, Johnston D, Salvatier J, Gavenčiak T, Stephenson AB, Leech G, Altman G, Mikulik V, Norman AJ, Monrad JT, Besiroglu T, Ge H, Hartwick MA, Teh YW, Chindelevitch L, Gal Y, Kulveit J. 2021. Inferring the effectiveness of government interventions against COVID-19. *Science (80-)* **371**. doi:10.1126/SCIENCE.ABD9338

Bulfone TC, Malekinejad M, Rutherford GW, Razani N. 2021. Outdoor Transmission of SARS-CoV-2 and Other Respiratory Viruses: A Systematic Review. *J Infect Dis* **223**:550–561. doi:10.1093/infdis/jiaa742

Bürkner PC. 2018. Advanced Bayesian multilevel modeling with the R package brms. *R J* **10**:395–411. doi:10.32614/rj-2018-017

Bürkner PC. 2017. brms: An R package for Bayesian multilevel models using Stan. *J Stat Softw* **80**:1–28. doi:10.18637/jss.v080.i01

Chang S, Pierson E, Koh PW, Gerardin J, Redbird B, Grusky D, Leskovec J. 2021. Mobility network models of COVID-19 explain inequities and inform reopening. *Nature* **589**:82–87. doi:10.1038/s41586-020-2923-3

Chu DK, Akl EA, Duda S, Solo K, Yaacoub S, Schünemann HJ, El-harakeh A, Bognanni A, Lotfi T, Loeb M, Hajizadeh A, Bak A, Izcovich A, Cuello-Garcia CA, Chen C, Harris DJ, Borowiack E, Chamseddine F, Schünemann F, Morgano GP, Muti Schünemann GEU, Chen G, Zhao H, Neumann I, Chan J, Khabsa J, Hneiny L, Harrison L, Smith M, Rizk N, Giorgi Rossi P, AbiHanna P, El-khoury R, Stalteri R, Baldeh T, Piggott T, Zhang Y, Saad Z, Khamis A, Reinap M. 2020. Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. *Lancet* **395**:1973–1987. doi:10.1016/S0140-6736(20)31142-9

Dahab M, van Zandvoort K, Flasche S, Warsame A, Ratnayake R, Favas C, Spiegel PB, Waldman RJ, Checchi F. 2020. COVID-19 control in low-income settings and displaced populations: what can realistically be done? *Confl Heal 2020 141* **14**:1–6. doi:10.1186/S13031-020-00296-8

Davies NG, Kucharski AJ, Eggo RM, Gimma A, Edmunds WJ, Jombart T, O'Reilly K, Endo A, Hellewell J, Nightingale ES, Quilty BJ, Jarvis CI, Russell TW, Klepac P, Bosse NI, Funk S, Abbott S, Medley GF, Gibbs H, Pearson CAB, Flasche S, Jit M, Clifford S, Prem K, Diamond C, Emery J, Deol AK, Procter

SR, van Zandvoort K, Sun YF, Munday JD, Rosello A, Auzenbergs M, Knight G, Houben RMGJ, Liu Y. 2020. Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital services in the UK: a modelling study. *Lancet Public Heal* **5**:e375–e385. doi:10.1016/S2468-2667(20)30133-X

De Negri F, Galiezz R, Miranda P, Koeller P, Zucoloto G, Costa J, Farias CM, Travassos GH, Medronho RA. 2021. Socioeconomic factors and the probability of death by Covid-19 in Brazil. *J Public Health (Bangkok)* 1–6. doi:10.1093/pubmed/fdaa279

Delamater PL, Street EJ, Leslie TF, Yang YT, Jacobsen KH. 2019. Complexity of the basic reproduction number (R0). *Emerg Infect Dis* **25**:1–4. doi:10.3201/eid2501.171901

Dodd PJ, Looker C, Plumb ID, Bond V, Schaap A, Shanaube K, Muyoyeta M, Vynnycky E, Godfrey-Faussett P, Corbett EL, Beyers N, Ayles H, White RG. 2015. Age- and Sex-Specific Social Contact Patterns and Incidence of *Mycobacterium tuberculosis* Infection. *Am J Epidemiol* **183**:kwv160. doi:10.1093/aje/kwv160

Downes MJ, Brennan ML, Williams HC, Dean RS. 2016. Development of a critical appraisal tool to assess the quality of cross-sectional studies (AXIS). *BMJ Open* **6**. doi:10.1136/bmjopen-2016

Dunne EM, Satzke C, Ratu FT, Neal EFG, Boelsen LK, Matanitobua S, Pell CL, Nation ML, Ortika BD, Reyburn R, Jenkins K, Nguyen C, Gould K, Hinds J, Tikoduadua L, Kado J, Rafai E, Kama M, Mulholland EK, Russell FM. 2018. Effect of ten-valent pneumococcal conjugate vaccine introduction on pneumococcal carriage in Fiji: results from four annual cross-sectional carriage surveys. *Lancet Glob Heal* **6**:e1375–e1385. doi:10.1016/S2214-109X(18)30383-8

Eubank S, Guclu H, Kumar VSA, Marathe M V., Srinivasan A, Toroczkai Z, Wang N. 2004. Modelling disease outbreaks in realistic urban social networks. *Nature* **429**:180–184. doi:10.1038/nature02541

Ferrari MJ, Bansal S, Meyers LA, Björnstad ON. 2006. Network frailty and the geometry of herd immunity. *Proc R Soc B Biol Sci* **273**:2743–2748. doi:10.1098/rspb.2006.3636

Firth JA, Hellewell J, Klepac P, Kissler S, Jit M, Atkins KE, Clifford S, Villabona-Arenas CJ, Meakin SR, Diamond C, Bosse NI, Munday JD, Prem K, Foss AM, Nightingale ES, Zandvoort K van, Davies NG, Gibbs HP, Medley G, Gimma A, Flasche S, Simons D, Auzenbergs M, Russell TW, Quilty BJ, Rees EM, Leclerc QJ, Edmunds WJ, Funk S, Houben RMGJ, Knight GM, Abbott S, Sun FY, Lowe R, Tully DC, Procter SR, Jarvis CI, Endo A, O'Reilly K, Emery JC, Jombart T, Rosello A, Deol AK, Quaife M, Hué S, Liu Y, Eggo RM, Pearson CAB, Kucharski AJ, Spurgin LG. 2020. Using a real-world network to model localized COVID-19 control strategies. *Nat Med* **26**:1616–1622. doi:10.1038/s41591-020-1036-8

Fontanet A, Cauchemez S. 2020. COVID-19 herd immunity: where are we? *Nat Rev Immunol*. doi:10.1038/s41577-020-00451-5

Fu Y, Wang D-W, Chuang J-H. 2012. Representative Contact Diaries for Modeling the Spread of Infectious Diseases in Taiwan. *PLoS One* **7**:e45113. doi:10.1371/journal.pone.0045113

Ghani AC, Aral SO. 2005. Patterns of sex worker - Client contacts and their implications for the persistence of sexually transmitted infections. *J Infect Dis* **191**:S34–S41. doi:10.1086/425276

Gimma A, Munday JD, Wong KL, Coletti P, Zandvoort K van, Prem K, group CC-19 working, Klepac P, Rubin GJ, Funk S, Edmunds WJ, Jarvis CI. 2021. CoMix: Changes in social contacts as measured by the contact survey during the COVID-19 pandemic in England between March 2020 and March 2021. *medRxiv* 2021.05.28.21257973. doi:10.1101/2021.05.28.21257973

Grijalva CG, Goeyvaerts N, Verastegui H, Edwards KM, Gil AI, Lanata CF, Hens N. 2015. A Household-Based Study of Contact Networks Relevant for the Spread of Infectious Diseases in the Highlands of Peru. *PLoS One* **10**:e0118457. doi:10.1371/journal.pone.0118457

Hoang T, Coletti P, Melegaro A, Wallinga J, Grijalva CG, Edmunds JW, Beutels P, Hens N. 2019. A Systematic Review of Social Contact Surveys to Inform Transmission Models of Close-contact

Infections. *Epidemiology* **30**:723–736. doi:10.1097/EDE.0000000000001047

Horton KC, Hoey AL, Béraud G, Corbett EL, White RG. 2020. Systematic review and meta-analysis of sex differences in social contact patterns and implications for tuberculosis transmission and control. *Emerg Infect Dis*. doi:10.3201/eid2605.190574

Huang Y, Cai X, Zhang B, Zhu G, Liu T, Guo P, Xiao J, Li X, Zeng W, Hu J, Ma W. 2020. Spatiotemporal heterogeneity of social contact patterns related to infectious diseases in the Guangdong Province, China. *Sci Rep* **10**:1–10. doi:10.1038/s41598-020-63383-z

Ibuka Y, Ohkusa Y, Sugawara T, Chapman GB, Yamin D, Atkins KE, Taniguchi K, Okabe N, Galvani AP. 2016. Social contacts, vaccination decisions and influenza in Japan. *J Epidemiol Community Health* **70**:162–167. doi:10.1136/jech-2015-205777

Jacquez JA, Simon CP, Koopman J, Sattenspiel L, Perry T. 1988. Modeling and analyzing HIV transmission: the effect of contact patterns. *Math Biosci* **92**:119–199. doi:10.1016/0025-5564(88)90031-4

Jarvis CI, Gimma A, van Zandvoort K, Wong KLM, Abbas K, Villabona-Arenas CJ, O'Reilly K, Quaife M, Rosello A, Kucharski AJ, Gibbs HP, Atkins KE, Barnard RC, Bosse NI, Procter SR, Meakin SR, Sun FY, Abbott S, Munday JD, Russell TW, Flasche S, Sherratt K, Eggo RM, Davies NG, Quilty BJ, Auzenbergs M, Hellewell J, Jombart T, Jafari Y, Leclerc QJ, Lowe R, Foss AM, Jit M, Deol AK, Hué S, Knight GM, Endo A, Prem K, Emery JC, Clifford S, Medley G, Funk S, Sandmann FG, Tully DC, Pearson CAB, Gore-Langton GR, Showering A, Houben RMGJ, Nightingale ES, Klepac P, Waterlow NR, Chan YWD, Rudge JW, Simons D, Diamond C, Williams J, Brady O, Liu Y, Edmunds WJ. 2021. The impact of local and national restrictions in response to COVID-19 on social contacts in England: a longitudinal natural experiment. *BMC Med* **19**:1–12. doi:10.1186/s12916-021-01924-7

Jarvis CI, Van Zandvoort K, Gimma A, Prem K, Auzenbergs M, O'Reilly K, Medley G, Emery JC, Houben

RMGJ, Davies N, Nightingale ES, Flasche S, Jombart T, Hellewell J, Abbott S, Munday JD, Bosse NI, Funk S, Sun F, Endo A, Rosello A, Procter SR, Kucharski AJ, Russell TW, Knight G, Gibbs H, Leclerc Q, Quilty BJ, Diamond C, Liu Y, Jit M, Clifford S, Pearson CAB, Eggo RM, Deol AK, Klepac P, Rubin GJ, Edmunds WJ. 2020. Quantifying the impact of physical distance measures on the transmission of COVID-19 in the UK. *BMC Med* **18**:1–10. doi:10.1186/s12916-020-01597-8

Keeling MJ, Eames KT. 2005. Networks and epidemic models. *J R Soc Interface* **2**:295–307. doi:10.1098/RSIF.2005.0051

Khan K, Arino J, Hu W, Raposo P, Sears J, Calderon F, Heidebrecht C, Macdonald M, Liauw J, Chan A, Gardam M. 2009. Spread of a Novel Influenza A (H1N1) Virus via Global Airline Transportation. *N Engl J Med* **361**:212–214. doi:10.1056/nejmc0904559

Kiti MC, Kinyanjui TM, Koech DC, Munywoki PK, Medley GF, Nokes DJ. 2014. Quantifying Age-Related Rates of Social Contact Using Diaries in a Rural Coastal Population of Kenya. *PLoS One* **9**:e104786. doi:10.1371/journal.pone.0104786

Kumar S, Gosain M, Sharma H, Swetts E, Amarchand R, Kumar R, Lafond KE, Dawood FS, Jain S, Widdowson M-A, Read JM, Krishnan A. 2018. Who interacts with whom? Social mixing insights from a rural population in India. *PLoS One* **13**:e0209039. doi:10.1371/journal.pone.0209039

Kwok KO, Cowling B, Wei V, Riley S, Read JM. 2018. Temporal variation of human encounters and the number of locations in which they occur: A longitudinal study of Hong Kong residents. *J R Soc Interface* **15**. doi:10.1098/rsif.2017.0838

Kwok KO, Cowling BJ, Wei VWI, Wu KM, Read JM, Lessler J, Cummings DA, Malik Peiris JS, Riley S. 2014. Social contacts and the locations in which they occur as risk factors for influenza infection. *Proc R Soc B Biol Sci* **281**. doi:10.1098/rspb.2014.0709

Lapidus N, de Lamballerie X, Salez N, Setbon M, Delabre RM, Ferrari P, Moyen N, Gougeon M-L, Vely F, Leruez-Ville M, Andreoletti L, Cauchemez S, Boëlle P-Y, Vivier É, Abel L, Schwarzinger M, Legeas M, Le Cann P, Flahault A, Carrat F. 2013. Factors Associated with Post-Seasonal Serological Titer and Risk Factors for Infection with the Pandemic A/H1N1 Virus in the French General Population. *PLoS One* **8**:e60127. doi:10.1371/journal.pone.0060127

le Polain de Waroux O., Cohuet S, Ndazima D, Kucharski AJ, Juan-Giner A, Flasche S, Tumwesigye E, Arinaitwe R, Mwanga-Amumpaire J, Boum Y, Nackers F, Checchi F, Grais RF, Edmunds WJ. 2018. Characteristics of human encounters and social mixing patterns relevant to infectious diseases spread by close contact: A survey in Southwest Uganda. *BMC Infect Dis* **18**:172. doi:10.1186/s12879-018-3073-1

le Polain de Waroux Olivier, Flasche S, Kucharski AJ, Langendorf C, Ndazima D, Mwanga-Amumpaire J, Grais RF, Cohuet S, Edmunds WJ. 2018. Identifying human encounters that shape the transmission of Streptococcus pneumoniae and other acute respiratory infections. *Epidemics* **25**:72–79. doi:10.1016/j.epidem.2018.05.008

Leung K, Jit M, Lau EHY, Wu JT. 2017. Social contact patterns relevant to the spread of respiratory infectious diseases in Hong Kong. *Sci Rep* **7**:4–8. doi:10.1038/s41598-017-08241-1

Liu CY, Berlin J, Kiti MC, Fava E Del, Grow A, Zagheni E, Melegaro A, Jenness SM, Omer S, Lopman B, Nelson K. 2021. Rapid review of social contact patterns during the COVID-19 pandemic. *medRxiv* 2021.03.12.21253410. doi:10.1101/2021.03.12.21253410

Loayza N V. 2020. Costs and Trade-Offs in the Fight Against the COVID-19 Pandemic, Costs and Trade-Offs in the Fight Against the COVID-19 Pandemic. World Bank, Washington, DC. doi:10.1596/33764

Mahikul W, Kripattanapong S, Hanvoravongchai P, Meeyai A, Iamsirithaworn S, Auewarakul P, Pan-ngum W. 2020. Contact Mixing Patterns and Population Movement among Migrant Workers in an Urban Setting in Thailand. *Int J Environ Res Public Health* **17**:2237. doi:10.3390/ijerph17072237

Mbala-Kingebeni P, Aziza A, Di Paola N, Wiley MR, Makiala-Mandanda S, Caviness K, Pratt CB, Ladner JT, Kugelman JR, Prieto K, Chitty JA, Larson PA, Beitzel B, Ayouba A, Vidal N, Karhemere S, Diop M, Diagne MM, Faye M, Faye O, Aruna A, Nsio J, Mulangu F, Mukadi D, Mukadi P, Kombe J, Mulumba A, Villabona-Arenas CJ, Pukuta E, Gonzalez J, Bartlett ML, Sozhamannan S, Gross SM, Schroth GP, Tim R, Zhao JJ, Kuhn JH, Diallo B, Yao M, Fall IS, Ndjoloko B, Mossoko M, Lacroix A, Delaporte E, Sanchez-Lockhart M, Sall AA, Muyembe-Tamfum JJ, Peeters M, Palacios G, Ahuka-Mundeke S. 2019. Medical countermeasures during the 2018 Ebola virus disease outbreak in the North Kivu and Ituri Provinces of the Democratic Republic of the Congo: a rapid genomic assessment. *Lancet Infect Dis* **19**:648–657. doi:10.1016/S1473-3099(19)30118-5

McCreesh N, Dlamini V, Edwards A, Olivier S, Dayi N, Dikgale K, Nxumalo S, Dreyer J, Baisley K, Siedner MJ, White RG, Herbst K, Grant AD, Harling G. 2021. Impact of the Covid-19 epidemic and related social distancing regulations on social contact and SARS-CoV-2 transmission potential in rural South Africa: analysis of repeated cross-sectional surveys. *BMC Infect Dis 2021 211* **21**:1–11. doi:10.1186/S12879-021-06604-8

Meeyai A, Praditsitthikorn N, Kotirum S, Kulpeng W, Putthasri W, Cooper BS, Teerawattananon Y. 2015. Seasonal Influenza Vaccination for Children in Thailand: A Cost-Effectiveness Analysis. PLOS Med 12:e1001829. doi:10.1371/journal.pmed.1001829

Mikolajczyk RT, Kretzschmar M. 2008. Collecting social contact data in the context of disease transmission: Prospective and retrospective study designs. *Soc Networks* **30**:127–135. doi:10.1016/j.socnet.2007.09.002

Mistry D, Litvinova M, Pastore y Piontti A, Chinazzi M, Fumanelli L, Gomes MFC, Haque SA, Liu QH, Mu K, Xiong X, Halloran ME, Longini IM, Merler S, Ajelli M, Vespignani A. 2021. Inferring high-resolution human mixing patterns for disease modeling. *Nat Commun* **12**:1–12. doi:10.1038/s41467-020-20544-y

Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, Massari M, Salmaso S, Tomba GS, Wallinga J, Heijne J, Sadkowska-Todys M, Rosinska M, Edmunds WJ. 2008. Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases. *PLoS Med* **5**:e74. doi:10.1371/journal.pmed.0050074

Neal EFG, Flasche S, Nguyen CD, Ratu FT, Dunne EM, Koyamaibole L, Reyburn R, Rafai E, Kama M, Ortika BD, Boelsen LK, Kado J, Tikoduadua L, Devi R, Tuivaga E, Satzke C, Mulholland EK, Edmunds WJ, Russell FM. 2020. Associations between ethnicity, social contact, and pneumococcal carriage three years post-PCV10 in Fiji. *Vaccine* **38**:202–211. doi:10.1016/j.vaccine.2019.10.030

Oguz MM, Camurdan AD, Aksakal FN, Akcaboy M, Altinel Acoglu E. 2018. Social contact patterns of infants in deciding vaccination strategy: A prospective, cross-sectional, single-centre study. *Epidemiol Infect* **146**:1157–1166. doi:10.1017/S0950268818001048

Potter GE, Wong J, Sugimoto J, Diallo A, Victor JC, Neuzil K, Halloran ME. 2019. Networks of face-to-face social contacts in Niakhar, Senegal. PLoS One 14:e0220443. doi:10.1371/journal.pone.0220443

Prem K, Cook AR, Jit M. 2017. Projecting social contact matrices in 152 countries using contact surveys and demographic data. *PLoS Comput Biol* **13**:e1005697. doi:10.1371/journal.pcbi.1005697

Prem K, Liu Y, Russell TW, Kucharski AJ, Eggo RM, Davies N, Flasche S, Clifford S, Pearson CAB, Munday JD, Abbott S, Gibbs H, Rosello A, Quilty BJ, Jombart T, Sun F, Diamond C, Gimma A, van Zandvoort K, Funk S, Jarvis CI, Edmunds WJ, Bosse NI, Hellewell J, Jit M, Klepac P. 2020. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. *Lancet Public Heal* **5**:e261–e270. doi:10.1016/S2468-2667(20)30073-6

Rader B, Scarpino S V., Nande A, Hill AL, Adlam B, Reiner RC, Pigott DM, Gutierrez B, Zarebski AE, Shrestha M, Brownstein JS, Castro MC, Dye C, Tian H, Pybus OG, Kraemer MUG. 2020. Crowding and the shape of COVID-19 epidemics. *Nat Med 2020 2612* **26**:1829–1834. doi:10.1038/s41591-020-1104-0

Read JM, Lessler J, Riley S, Wang S, Tan LJ, Kwok KO, Guan Y, Jiang CQ, Cummings DAT. 2014. Social mixing patterns in rural and urban areas of Southern China. Proc R Soc B Biol Sci 281. doi:10.1098/rspb.2014.0268

Routledge I, Epstein A, Takahashi S, Hakim J, Janson O, Duarte E, Turcios K, Vinden J, Sujishi K, Rangel J, Coh M, Besana L, Ho W-K, Oon C-Y, Ong CM, Yun C, Lynch K, Wu A, Wu W, Karlon W, Thornborrow E, Peluso M, Henrich T, Pak J, Briggs J, Greenhouse B, Rodriguez-Barraquer I. 2021. Citywide serosurveillance of the initial SARS-CoV-2 outbreak in San Francisco. *Res Sq*. doi:10.21203/rs.3.rs-180966/v1

Stein ML, van Steenbergen JE, Buskens V, van der Heijden PGM, Chanyasanha C, Tipayamongkholgul M, Thorson AE, Bengtsson L, Lu X, Kretzschmar MEE. 2014. Comparison of Contact Patterns Relevant for Transmission of Respiratory Pathogens in Thailand and the Netherlands Using Respondent-Driven Sampling. *PLoS One* **9**:e113711. doi:10.1371/journal.pone.0113711

Thompson HA, Mousa A, Dighe A, Fu H, Arnedo-Pena A, Barrett P, Bellido-Blasco J, Bi Q, Caputi A, Chaw L, De Maria L, Hoffmann M, Mahapure K, Ng K, Raghuram J, Singh G, Soman B, Soriano V, Valent F, Vimercati L, Wee LE, Wong J, Ghani AC, Ferguson NM. 2021. SARS-CoV-2 setting-specific transmission rates: a systematic review and meta-analysis. *Clin Infect Dis*. doi:10.1093/cid/ciab100

United Nations Department of Economic and Social Affairs Population Division. 2019. Database on Household Size and Composition 2019. https://population.un.org/Household/index.html#/countries/840

Walker PGT, Whittaker C, Watson OJ, Baguelin M, Winskill P, Hamlet A, Djafaara BA, Cucunubá Z, Mesa DO, Green W, Thompson H, Nayagam S, Ainslie KEC, Bhatia S, Bhatt S, Boonyasiri A, Boyd O, Brazeau NF, Cattarino L, Cuomo-Dannenburg G, Dighe A, Donnelly CA, Dorigatti I, Van Elsland SL, FitzJohn R, Fu H, Gaythorpe KAM, Geidelberg L, Grassly N, Haw D, Hayes S, Hinsley W, Imai N, Jorgensen D, Knock E, Laydon D, Mishra S, Nedjati-Gilani G, Okell LC, Unwin HJ, Verity R, Vollmer M, Walters CE, Wang H, Wang Y, Xi X, Lalloo DG, Ferguson NM, Ghani AC. 2020. The impact of COVID-19 and strategies for mitigation and suppression in low- And middle-income countries. *Science (80-)* **369**:413–422. doi:10.1126/science.abc0035

Ward H, Cooke G, Whitaker M, Redd R, Eales O, Brown JC, Collet K, Cooper E, Daunt A, Jones K, Moshe M, Willicombe M, Day S, Atchison C, Darzi A, Donnelly CA, Riley S, Ashby D, Barclay WS, Elliott P. 2021. REACT-2 Round 5: increasing prevalence of SARS-CoV-2 antibodies demonstrate impact of the second wave and of vaccine roll-out in England. *medRxiv* 2021.02.26.21252512. doi:10.1101/2021.02.26.21252512

Watson CH, Coriakula J, Ngoc DTT, Flasche S, Kucharski AJ, Lau CL, Thieu NTV, le Polain de Waroux O, Rawalai K, Van TT, Taufa M, Baker S, Nilles EJ, Kama M, Edmunds WJ. 2017. Social mixing in Fiji: Who-eats-with-whom contact patterns and the implications of age and ethnic heterogeneity for disease dynamics in the Pacific Islands. PLoS One 12:e0186911. doi:10.1371/journal.pone.0186911

Winskill P, Whittaker C, Walker P, Watson O, Laydon D, Imai N, Cuomo-Dannenburg G, Ainslie K, Baguelin M, Bhatt S, Boonyasiri A, Cattarino L, Ciavarella C, Cooper L V, Coupland H, Cucunuba Z, Van Elsland SL, Fitzjohn R, Flaxman S, Gaythorpe K, Green W, Hallett T, Hamlet A, Hinsley W, Knock E, Lees J, Mellan T, Mishra S, Nedjati-Gilani G, Nouvellet P, Okell L, Parag K V, Thompson HA, Juliette H, Unwin T, Vollmer M, Wang Y, Whittles L, Xi X, Ferguson N, Donnelly C, Ghani A. 2020. Report 22: Equity in response to the COVID-19 pandemic: an assessment of the direct and indirect impacts on disadvantaged and vulnerable populations in low-and lower middle-income countries. doi:10.25561/78965

Wood R, Racow K, Bekker L-G, Morrow C, Middelkoop K, Mark D, Lawn SD. 2012. Indoor Social Networks in a South African Township: Potential Contribution of Location to Tuberculosis Transmission. *PLoS One* **7**:e39246. doi:10.1371/journal.pone.0039246

World Bank Group. 2020. World Bank Country and Lending Groups – World Bank Data Help Desk. https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups

World Bank International Comparison Programme. 2021. World Development Indicators database Eurostat-OECD PPP Programme.

Zhang J, Litvinova M, Liang Y, Wang Y, Wang W, Zhao S, Wu Q, Merler S, Viboud C, Vespignani A, Ajelli M, Yu H. 2020. Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. *Science (80-)* **368**:1481–1486. doi:10.1126/science.abb8001

Table 1- Summary table of total daily contacts. The total number of observations, as well as the mean, median and interquartile range (p25 and p75) of total daily contacts shown by participant and study characteristics.

			N	Mean	p25	Median	p75
Overall			28,503	14.5	5	9	17
Gender		**Male**	13,218	15.3	5	9	18
		Female	14,598	13.7	5	9	16
Age		**<15**	8,561	14.6	6	10	19
		15 to 65	17,841	14.9	5	9	17
		>65	2,047	10.4	3	6	12
Income status		**LIC/LMIC**	9,906	15.4	5	10	17
		UMIC	8,330	14.4	5	8	16
		HIC	10,267	13.7	5	9	17
Survey Methodology		**Diary**	12,226	13.9	6	10	18
		Interview/Survey	16,227	15.0	4	8	16
Day type		**Weekend**	4,308	14.7	5	9	16
		Weekday	21,579	14.1	5	9	17
Employment		**Yes**	8,879	15.4	5	9	17
(in those aged >18)		**No**	6,158	9.8	4	7	12
Student		**Yes**	4,438	18.4	8	14	24
(in those aged 5 to 18)		**No**	600	10.4	5	8	14
Household size		**1**	1,479	10.4	3	6	12
		2	3,220	11.8	4	7	14
		3	4,130	12.0	4	7	14
		4	5,240	13.4	5	8	17
		5	3,109	12.5	4	8	14
		6+	8,873	17.7	7	11	20
Study	**Belgium**	**Mossong**	750	11.8	5	9	15
	China	**Read**	1,821	18.6	7	13	22
	China	**Zhang**	965	18.8	4	10	30
	Fiji	**Neal**	2,019	6.4	4	6	8
	Finland	**Mossong**	1,006	11.1	5	9	15
	Germany	**Mossong**	1,341	7.9	4	6	10
	Hong Kong	**Kwok (2014)**	762	18.3	5	9	18
	Hong Kong	**Kwok (2018)**	1,066	11.9	3	7	13
	Hong Kong	**Leung**	1,149	14.4	3	7	15
	India	**Kumar**	2,943	27.0	12	17	26
	Italy	**Mossong**	849	19.8	10	17	27
	Kenya	**Kiti**	568	17.7	10	15	23
	Luxembourg	**Mossong**	1,051	17.5	8	14	24
	Netherlands	**Mossong**	269	13.9	6	11	19
	Peru	**Grijalva**	588	15.3	8	12	20
	Poland	**Mossong**	1,012	16.3	7	13	22.5
	Russia	**Ajelli**	502	18.0	6	11	19
	South Africa	**Dodd**	1,276	5.2	4	5	7
	South Africa	**Wood**	571	15.6	9	14	20
	Senegal	**Potter**	1,417	19.7	10	15	25
	Thailand	**Mahikul**	369	22.6	13	20	31
	Thailand	**Stein**	219	58.5	15	24	55
	Uganda	**Le Polain de Waroux**	568	7.0	5	7	9
	United	**Mossong**	1,012	11.7	6	10	16
	Vietnam	**Horby**	865	7.7	5	7	9
	Zambia	**Dodd**	2,300	4.8	3	4	6
	Zimbabwe	**Melegaro**	1,245	10.7	6	9	14

Figure 1 – Total number of contacts. Sample median total number of contacts shown by gender (right) and 5-year age groups up to ages 80+ shown for A) LICs/LMICs, B) UMICs and C) HICs. Grey lines denote individual studies, and the solid black line is the median across all studies of within that income group. Studies with a diary-based methodology are represented by a solid grey line and those with a questionnaire or interview design are shown as a dashed line. For UMICs, one study outlier with extremely high number of contacts is excluded (online Thai survey with a "snowball" design by Stein et al., 2014). Contact Rate Ratios and associated 95% Credible intervals from a negative binomial model with random study effects are shown in D (LICs/LMICs), E (UMICs) and F (HICs). All models were adjusted for age and gender and were ran separately for each key variable (weekday/weekend, household size, survey methodology, student/employment status).

Figure 2- Contact location and household size. A) Sample median number of contacts by household size in review data, stratified by income strata. Shaded area denotes the interquartile range. B) sample mean % of contacts made at each location (home, school, work, other) by income group. C) total daily contacts (sample mean number) made at each location by 5-year age group. D) Sample median number of contacts made at home by 5-year age groups and income strata. Shaded area denotes the interquartile range. E) Average household size and GDP; red circles represent median household size in single studies from the review. GDP information was obtained from the World Bank Group and global household size data from the Department of Economic and Social Affairs, Population Division, United Nations.

Figure 3- Physical contacts. Mean proportion of contacts that are physical shown by gender (right) and 5-year age groups up to ages 80+ shown for A) LICs/LMICs, B) UMICs and C) HICs. Grey lines denote individual studies, and the solid black line is the mean across all studies of within that income group. Studies with a diary-based methodology are represented by a solid grey line and those with a questionnaire or interview design are shown as a dashed line. Odds Ratios and associated 95% Credible intervals from a logistic regression model with random study effects are shown in D (LICs/LMICs), E (UMICs) and F (HICs). All models were adjusted for age and gender and were ran separately for each key variable (weekday/weekend, household size, survey methodology, student/employment status).

Figure 4- Contact duration. Mean proportion of contacts that last at least an hour shown by gender (right) and 5-year age groups up to ages 80+ shown for A) LICs/LMICs, B) UMICs and C) HICs. Grey lines denote individual studies and the solid black line is the mean across all studies of within that income group. Studies with a diary-based methodology are represented by a solid grey line and those with a questionnaire or interview design are shown as a dashed line. Odds Ratios and associated 95% Credible intervals from a logistic regression model with random study effects are shown in D (LICs/LMICs), E (UMICs) and F (HICs). All models were adjusted for age and gender and were ran separately for each key variable (weekday/weekend, household size, survey methodology, student/employment status).

Appendices

Appendix 1

Appendix 2

APPENDIX 1

Systematic review methods and additional information

The search string used to identify eligible studies is shown in Appendix 1 – Table 1. All search terms were searched in all fields.

Appendix 1 - Table 1: Search string

1	(social contact* or contact pattern* or contact network* or contact mixing or contact survey* or contact data or contact rate* or contact matri* or contact parameter* or physical contact* or social mixing or mixing behavio* or mixing pattern* or mixing matri* or assortative mixing or disassortative mixing or mixing parameter*).af.
2	(**LIC** or **low-income*** or **low income*** or **low- income*** or **developing countr*** or Afghanistan or Guinea-Bissau or Sierra Leone or Benin or Haiti or Somalia or Burkina Faso or Korea or Sudan or Burundi or Liberia or Syria* or Central African Republic or CAR or Madagascar or Tajikistan or Chad or Malawi or Tanzania or Congo or DRC or Mali or Togo or Eritrea or Mozambique or Uganda or Ethiopia or Nepal or Yemen or Gambia or Niger or Guinea or Rwanda or LMIC or Angola or India or Papua New Guinea or Bangladesh or Indonesia or Philippines or Bhutan or Kenya or (Sao Tome and Principe) or Bolivia or Kiribati or Senegal or Cabo Verde or Kyrgyz Republic or Solomon Islands or Cambodia or Laos or Cameroon or Lesotho or Timor-Leste or Comoros or Mauritania or Tunisia or Micronesia or Ukraine or Cote d'Ivoire or Moldova or Uzbekistan or Djibouti or Mongolia or Vanuatu or Egypt or Morocco or Vietnam or El Salvador or Myanmar or “West Bank and Gaza” or Eswatini or Nicaragua or Zambia or Ghana or Nigeria or Zimbabwe or Honduras or Pakistan or **middle income*** or **middle-income*** or **middle- income** or **UMIC** or Albania or Algeria or American Samoa or Argentina or Armenia or Azerbaijan or Belarus or Belize or (Bosnia and Herzegovina) or Botswana or Brazil or Bulgaria or China or Colombia or Costa Rica or Cuba or Dominica or Dominican Republic or Guinea or Ecuador or Fiji or Gabon or Georgia or Grenada or Guatemala or Guyana or Iran or Iraq or Jamaica or Jordan or Kazakhstan or Kosovo or Lebanon or Libya or Malaysia or Maldives or Marshall Islands or Mauritius or Mexico or Montenegro or Namibia or Nauru or North Macedonia or Paraguay or Peru or Romania or Russia* or Samoa or Serbia or Sri Lanka or South Africa or Lucia or (Vincent and the Grenadines) or Suriname or Thailand or Tonga or Turkey or Turkmenistan or Tuvalu or Venezuela).af.
3	1 and 2

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist of items specific to IPD meta-analyses, relevant to this study can be found in Appendix 1 - Table 2). We assessed the risk of bias using the AXIS critical appraisal tool (Appendix 1 - Table 3).

Additional information on each study and specific data assumptions are provided in Appendix 1- Table 4. The data dictionary associated with the combined dataset shared on github (https://github.com/mrc-ide/contact_patterns), is shown in Appendix 1- Table 5.

Appendix 1 - Table 2: PRISMA-IPD Checklist of items to include when reporting a systematic review and meta-analysis of individual participant data (IPD)

PRISMA-IPD Section/topic	Item No	Checklist item	Reported on page
Title			
Title	1	Identify the report as a systematic review and meta-analysis of individual participant data.	Title page
Abstract			
Structured summary	2	Provide a structured summary including as applicable:	Abstract page
		Background: state research question and main objectives, with information on participants, interventions, comparators and outcomes.	
		Methods: report eligibility criteria; data sources including dates of last bibliographic search or elicitation, noting that IPD were sought; methods of assessing risk of bias.	
		Results: provide number and type of studies and participants identified and number (%) obtained; summary effect estimates for main outcomes (benefits and harms) with confidence intervals and measures of statistical heterogeneity. Describe the direction and size of summary effects in terms meaningful to those who would put findings into practice.	
		Discussion: state main strengths and limitations of the evidence, general interpretation of the results and any important implications.	
		Other: report primary funding source, registration number and registry name for the systematic review and IPD meta-analysis.	
Introduction			
Rationale	3	Describe the rationale for the review in the context of what is already known.	Last 2 paragraphs of the introduction
Objectives	4	Provide an explicit statement of the questions being addressed with reference, as applicable, to participants, interventions, comparisons, outcomes and study design (PICOS). Include any hypotheses that relate to particular types of participant-level subgroups.	Last paragraph of the introduction
Methods			
Protocol and registration	5	Indicate if a protocol exists and where it can be accessed. If available, provide registration information including registration number and registry name. Provide publication details, if applicable.	The protocol is available through PROSPERO (registration number:

			CRD42020191197) and this is stated in the first paragraph of the Methods
Eligibility criteria	6	Specify inclusion and exclusion criteria including those relating to participants, interventions, comparisons, outcomes, study design and characteristics (e.g. years when conducted, required minimum follow-up). Note whether these were applied at the study or individual level i.e. whether eligible participants were included (and ineligible participants excluded) from a study that included a wider population than specified by the review inclusion criteria. The rationale for criteria should be stated.	Methods: the first paragraph under “Systematic Review”
Identifying studies - information sources	7	Describe all methods of identifying published and unpublished studies including, as applicable: which bibliographic databases were searched with dates of coverage; details of any hand searching including of conference proceedings; use of study registers and agency or company databases; contact with the original research team and experts in the field; open adverts and surveys. Give the date of last search or elicitation.	Methods: the first two paragraphs under “Systematic Review”
Identifying studies - search	8	Present the full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Appendix 1 - Table 1 .
Study selection processes	9	State the process for determining which studies were eligible for inclusion.	Appendix 1 - Figure 1 (PRISMA flow diagram)
Data collection processes	10	Describe how IPD were requested, collected and managed, including any processes for querying and confirming data with investigators. If IPD were not sought from any eligible study, the reason for this should be stated (for each such study).	Methods: second Paragraph (under “Data extraction”)
		If applicable, describe how any studies for which IPD were not available were dealt with. This should include whether, how and what aggregate data were sought or extracted from study reports and publications (such as extracting data independently in duplicate) and any processes for obtaining and confirming these data with investigators.	
Data items	11	Describe how the information and variables to be collected were chosen. List and define all study level and participant level data that were sought, including baseline and follow-up information. If applicable, describe methods of standardising or translating variables within the IPD datasets to ensure common scales or measurements across studies.	First paragraph under “Statistical Analysis”, Appendix 1 – Table 4
IPD integrity	A1	Describe what aspects of IPD were subject to data checking (such as sequence generation, data consistency and completeness, baseline imbalance) and how this was done.	Appendix 1 – Table 4
Risk of bias assessment in	12	Describe methods used to assess risk of bias in the individual studies and whether this was applied separately for each outcome. If applicable, describe how findings of IPD checking were used to inform the assessment. Report if	2nd paragraph of methods

individual studies.		and how risk of bias assessment was used in any data synthesis.	and Appendix 1 - Table 3
Specification of outcomes and effect measures	13	State all treatment comparisons of interests. State all outcomes addressed and define them in detail. State whether they were pre-specified for the review and, if applicable, whether they were primary/main or secondary/additional outcomes. Give the principal measures of effect (such as risk ratio, hazard ratio, difference in means) used for each outcome.	Final 2 paragraphs of methods section under "Statistical Analysis"
Synthesis methods	14	Describe the meta-analysis methods used to synthesise IPD. Specify any statistical methods and models used. Issues should include (but are not restricted to): • Use of a one-stage or two-stage approach. • How effect estimates were generated separately within each study and combined across studies (where applicable). • Specification of one-stage models (where applicable) including how clustering of patients within studies was accounted for. • Use of fixed or random effects models and any other model assumptions, such as proportional hazards. • How (summary) survival curves were generated (where applicable). • Methods for quantifying statistical heterogeneity (such as I^2 and τ^2). • How studies providing IPD and not providing IPD were analysed together (where applicable). • How missing data within the IPD were dealt with (where applicable).	Final 2 paragraphs of methods section under "Statistical Analysis"
Exploration of variation in effects	A2	If applicable, describe any methods used to explore variation in effects by study or participant level characteristics (such as estimation of interactions between effect and covariates). State all participant-level characteristics that were analysed as potential effect modifiers, and whether these were pre-specified.	NA
Risk of bias across studies	15	Specify any assessment of risk of bias relating to the accumulated body of evidence, including any pertaining to not obtaining IPD for particular studies, outcomes or other variables.	Appendix 1.
Additional analyses	16	Describe methods of any additional analyses, including sensitivity analyses. State which of these were pre-specified.	Second paragraph under statistical analysis.
Results			
Study selection and IPD obtained	17	Give numbers of studies screened, assessed for eligibility, and included in the systematic review with reasons for exclusions at each stage. Indicate the number of studies and participants for which IPD were sought and for which IPD were obtained. For those studies where IPD were not available, give the numbers of studies and participants for which aggregate data were available. Report reasons for non-availability of IPD. Include a flow diagram.	Appendix 1 - Figure 1, Appendix 1 - Table 7 and first paragraph under Results.

Study characteristics	18	For each study, present information on key study and participant characteristics (such as description of interventions, numbers of participants, demographic data, unavailability of outcomes, funding source, and if applicable duration of follow-up). Provide (main) citations for each study. Where applicable, also report similar study characteristics for any studies not providing IPD.	Appendix 1 - Table 6Table 1
IPD integrity	A3	Report any important issues identified in checking IPD or state that there were none.	Appendix 1 - Table 4
Risk of bias within studies	19	Present data on risk of bias assessments. If applicable, describe whether data checking led to the up-weighting or down-weighting of these assessments. Consider how any potential bias impacts on the robustness of meta-analysis conclusions.	Appendix 1 - Table 3
Results of individual studies	20	For each comparison and for each main outcome (benefit or harm), for each individual study report the number of eligible participants for which data were obtained and show simple summary data for each intervention group (including, where applicable, the number of events), effect estimates and confidence intervals. These may be tabulated or included on a forest plot.	Table 1
Results of syntheses	21	Present summary effects for each meta-analysis undertaken, including confidence intervals and measures of statistical heterogeneity. State whether the analysis was pre-specified, and report the numbers of studies and participants and, where applicable, the number of events on which it is based.	Appendix 2- Figure 1, Appendix 2 -Figures 4 and 5, Main manuscript: Figure 1, 3-4.
		When exploring variation in effects due to patient or study characteristics, present summary interaction estimates for each characteristic examined, including confidence intervals and measures of statistical heterogeneity. State whether the analysis was pre-specified. State whether any interaction is consistent across trials.	
		Provide a description of the direction and size of effect in terms meaningful to those who would put findings into practice.	
Risk of bias across studies	22	Present results of any assessment of risk of bias relating to the accumulated body of evidence, including any pertaining to the availability and representativeness of available studies, outcomes or other variables.	1st paragraph under Results and Appendix 1
Additional analyses	23	Give results of any additional analyses (e.g. sensitivity analyses). If applicable, this should also include any analyses that incorporate aggregate data for studies that do not have IPD. If applicable, summarise the main meta-analysis results following the inclusion or exclusion of studies for which IPD were not available.	End of 3rd paragraph under section “Total number of contacts and contact location” under Results.
Discussion			
Summary of	24	Summarise the main findings, including the strength of evidence for each main outcome.	Figure 1, 3 and 4 and

evidence			Discussion 1st paragraph.
Strengths and limitations	25	Discuss any important strengths and limitations of the evidence including the benefits of access to IPD and any limitations arising from IPD that were not available.	Strengths and limitations of the evidence discussed throughout discussion. Only for 3 studies, IPD was unavailable but they were all explored qualitatively (see Appendix 1 and Appendix 1 – Table 6)
Conclusions	26	Provide a general interpretation of the findings in the context of other evidence.	Throughout discussion section
Implications	A4	Consider relevance to key groups (such as policy makers, service providers and service users). Consider implications for future research.	Throughout discussion section
Funding			
Funding	27	Describe sources of funding and other support (such as supply of IPD), and the role in the systematic review of those providing such support.	Funding statement

A1 – A3 denote new items that are additional to standard PRISMA items. A4 has been created as a result of re-arranging content of the standard PRISMA statement to suit the way that systematic review IPD meta-analyses are reported.

Appendix 1 - Table 3: Risk of bias table (AXIS critical appraisal tool)

	LIC	LMIC						UMIC													HIC			
	Le Polain de Waroux 2018	Kumar 2018	Kiti 2014	Potter 2019	Horby 2011	Dodd 2016	Melegaro 2017	Read 2014	Zhang 2019	Huang 2020	Watson 2017	Neal 2020	Grijalva 2015	Ajelli 2017	Dodd 2016	Wood 2012	Mahikul 2020	Stein 2014	Meeyai 2015	Oguz 2018	Mossong 2008	Kwok 2014	Kwok 2018	Leung 2017
1. Were the aims/ objectives of the study clear?	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2. Was the study design appropriate for the stated aims?	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
3. Was the sample size justified? (e.g sample size calculation)	1	1	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
4. Was the target/ reference population clearly defined? (e.g. sufficient information on the setting)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	1	0
5. Was the sample frame taken from an appropriate population base so that it closely represented the target/reference population relevant to our review? (general population)	1	0	1	0	1	0	1	1	1	1	1	0	1	0	0	1	0	0	1	0	1	1	1	1
6. Was the selection process likely to select subjects/ participants that were representative of the target/ reference population ? (random sampling)	1	0	1	0	1	1	1	1	0	1	1	1	0	0	1	1	0	0	0	1	1	1	1	1
7. Were demographic variables collected appropriate to the aims of the study? (participant's age and gender)	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1
8. Was information on potential confounders collected (e.g. household size/ employment or student status)	1	1	1	1	1	1	1	0	1	1	0	1	1	1	1	1	1	1	0	1	1	1	1	1
9. Were the risk factor and outcome variables measured correctly using appropriate survey methodology?	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1
10. Was the final sample size large? (>500 participants)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1
11. Were basic demographic data adequately described?	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
12.Does the study state a response rate of >70%?	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0	1	0	0	0	0
13. Were the results for the analyses described in the methods presented?	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
14. Were the authors' discussions and conclusions justified by the results?	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
15. Were the limitations of the study discussed?	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
16. Were there any funding sources or conflicts of interest that may affect the authors' interpretation of the results?	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
17.Was ethical approval or consent of participants attained?	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1
Quality score (%)	100	88	94	76	82	82	94	82	82	88	88	88	82	71	82	94	76	65	65	88	82	82	88	82

Appendix 1 – Table 4: Study additional Information and data assumptions

Study	Additional data information and assumptions	Data source
China(Read et al., 2014)	The survey was carried out between 2009 and 2010, with interviewer led questionnaires in which study participants reported all the people they encountered the previous day. Parents were interviewed on behalf of children deemed too young to provide reliable information for both the individual participant questionnaire and the contact diary. The study did not differentiate between unique contacts across contact events, and sometimes the same individual may appear in multiple events reported by a participant. The total number of contacts reported by the participant was obtained using variable "**c.all"** in the original publication data. The number of non-physical contacts was derived by subtracting the number of reported physical contacts from the total number of contacts.	Supplementary data from publication
China(Zhang et al., 2020)	The survey was carried out between December 2017 and May 2018 in Shanghai. Participants were requested to record each contact they made on the assigned day (during the 24 hours before going to bed). The study used both prospectively completed paper diaries and retrospectively collected telephone interview surveys. A household size of 0 was assumed to be an error and was set to missing. The questionnaire allowed participants to report a maximum number of 40 individual contacts. Participants were asked to report group contacts (Zenodo variable "**group_n"**), defined as a contact with a group of at least 20 individuals. Additionally, they were asked to include the number of contacts they left out (Zenodo variable "**num_left_out**"). Contacts entered in "**group_n"** and "**num_left_out"** were mutually exclusive and both were added to the individual contacts in the main analysis (and excluded in the sensitivity analysis)	Zenodo
European(Mossong et al., 2008)	The surveys were conducted between May 2005 and September 2006 and were implemented by different commercial companies or public health institutes in eight European countries. Participants were assigned a random day of the week to record every person they had contact with between 5 a.m. and 5 a.m. the following morning. Participants were instructed to record contacted individuals only once in the diary. Diaries for young children were filled in by a parent or guardian on their behalf. The maximum number of contact entries in the diary varied between 29 (UK) and 90 (Belgium). Participants were instructed to not record professional contacts in the diary (eg. with clients) for 4 of the countries if they were above a certain number (Belgium if >20 professional contacts; Germany, Finland and Netherlands if >10 professional contacts). This instruction may have led to some underreporting of contact frequencies in these countries. The number of these additional professional contacts in addition to the estimated number of contacts left out of the survey, which was reported for some surveys (e.g UK), are not available in the public domain (and hence were not used). For further differences in the way surveys were conducted, refer to the supplementary material of the original publication. In the present study, we obtained employment and student status for participants who reported an occupation (variable "**part_occupation**"). Participants were considered as being in employment if they reported "working") and as being a student if they reported "in fulltime or further education".	Zenodo
Fiji(Neal et al., 2020)	A social contact questionnaire was administered between August and November 2015. An experienced study nurse interviewed caregivers by telephone and recorded contact details for them and each of their child participants over the previous 24 h (following a nasopharyngeal carriage survey). Data on contacts made at work were unavailable as this study focused on children and their caregivers. Participants for which an occupation was reported were considered in employment, unless they were one of the following: "Housewife", "Retired", "Unemployed", children under 6 years, or "secondary/college student" (considered as not being in employment). If a participant was not reported as a student and was in employment, they were considered as not being a student.	Data received by the authors of the original publication
Hong Kong(Kwok et al., 2014)	An interviewer-led social-contact questionnaire was carried out between 2009 and 2010. An estimated range for total contacts was reported by the respondents and was available from the original dataset (variables **contactalltotalmax** and **contactalltotalmin**). For the present study, the number of daily contacts was calculated as the mean of **contactalltotalmax** and **contactalltotalmin,** rounded.	Supplementary data from publication, with additional data received

	Similarly, the total number of physical contacts was estimated as the mean of **contactclosemin** and **contactclosemax**, rounded. The same method was used for estimating the number of contacts of a particular duration (e.g. total number of contacts made that lasted over an hour). Participants reporting an occupation were considered to be in employment unless they reported "housewife", "retired", "student", "economically inactive person" or "maid" (considered as not being in employment). If a participant did not report being a student, and they reported an occupation, they were considered as not being a student.	from authors on participant gender household size, day type, student status and employment.
Hong Kong(Kwok et al., 2018)	The surveys took place between May 2012 and September 2013. Only data from the first wave were used (N=1,066). Participants would be interviewed about a randomly assigned day within 4 days after the assigned reporting day. For a single contact made multiple times during the same day, multiple locations could be reported. The coding for occupation was identical to the Kwok 2014 study, and the same procedures were used as explained above to code employment and student status. The number of contacts was reported for individual locations ie. home, school, work, and other ("eat", "play", "shop", "transport", "meet", "other"). The total number of daily contacts was obtained from original variable "**ncontacttotal**". The number of non-physical contacts was derived by subtracting the number of reported physical contacts from the total number of contacts.	Supplementary data from publication, with additional data received from authors on household size, contact location, student status and employment.
Hong Kong(Leung et al., 2017)	The contact survey took place between 2015 and 2016. A total of 430 participants filled in a paper diary and 719 filled in an online questionnaire. For a single contact made multiple times during the same day, multiple locations could be reported. There were cases where the participant had not reported all the contacts encountered on the assigned day and these are recorded in a variable named "**num_left_out**" (estimated number of contacts left out). These were included in the main analysis and excluded in a sensitivity analysis. Participants reporting (in original variable **work_role**) being an "employee", an "employer", or "self-employed" were considered as being in employment and those reporting being "students", "homemakers" or "retired" were considered as not being in employment. "others", "don't know" or "unwilling to answer" were coded as missing for both student and employment statuses. If a participant did not report being a student, and they reported an occupation, they were considered as not being a student.	Zenodo
India(Kumar et al., 2018)	Interviews took place between October 20, 2015 and February 29, 2016. All individuals in each household were interviewed and asked about their contacts in the past 24 hours. A caregiver responded for children five years old or younger, whereas children 6–10 years old responded in the presence of a caregiver. For a single contact made multiple times during the same day, multiple locations could be reported. Respondents could report an encounter with multiple individuals as a "group" contact. These were included in the main analysis and excluded in a sensitivity analysis. Participants were asked "Are you enrolled in school OR college? (yes=student, no=not a student) and "Are you employed outside the home?" (yes=in employment, no=not in employment). Additional data on occupation were used to help define employment status. If participants reported an occupation they were considered in employment unless they reported being "Dependents (Still studying)" , "Aged individuals", "Retired", "House Wife", "Girls not studying but doing household chores" or "Unemployed", even if they had answered "no" to "Are you employed outside the home?". If the answer to this question was missing, and participants reported any of the above in the occupation field, they were considered as not being in employment. Similarly, if occupation was recorded as "Dependents (Still studying)" and the answer to "Are you enrolled in school OR college?" was missing, they were considered as students.	OSF, and additional data received from authors on household size and occupation
Kenya(Kiti et al., 2014)	The study took place over the period 17th August 2011 to 31st January 2012. For participants under 10 years old who were unable to read and write, a "shadow" was asked to report their contacts. Participants were expected to keep the diary for a day, defined as the period between first waking and going to bed for the night. Each contact was recorded only once in the diary during the day of study, and repeat encounters were recorded as tallies. Participant occupation was recorded by original publication variable "**part_job**". If participants reported an occupation they were considered in employment unless they reported being a "Student , "Pre-School", "Retired"	Supplementary data from publication, with the addition of exact participant age received from authors.

	or “Unemployed” (considered not in employment).	
Peru(Grijalva et al., 2015)	The contact survey was carried out from May 2009 through September 2011 and participants reported the number of contacts made over a 24 hour period (5am to 5 am). Contact information for children younger than 10 years was provided by the parents. For a single contact made multiple times during the same day, multiple locations could be reported. Original variables **contd005** and **contd005s** that record the participant’s occupation were used to classify employment. For those who reported an occupation, they were assumed to be “employed” unless they reported the following in contd005 or contd005s: "ESTUDIANTE", “ESTUDIASUPERIOR", "En la escuela", "Su casa" (all 3 of which were coded as student), "ABUELAPATERNA", "ENELJARDIN", "JARDIN", "JARDINLIBRE", "WAWASI" (these were coded as “not in employment”). Participants who reported an occupation, were coded as not being a student.	Zenodo
Russia(Ajelli and Litvinova, 2017)	The survey was conducted between January 28, 2016 and February 26, 2016. The data which were collected during as school closure reactive to pathogen transmission (variable **period**) were excluded (ie. period== “school_closure”). For a single contact made multiple times during the same day, multiple locations could be reported. All surveys were reporting for weekdays. In the original data the variable activity_status reported whether a participant was a “Student”, a “Worker” or “Inactive”. “Student” was assumed to be a student and not in employment, “Worker” was assumed to be in employment and not a student, and “Inactive” was assumed to be neither a student nor in employment. For workers who have large number of contacts at work (such as cashiers of supermarkets, waitresses), study participants were asked to provide an estimate of the number of people that they have contacted at work, referred to as additional professional contacts (found in original variable “**number_of_other_work_contacts”** in Zenodo). These contacts were included in the main analysis and excluded in a sensitivity analysis.	Zenodo and additional data received by authors on 52 additional diaries
S Africa(Wood et al., 2012)	The contact survey was carried out over a period of 4 months in 2010 and participants reported the number of contacts made over a 24 hour period (5am to 5 am). For participants under 11 years of age, parents/guardians completed the diary survey together with the child. Participants were asked to report each contact once, but also recorded whether it was the first time each close contact had been met within the 24-hour period. For a single contact made multiple times during the same day, multiple locations could be reported. Data received in the required format, so no data assumptions were made.	Data received by the authors of the original publication
Senegal(Potter et al., 2019)	Data were collected between August 1, 2009 to February 1, 2010 and respondents reported the number of people they contacted in their own compound on both the morning (AM) and the evening (PM) of the survey day. Next, they were asked whether they had visited a list of twelve geographic locations on the survey day, including up to five (non-home) compounds, a field, market, mosque/church, and others. For each location that was visited, respondents reported the time of day (AM, PM, or both) and the number of people contacted in that location. The same information was collected for the preceding two days. For children too young to respond to the survey, the questions were answered by a parent or guardian. In the current meta-analysis study, only contacts reported for the day before the survey (ie. “yesterday”) were considered, so that is comparable with other studies. Additionally, a complete case analysis was used that ignores those who have missing number of contacts made outside or inside the home for any of the locations (variables **spokenumberyesterday1**,…. , **spokenumberyesterday12**, **contactsnumberyesterdayam**, **contactsnumberyesterdaypm).** A complete case analysis reduces the data to a subset of 1,417, which were used. A limitation of this approach is that it may underestimate the number of contacts as those who reported having visited a location but did not report the contact number were not used in the analysis. No “work” location was recorded, and the total number of contacts was calculated as the sum of contacts made by an individual in their own compound and in other locations. The total number of contacts made by a participant in a day was calculated as the sum of contacts made by a single participant in the morning and in the afternoon (e.g. number of contacts made at home =sum of **contactsnumberyesterdayam** and	OSF

	contactsnumberyesterdaypm variables in the OSF data. **For further details on methodology and citing these data, please refer to the original publication.**	
Thailand(Mahikul et al., 2020)	The survey was conducted in various workplaces in Pathum Thani, Thailand, between September and November 2015 and participants were asked to record their contacts in the past 24 hrs. For a single contact made multiple times during the same day, multiple locations could be reported. Variable **Occupation** in the original study was recorded as : "Agricultural worker", "general labor" or "merchant" (all 3 of which were coded as being in employment), "other", "NA", both coded as missing. Contact age was available in broad age groups: <5, 5 to 15, 15 to 40, >40.	Zenodo
Thailand(Stein et al., 2014)	Online surveys were used to collect contact network patterns in Thailand between November 2012 and May 2013. Participants were asked to record the number of contacts they had the day before. Household size was available from the pilot study and the remaining data were obtained from the second publication. The total number of contacts were obtained from variable **degreeyourspace** in the original data. Two observations with extremely high number of contacts (2,233 and 4,456 contacts made within one day were considered as data entry errors and were set to missing. An additional 36 participants had missing number of contacts and were excluded from the analysis.	Data from two publications (pilot and follow-up study) both available on figshare.
Uganda(le Polain de Waroux et al., 2018)	The survey took place between January and March 2014 and participants were asked about their social contacts in the 24 h preceding the survey. For children < 5 years, parents were asked about their child's encounters. Encounters reported with the same individual in different settings counted as one contact only. Participants were asked to select their occupation (variable **occupation** in supplementary file of published study). Those reported as "school college university student" were coded as students. The following occupations were considered to be in employment: agriculture, manual worker, office worker, shop worker. If one of these occupations were selected, participants were coded as not being students. If participants were recorded in the **occupation** variable as "unemployed" or "housewife" or "pre-school child" or "retired" or "school college university student", they were considered as not being in employment.	Publication data supplement and additional data received from authors on gender, exact age and household size.
Vietnam(Horby et al., 2011)	The survey took place in 2007, and subjects recorded the details of each contact made on the day preceding the interview. If an individual was contacted multiple times during the day, the individual was recorded only once but the total time spent with that person during the day was entered. For a single contact made multiple times during the same day, multiple locations could be reported. If a participant reported being a "student" when listing their occupation variable (variable **part_occupation_detail** in Zenodo file), they were reported as being a student. If an occupation was reported, participants were coded as not being students. Those reporting any occupation were assumed to be in employment and for those reporting "unemployed" or "student" were assumed not to be in employment.	Zenodo
Zambia and South Africa(Dodd et al., 2015)	Interviews took place in February and March 2011 in Zambia and in May and July 2011 in South Africa. Interviewees were asked to report contacts that occurred in the 24 hours preceding the midnight before the interview. For determining employment status, the following question was asked: "How have you contributed to household living during the past year?" (variables **q56_job_0** to **q56_job_7** in Zenodo). Participants with the following answers were considered as being in employment: "working own land", "occasional/seasonal employment", "employed" or "own business". Participants who only reported "No contribution", "Housewife-home-maker" or "Welfare grant" or "student" were considered as not being in employment. Participants who did not report being a "student" were considered as not being a student, only if any information on employment was recorded.	Zenodo
Zimbabwe(Melegaro et al., 2017)	Data were collected from March 2013 to August 2013 (which included a school holiday closure from March 28th to May 6th). Multiple contacts with the same individual were reported only once per day. For illiterate adults and children < 10 years, a designated "shadow" filled in the questionnaire on behalf of the study participant. Contacts made by individuals were reported for two consecutive days. For the present study, only the contacts made within one day of the survey were used (i.e variable **studyDay**==2 using the **contact_extra.csv** file uploaded on Zenodo.	Zenodo

	The employment status of the majority of participants (81%) was unknown. Sector of employment was recorded as: teacher, office worker, agriculture/fishing, retail, casual labour, retired, unemployed or “other” (variable “**work_sector**” in Zenodo file). For “other” work sectors, participants were asked to report their occupation (variable “**work_sector_detail”** in Zenodo file, but this was not filled in by most, and was recorded in this study as missing.	
General assumptions for all studies	For the methodology, if a study employed both a diary-based method AND an interview at the end of the day, then methodology was considered as “Diary”. Structured questionnaires filled in during an interview retrospectively were considered as “Interview”. Participants reporting a contact at work who have missing employment status were assumed to be employed. Participants reporting a contact at school who have a missing student status were assumed to be students (for ages <=18). Entries for children aged<10 which were recorded as “employed” or for which a number of contacts at work was given, were set to missing. Contact duration was categorized into <1hr and 1hr+ to utilize all data on contact duration. For contact-level datasets, the number of total contacts per participant was calculated as the sum across contact rows for a given location (tot_home, tot_school, tot_work, tot_other). For a contact where “False” was recorded for all locations, location was coded as missing. For the total number of contacts per participant per group (eg duration <1hr), the sum of contacts was calculated for each participant, unless all were missing.	NA

Appendix 1 – Table 5: Data dictionary for participant-level data

Variable name	**Description**
study	Survey country and first author of original publication
income	Country-level income group, as defined by the World Bank (LIC/LMIC, UMIC, HIC)
method	Survey methodology (Diary, Interview). “Interview” also includes retrospectively reported phone/online surveys.
part_gender	Participant gender (Female/Male
part_age	Participant age
age3cat	Participant age group (<15, 15 to <65, 65+)
hh_size	Household size
student	Student status (0=no, 1=yes)
employment	Employment status (0=no, 1=yes)
weekday	Survey day type (0=weekend, 1=weekday)
tot_contacts	Total number of daily contacts made by a participant. This includes additional contacts, including additional work contacts, group contacts and number of contacts left out. This variable is used in the main analysis.
tot_contacts_no_add	Total number of daily contacts made by a participant. This excludes additional contacts, such as additional work contacts, group contacts and number of contacts left out. This variable is used in a sensitivity analysis.
tot_phys	Total number of contacts made by a participant that were physical
tot_nonphys	Total number of contacts made by a participant that were not physical
tot_dur_under_1hr	Total number of contacts made by a participant which lasted under 1 hour
tot_dur_1hr_plus	Total number of contacts made by a participant which lasted an hour or longer
tot_home	The number of contacts made by a participant at home
tot_school	The number of contacts made by a participant at school
tot_work	The number of contacts made by a participant at work including additional work contacts (main analysis)
tot_work_no_additional	The number of contacts made by a participant at work without the inclusion of additional work contacts (sensitivity analysis)
tot_other	The number of contacts made by a participant at other locations
tot_miss	The number of contacts made by a participant with a missing location
prop_home	Proportion of contacts that occurred at home, among those with a known location

prop_school	Proportion of contacts that occurred at school, among those with a known location
prop_work	Proportion of contacts that occurred at work, among those with a known location
prop_other	Proportion of contacts that occurred at other locations, among those with a known location
prop_cont_male	Proportion of contacts that are male
prop_cont_female	Proportion of contacts that are female
prop_cont_age1 prop_cont_age2 prop_cont_age3	Proportion of a participant's contacts that belong to each of the 3 broad age groups (group 1= children aged 0 to 12-15; group 2= younger adults aged 13-16 to 40-49; group 3=older adults aged 41-50 or over) Contact age was given as an exact age (green) or an estimated range or age group (yellow) and was categorized into three broad age groups. A total of 5,724 contacts out of 269,662 with available age information, but where the age range given was overlapping across the category bounds, were excluded in the assortativity analysis. For more information see Supplementary Text 3.

	Children 0 to 12-15	**Younger adults (13-16 to 40-49)**	**Older adults (41-50 to max)**
European, Mossong	0 to <15	15 to <45	45+
China, Zhang	0 to <15	15 to <45	45+
Hong Kong, Leung	0 to <15	15 to <45	45+
India, Kumar	0 to <15	15 to <45	45+
Kenya, Kiti	<1,1-5,6-15	16-19,20-49	50+
Peru, Grijalva	0 to <15	15 to <45	45+
Russia, Ajelli	0 to <15	15 to <45	45+
South Africa, Wood	0 to <15	15 to <45	45+
Uganda, Le Polain	<2, 2-4, 5-9, 10-14	15-24, 25-34, 35-44,	45-54, 55-64, 65+
Vietnam, Horby	0-5, 6-15	16-25, 26-34, 35-49	50-64, 65+
Zambia, Dodd	0-4, 5-12	13-25, 26-45	46+
South Africa, Dodd	0-4, 5-12	13-25, 26-45	46+
Zimbabwe, Melegaro	0 to <15	15 to <45	45+
Fiji, Neal	0 to <15	15 to <45	45+
Thailand, Majikul	0-4, 5-14	15-40	41+

Systematic review findings

A total of 3,409 titles and abstracts were retrieved from the databases, and 313 full-text articles were screened for eligibility (Appendix 1 - Figure 1). This search identified 19 studies with suitable contact data from LIC, LMIC and UMIC settings. Details of the identified studies can be found in Appendix 1- Table 6. One study identified in the systematic review and included in the IPD meta-analysis was conducted in both a LMIC (Zambia) and an UMIC (South Africa) setting (Dodd et al., 2015). The studies for which authors did not respond (n=3) included a contact study of infants (Oguz et al., 2018). One contact survey with available data was explored qualitatively and was not included in the meta-analysis as the study focused on meal contacts only (Watson et al., 2017). For the IPD meta-analysis there were 11 additional datasets from HICs (Kwok et al., 2018, 2014; Leung et al., 2017; Mossong et al., 2008) included which are not shown in the PRISMA diagram (Appendix 1 - Figure 1).

The majority of the studies collected data representative of the general population, some of which were conducted in rural sites (Horby et al., 2011; Kiti et al., 2014; le Polain de Waroux et al., 2018), urban sites(Mahikul et al., 2020; Zhang et al., 2020) or a combination of both(Dodd et al., 2015; Melegaro et al., 2017; Neal et al., 2020; Read et al., 2014) (rurality was not explicitly stated for all studies). Some of the studies were carried out in the context of another study or trial, such as a flu vaccine trial in Senegal(Potter et al., 2019), the community-based nasopharyngeal carriage surveys in Fiji(Dunne et al., 2018; Neal et al., 2020)and Uganda(le Polain de Waroux et al., 2018), the Study of Respiratory Infections in Andean Peruvian children (RESPIRA PERU(Grijalva et al., 2015)) and the Manicaland HIV/STD Prevention Study(Melegaro et al., 2017).

Overall, the age range of participants was 0 to 105 years. Although most studies included respondents of all ages, one study restricted their participants to ages over 18 years(Dodd et al., 2015), one to ages over 15 years(Mahikul et al., 2020), one to ages over 6 months(Huang et al., 2020), one study only collected contact data on infants under 6 months(Oguz et al., 2018)and another on contacts of children under 6 years and their caregivers(Neal et al., 2020). The distribution of participant age groups in each study was also dependent on the sampling method. For instance, two studies focused on school and university students and their contacts, thereby oversampling older children and young adults(Ajelli and Litvinova, 2017; Stein et al., 2014). In an online survey conducted in Thailand(Stein et al., 2014), further participants were invited from convenience samples of university students (snowball sampling) and another study conducted in Russia recruited mostly students and one of their parents(Ajelli and Litvinova, 2017). Purposive or quota sampling were also used in four of the studies; one study explored contact patterns of migrant workers in Thailand(Mahikul et al., 2020), another of children and their caregivers in Fiji(Neal et al., 2020), and two studies, one in Hong Kong(Leung et al., 2017) and one in Russia(Ajelli and Litvinova, 2017) where children under 18 and university students were oversampled. Most studies (N=10) deployed a random sampling method (e.g., through population registers), often stratified to include sufficient numbers for each age group. Five studies used a convenience sample(Grijalva et al., 2015; Kumar et al., 2018; Potter et al., 2019; Stein et al., 2014; Zhang et al., 2020) and one had no description of sampling methods(Meeyai et al., 2015).

Sampling weights to account for selection bias were used by ten of the studies to account for over-sampling or under-sampling particular characteristics. Using inverse probability weights, most studies adjusted for the age and gender structure of the target population using national census data(Ajelli and Litvinova, 2017; Dodd et al., 2015; Horby et al., 2011; le Polain de Waroux et al., 2018; Leung et al., 2017; Melegaro et al., 2017; Mossong et al., 2008; Potter et al., 2019; Stein et al., 2014). Less often, studies accounted for selection bias in the level of education(Stein et al., 2014), rurality(Kiti et al., 2014; Neal et al., 2020) and household size(Horby et al., 2011; Mossong et al., 2008; Potter et al., 2019; Stein et al., 2014). Weights were calculated either in a one-step or two-stepped approach, depending on the sampling design (eg. two-stage or stratified design). However, these weights were sometimes not included in the main analysis of a study. A study in Kenya(Kiti et al., 2014) accounted for oversampling of semi-urban locations and under-sampling of rural locations, though differences on the estimated contact rates were negligible with the use of weights. In another study which oversampled school- and university-aged students found no substantial effect of accounting for oversampling(Ajelli and Litvinova, 2017). Studies using random sampling, such as one conducted in China(Zhang et al., 2020), ensured that the age and gender structure of the sample was not significantly different to the one of the general population.

Participant response rates were typically high but variable, ranging from 50% to 98%. All studies reported the number of contacts made in the past 24 hours of (or day preceding) the survey, with some studies reporting the number of contacts over 2(Melegaro et al., 2017)or 3 days(Potter et al., 2019). The definitions of contacts were broadly similar (Appendix 1- Table 6). Specifically, contacts were defined as skin-to-skin (physical) contact or a two-way conversation in the physical presence of another person, with some studies specifying a minimum duration, distance of contact or number of words exchanged. Ten of the studies identified had a retrospective design without the use of diaries, such as an interview-based questionnaire or online survey. Interviews were conducted both face-to-face and over the phone. Eight studies adopted a diary-based design only, often reporting contacts prospectively, and one study included both interview and diary-based methods(Zhang et al., 2020) (Appendix 1- Table 6). Data availability by outcome, study- , participant- and contact-level characteristics are shown in Appendix 1 – Table 7. All studies scored above 65% of the items on the AXIS risk of bias tool, suggesting good or fair quality (Appendix 1 - Table 3).

Appendix 1 - Table 6: Extraction table of study characteristics

Income status	Author, Year	Country, Area/District	Method	N (participants)	N (Contacts)	Contact definition	Data type
LIC	(le Polain de Waroux et al., 2018)	Uganda, Southwest Uganda, Sheema	Interview	568	3,964	Two-way conversational encounters lasting for ≥5 min	Contact-level
LMIC	(Kumar et al., 2018)	India, Haryana, Faridabad district	Face-to-face Interview	2,943	79,374	A face-to-face conversation within 3 feet	Contact-level
	(Kiti et al., 2014)	Kenya, Kilifi	Diary-based	568	10,042	Direct physical contact involving skin-to-skin touch	Contact-level
	(Potter et al., 2019)	Senegal, Niakhar	Face-to-face Interview	1,417	27,930	A face-to-face conversation	Participant-level
	(Horby et al., 2011)	Vietnam, Red River Delta	Diary-based	865	6,675	Skin-to-skin contact or a face-to-face two-way conversation	Contact-level
	(Dodd et al., 2015)	Zambia, multiple locations	Face-to-face Interview	2,300	11,028	Face-to-face conversation that was longer than a greeting and within an arm's reach.	Contact-level
	(Melegaro et al., 2017)	Zimbabwe, Manicaland	Diary-based	1,245	13,282	Skin-to-skin contact or a face-to-face two-way conversation	Contact-level
UMIC	(Read et al., 2014)	China, Guangzhou	Face-to-face Interview	1,821	33,789	A face-to-face conversation or skin-on-skin touch	Participant-level
	(Zhang et al., 2020)	China, Shanghai (multiple locations)	Both telephone interview and diary-based	965	18,116	Skin-to-skin contact or a face-to-face two-way conversation	Contact-level

(Huang et al., 2020)	China, Guangdong, Pearl River Delta	Face-to-face Interview	5,818	~96,500-97,100	A conversation with three or more words or physical contact.	Not available
(Watson et al., 2017)	Fiji, multiple locations across Central, Northern and Western divisions	Face-to-face Interview	1,814	9,650	Sharing a meal or a table (meal time contacts) or physical (skin-to-skin) contacts.	Not available
(Neal et al., 2020)	Fiji, Suva	Telephone Interview	2,019	12,932	Skin-to-skin contact or all other contact in the physical presence of another person	Contact-level
(Grijalva et al., 2015)	Peru,San Marcos, Cajamarca	Face-to-face Interview	588	9,009	Skin-to-skin contact or a face-to-face two-way conversation no further than 3m apart	Contact-level
(Ajelli and Litvinova, 2017)	Russia,Tomsk	Diary-based	502	9,026	A face-to-face two-way conversation of at least five words	Contact-level
(Dodd et al., 2015)	South Africa, multiple locations	Face-to-face Interview	1,276	6,694	"close contact": face-to-face conversation that was longer than a greeting and within an arm's reach.	Contact-level
(Wood et al., 2012)	South Africa, Cape Town	Diary-based	571	8,919	physical touch or a face-to-face 2-way conversation with 3 or more words	Contact-level
(Mahikul et al., 2020)	Thailand, Pathum Thani	Diary-based	369	8,356	either skin-to-skin contact or a two-way conversation, approximately one meter apart	Contact-level

	(Stein et al., 2014)	Thailand, Bangkok	Online survey	219	12,812	A person sitting or standing within arm's length of the participant for 30 seconds or longer	Participant-level
	(Meeyai et al., 2015)	Thailand	Diary-based	NA	NA	Physical skin-to-skin contacts or a face-to-face two-way conversation	Not available
	(Oguz et al., 2018)	Turkey, Ankara	Diary-based	1,006	4,706	physical skin-to-skin contacts or interaction in close proximity with three or more words directed to the infant	Not available
HIC*	(Mossong et al., 2008)	Belgium	Diary-based	750	8,878	Skin-to-skin contact or a face-to-face 2-way conversation with 3 or more words	Contact-level
		Finland		1,006	11,128		
		Germany		1,341	10,659		
		Italy		849	16,784		
		Luxembourg		1,051	18,352		
		The Netherlands		269	3,726		
		Poland		1,012	16,501		
		United Kingdom		1,012	11,876		
	(Kwok et al., 2014)	Hong Kong	Interview	762	13,980	Face-to-face conversation or skin-on-skin contact	Participant-level
	(Kwok et al., 2018)	Hong Kong	Interview	1,066	13,696	Face-to-face conversation or skin-on-skin contact	Participant-level
	(Leung et al., 2017)	Hong Kong	Diary-based and online	1,149	16,541	Skin-to-skin contact or a face-to-face 2-way conversation with 3 or more words	Contact-level

∗studies conducted in HIC, were not part of the systematic review, but were used as a comparison in the individual participant meta-analysis.

Appendix 1 - Table 7: Data availability by study

		Outcomes				Study/ participant characteristics							Contact characteristics	
Study	**Income**	**Total contacts**	**Contact location**	**Contact type (physical/ non-physical)**	**Contact duration**	**Age**	**Gender**	**Day type (weekend or weekday)**	**Household size**	**Student status**	**Employment status**	**Method (diary/ interview)**	**Gender**	**Age**
China (Read et al., 2014)	**UMIC**	✓	✗	✓	✓	✓	✓	✗	✗	✗	✗	✓	✗	✗
China (Zhang et al., 2020)	**UMIC**	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗	✓	✓	✓ (exact)
European (Mossong et al., 2008)	**HIC**	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓ (exact)
Fiji (Neal et al., 2020)	**UMIC**	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓ (exact)
Hong Kong (Kwok et al., 2014)	**HIC**	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗
Hong Kong (Kwok et al., 2018)	**HIC**	✓	✓	✗	✗	✓	✓	✓	✓	✓	✓	✓	✗	✗
Hong Kong (Leung et al., 2017)	**HIC**	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓ (exact)
India (Kumar et al., 2018)	**LIC/LMIC**	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓ (exact)
Kenya (Kiti et al., 2014)	**LIC/LMIC**	✓	✗	✓	✗	✓	✓	✓	✓	✓	✓	✓	✗	✓ (age groups)
Peru (Grijalva et al., 2015)	**UMIC**	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓ (exact)
Russia (Ajelli and Litvinova, 2017)	**UMIC**	✓	✓	✗	✗	✓	✗	✓	✓	✓	✓	✓	✗	✓ (exact)
S Africa (Dodd et al., 2015)	**UMIC**	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓ (age groups)
S Africa (Wood et al., 2012)	**UMIC**	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓ (exact)
Senegal (Potter et al., 2019)	**LIC/LMIC**	✓	✓	✗	✗	✓	✓	✓	✓	✗	✗	✓	✗	✗
Thailand (Mahikul et al., 2020)	**UMIC**	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓ (age groups)
Thailand (Stein et al., 2014)	**UMIC**	✓	✓	✗	✗	✓	✓	✓	✓	✗	✗	✓	✗	✗
Uganda (le Polain de Waroux et al., 2018)	**LIC/LMIC**	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓ (age groups)
Vietnam (Horby et al., 2011)	**LIC/LMIC**	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓ (age groups)
Zambia (Dodd et al., 2015)	**LIC/LMIC**	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓ (age groups)
Zimbabwe (Melegaro et al., 2017)	**LIC/LMIC**	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓ (age groups)

List of Figures and Tables:

APPENDIX 2

Additional IPD results and sensitivity analyses

The median number of total daily contacts by gender, day types, participant age, household size, survey methodology and student/employment status is shown in Appendix 2 – Figure 1. Boxes indicate the interquartile range. The overlaid violin plots indicate the probability density of the data at different values.

The total number of contacts made at home was proportional to the participant's household size (Appendix 2 – Figure 2). This figure shows the ratio of median number of home contacts to household size, with the shaded area denoting the interquartile range of the ratio. Ratios of >1 (y-axis) indicate more home contacts than household members.

The sensitivity analysis shown in Appendix 2 – Figure 3 excluded additional contacts (such as additional work contacts, group contacts, and number missed out, which were recorded separately and in less detail by participants compared to their other contacts (Ajelli and Litvinova, 2017; Kumar et al., 2018; Leung et al., 2017; Zhang et al., 2020). The coefficients being plotted are those reported in the forest plots in Figure 1. In all instances, coefficients were strongly correlated, with Pearson's correlation values of 0.77, 0.97 and 0.99 for LICs/LMICs, UMICs and HICs, respectively.

The relationships between proportion of contacts made at each location and the different participant or survey characteristics (i.e age, gender, household size, day of the week, employment/student status) are shown in Appendix 2 – Figure 4. Stacked bar charts for the absolute number of daily contacts by location are shown in Appendix 2 – Figure 5. The relationship between contact location with (a) the proportion of contacts which were physical, (b) the proportion of contacts which lasted a minimum of one hour are shown Appendix 2 – Figure 6.

A sensitivity analysis was used to compare all estimated coefficients in the main analysis to an analysis weighting each study equally within an income group (Appendix 2 – Figure 7). This sensitivity analysis uses the same total sample size for an income group and weighs each study equally within an income group, irrespective of its observed sample size. The coefficients being plotted in this figure are those reported in the forest plots in Figures 1, 3 and 4 in the main text, for total daily contacts (A), whether a contact was physical (B), and duration of contacts (C).

In the weighted analysis, a reverse effect was observed for methodology and the effect of weekday in the total number of daily contacts for UMICs, as a result of disparities in each study's sample size. In the weighted analysis more contacts were observed in interview surveys as compared to diary-based studies (CRR=1.08, 95% CrI: 1.04-1.13), whereas in the main analysis less contacts were observed in interview surveys (CRR=0.73, 95% CrI: 0.70-0.77). In UMICs, the number of contacts made on weekdays was higher than those made on weekends in the main analysis (CRR=1.09, 95% CrI:1.02-1.17), but in the weighted analysis the opposite was true (CRR=0.92, 95%CrI=0.87-0.98). For the remaining coefficients, only small quantitative changes (in the effect sizes), but no qualitative changes were observed. In all instances, coefficients were strongly correlated, with Pearson's rho correlation values ranging between 0.93 and 1.00, depending on outcome and income group (see Appendix 2 – Table 1).

Appendix 2 - Table 1: Correlation (Pearson's rho) between coefficients estimated in the main analysis and those from the sensitivity analysis weighing each study equally within an income group.

	Income group	Pearson's rho correlation coefficient
Total daily contacts	LIC/LMIC	0.927
	UMIC	0.962
	HIC	0.996
Duration of contacts	LIC/LMIC	0.987
	UMIC	0.984
	HIC	0.998
Physical contacts	LIC/LMIC	0.977
	UMIC	0.974
	HIC	0.998

Assortativity analysis

Among all studies included in the IPD meta-analysis, 12 studies had collected information on the gender of the contact and 15 studies on the contact's age (Appendix 1 – Table 7). Eight of those studies provided the contact's exact age or a minimum and maximum estimate, where exact age was unknown. In the remaining seven studies, contact age was provided as predefined categories which varied across studies.

There were three broad contact age categories: 1= "children" aged 0 to 12-15, 2= "younger adults" aged 13-16 to 40-49 and 3= "older adults" ages 41-49 to maximum age. The minimum and maximum for each broad age category is given as a range instead of fixed values to utilize data from all studies providing any information on contact age. Age of contact was usually given as a category, and these categories were different for each study. Participant age groups were the following, where exact age was known: <15, 15 to <45, 45+.

Assortativity was explored in two ways: A) Each participant in the combined data contributing equally to the matrix proportions (thereby implicitly weighing by study size; Appendix 2 – Table 2) and B) Each study contributing equally to the matrix proportions presented (Appendix 2 – Table 3).

Method A:

Each cell ($m_{r,c}$) in the matrix is defined as the mean proportion of contacts a respondent in age group r makes with a contact in age group c. This weighs each respondent equally and does not take study into account.

i indicates the index of the respondent/ participant

r indicates the age group of the respondent (1 to 3)

c indicates the age group of the contact (1 to 3)

$$m_{r,c} = \frac{1}{n_r} \times \sum_{i=1}^{n_r} p_{r,c,i}$$

where

$$n_r = total\ number\ of\ participants\ in\ age\ group\ \boldsymbol{r}$$

$$p_{r,c,i} = proportion\ of\ contacts\ that\ are\ in\ age\ group\ \boldsymbol{c}\ among\ all\ contacts\ made\ by\ the\ \boldsymbol{i^{th}}\ participant\ in\ age\ group\ \boldsymbol{r}$$

Appendix 2 – Table 2: Assortativity by age and sex, weighing by study sample size (method A)

AGE CATEGORY / GENDER

LIC/LMIC

Participant age	Contact age: 1	2	3
1	0.47	0.41	0.12
2	0.22	0.64	0.14
3	0.20	0.51	0.29

Participant gender	Contact gender: Male	Female
Male	0.59	0.41
Female	0.41	0.59

UMIC

Participant age	Contact age: 1	2	3
1	0.34	0.51	0.15
2	0.20	0.62	0.17
3	0.14	0.41	0.45

Participant gender	Contact gender: Male	Female
Male	0.52	0.48
Female	0.46	0.54

HIC

Participant age	Contact age: 1	2	3
1	0.55	0.31	0.14
2	0.24	0.53	0.23
3	0.15	0.33	0.51

Participant gender	Contact gender: Male	Female
Male	0.51	0.49
Female	0.42	0.58

Method B:

Each cell ($M_{r,c}$) in the matrix is defined as the mean proportion a respondent in age group *r* makes with a contact in age group *c*.

s indicates index of the study

$$M_{r,c} = \frac{1}{N} \times \sum_{s=1}^{N} \left[\frac{1}{n_{r,s}} \times \sum_{i=1}^{n_{rs}} p_{r,c,i,s} \right]$$

where

$n_{r,s} = total\ number\ of\ participants\ in\ age\ group\ \boldsymbol{r}\ in\ study\ \boldsymbol{s}$

$N = total\ number\ of\ studies$

$p_{r,c,i,s} = proportion\ of\ contacts\ that\ are\ in\ age\ group\ \boldsymbol{c}\ among\ all\ contacts\ made\ by\ the\ \boldsymbol{i^{th}}\ participant\ in\ age\ group\ \boldsymbol{r}\ in\ study\ \boldsymbol{s}$

Appendix 2 – Table 3: Assortativity by age and sex, weighing each study equally (method B)

LIC/LMIC

		Contact age		
		1	2	3
Participant age	1	0.48	0.41	0.11
	2	0.27	0.60	0.13
	3	0.21	0.51	0.28

		Contact gender	
		Male	Female
Participant gender	Male	0.55	0.45
	Female	0.42	0.58

UMIC

		Contact age		
		1	2	3
Participant age	1	0.38	0.46	0.16
	2	0.21	0.61	0.18
	3	0.18	0.50	0.32

		Contact gender	
		Male	Female
Participant gender	Male	0.54	0.46
	Female	0.44	0.56

HIC

		Contact age		
		1	2	3
Participant age	1	0.54	0.31	0.14
	2	0.27	0.51	0.22
	3	0.21	0.31	0.48

		Contact gender	
		Male	Female
Participant gender	Male	0.51	0.49
	Female	0.42	0.58

List of Figures and Tables

Appendix 2 - Figure 4: Location of contacts as a percentage of total daily contacts by (A) participant’s age, (B) participant’s gender, (C) day of the week, (D) household size, (E) employment status (in participants aged 18 or over) and (F) student status in participants aged 5 to <20 years.

Appendix 2 - Figure 5: Total number of daily contacts in each location by (A) participant’s age, (B) participant’s gender, (C) day of the week, (D) household size, (E) employment status (in participants aged 18 or over) and (F) student status in participants aged 5 to <20 years.

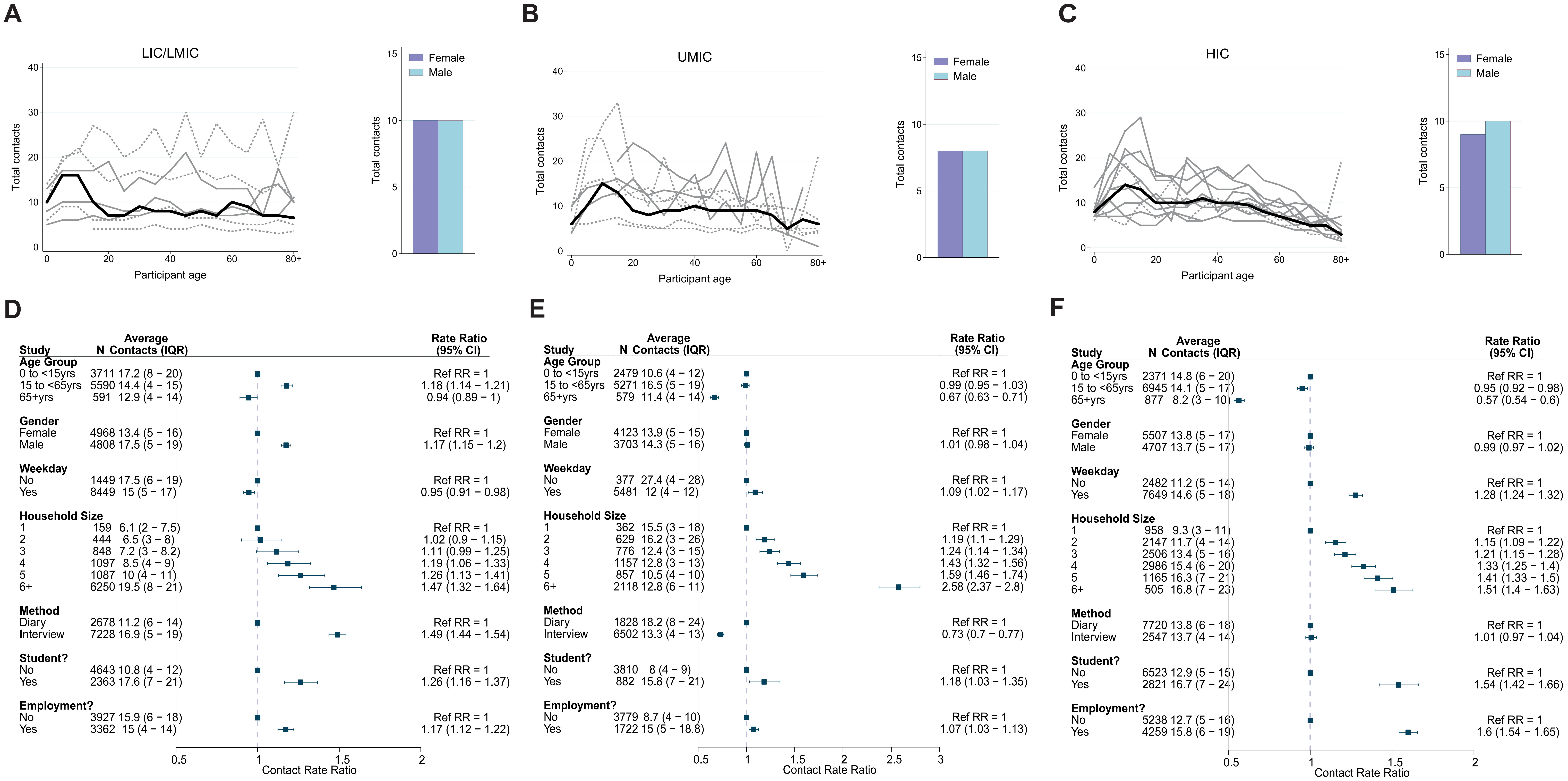
Appendix 2 - Figure 6: Contact location and A) Type of contacts and B) Duration of contact, by income group.

Appendix 2 - Figure 7: Comparison of estimated regression coefficients in the main analysis and sensitivity analysis weighing each study equally within an income group.

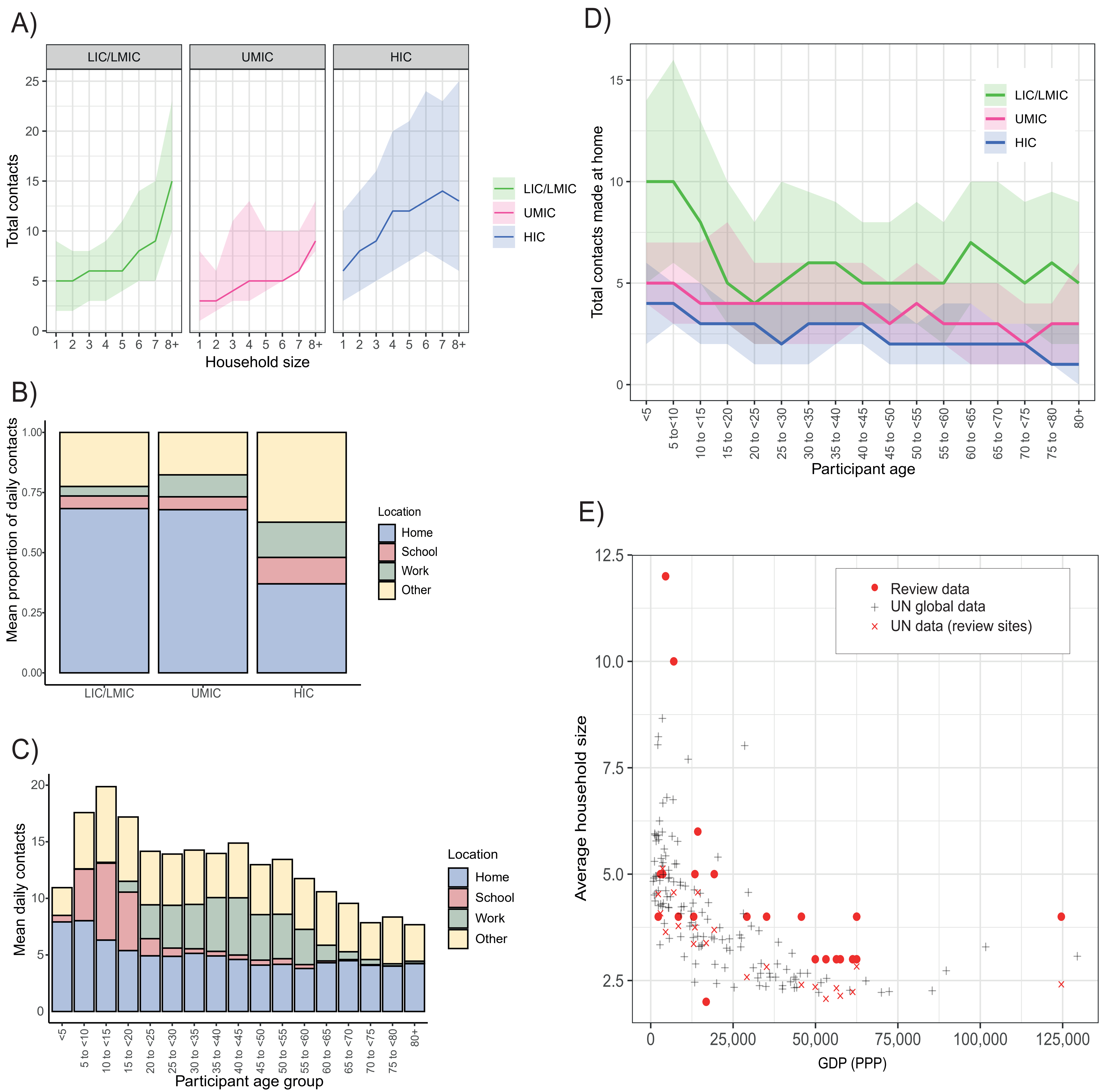
Appendix 2 - Table 1: Correlation (Pearson's rho) between coefficients estimated in the main analysis and those from the sensitivity analysis weighing each study equally within an income group.

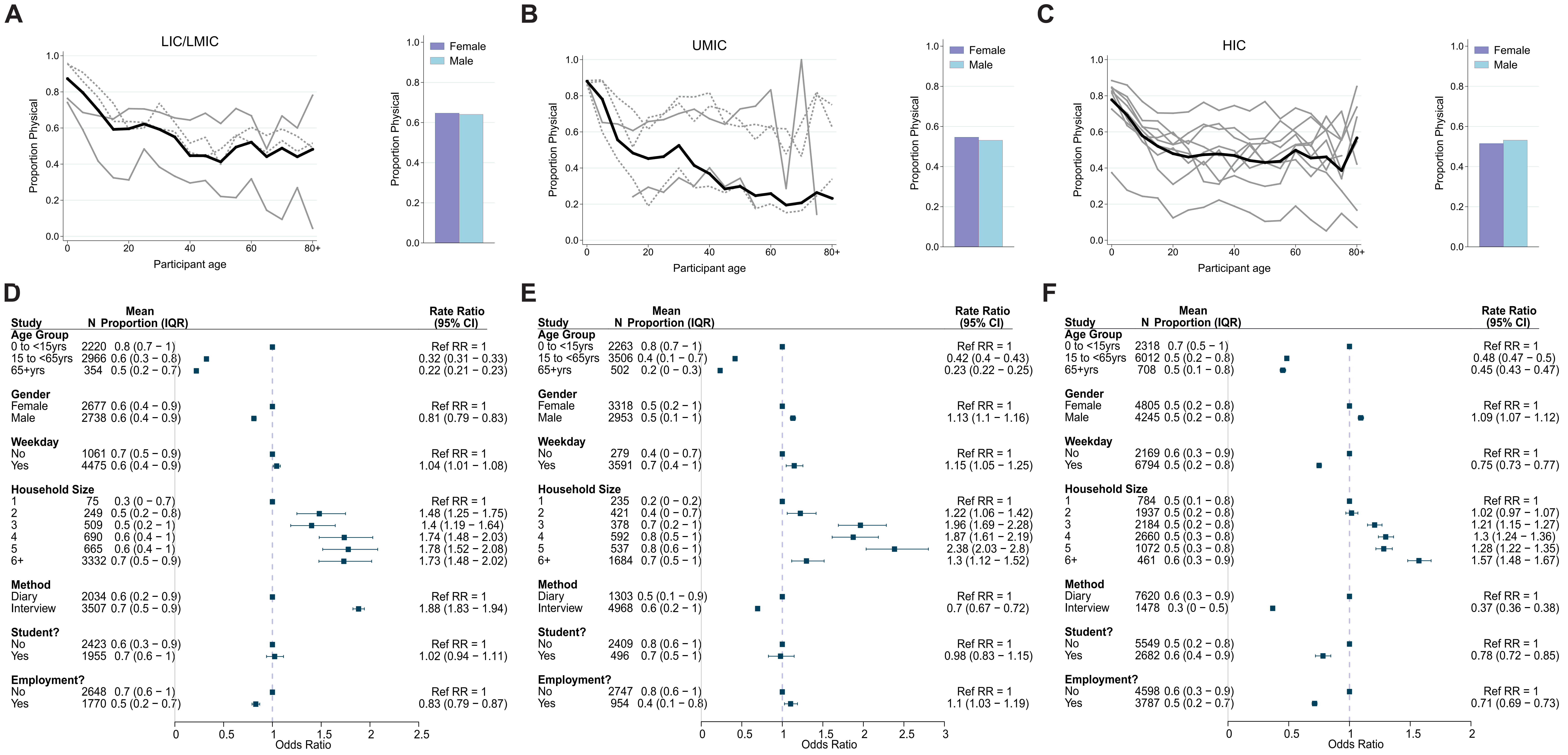
Appendix 2 - Table 2: Assortativity by age and sex, weighing by study sample size (method A)

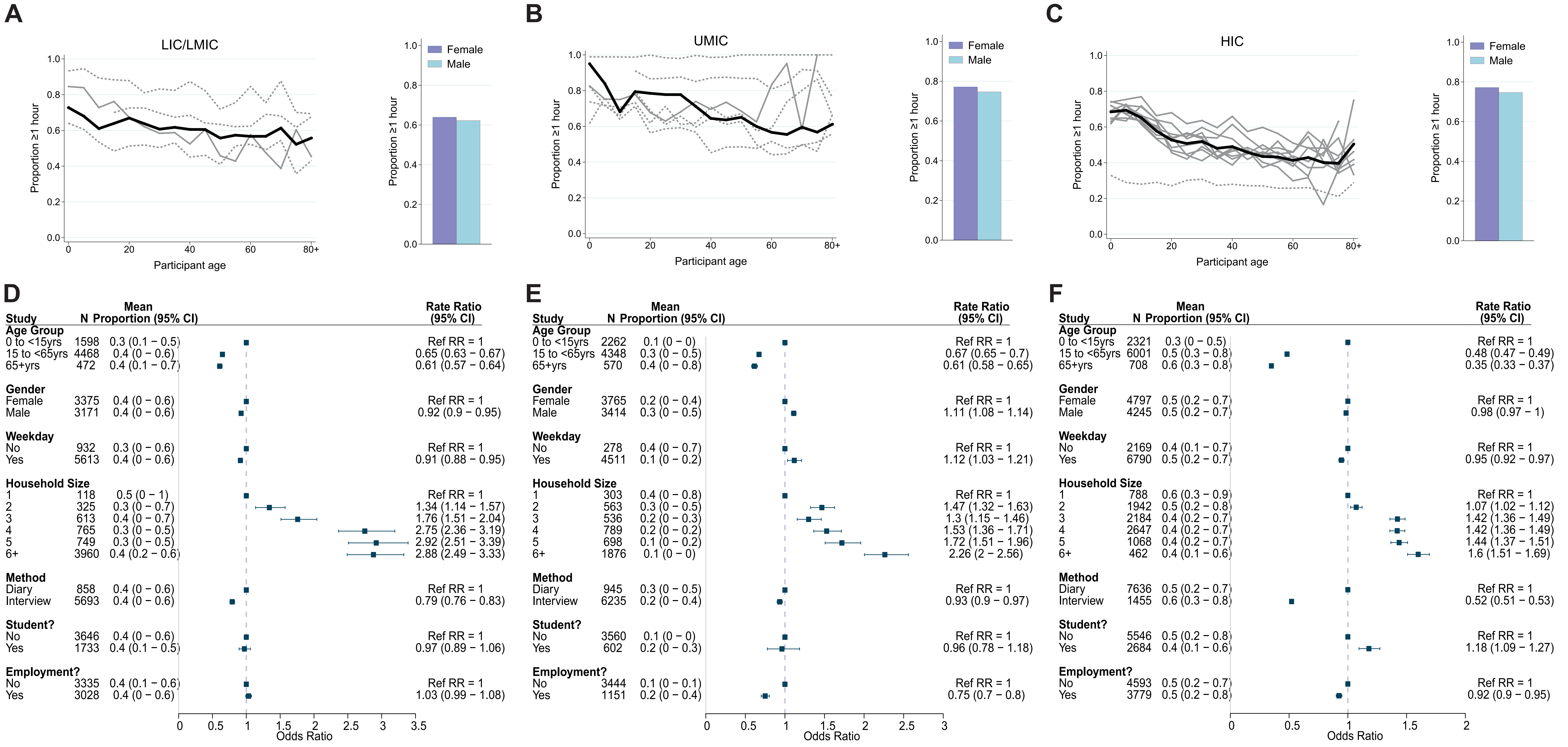
Appendix 2 - Table 3: Assortativity by age and sex, weighing each study equally (method B)



A
LIC/LMIC
Total contacts
Participant age
Female
Male
B
UMIC
Total contacts
Participant age
Female
Male
C
HIC
Total contacts
Participant age
Female
Male
D
Study N Average Contacts (IQR) Rate Ratio (95% CI)
Age Group
0 to <15yrs 3711 17.2 (8 − 20) Ref RR = 1
15 to <65yrs 5590 14.4 (4 − 15) 1.18 (1.14 − 1.21)
65+yrs 591 12.9 (4 − 14) 0.94 (0.89 − 1)
Gender
Female 4968 13.4 (5 − 16) Ref RR = 1
Male 4808 17.5 (5 − 19) 1.17 (1.15 − 1.2)
Weekday
No 1449 17.5 (6 − 19) Ref RR = 1
Yes 8449 15 (5 − 17) 0.95 (0.91 − 0.98)
Household Size
1 159 6.1 (2 − 7.5) Ref RR = 1
2 444 6.5 (3 − 8) 1.02 (0.9 − 1.15)
3 848 7.2 (3 − 8.2) 1.11 (0.99 − 1.25)
4 1097 8.5 (4 − 9) 1.19 (1.06 − 1.33)
5 1087 10 (4 − 11) 1.26 (1.13 − 1.41)
6+ 6250 19.5 (8 − 21) 1.47 (1.32 − 1.64)
Method
Diary 2678 11.2 (6 − 14) Ref RR = 1
Interview 7228 16.9 (5 − 19) 1.49 (1.44 − 1.54)
Student?
No 4643 10.8 (4 − 12) Ref RR = 1
Yes 2363 17.6 (7 − 21) 1.26 (1.16 − 1.37)
Employment?
No 3927 15.9 (6 − 18) Ref RR = 1
Yes 3362 15 (4 − 14) 1.17 (1.12 − 1.22)
Contact Rate Ratio
E
Study N Average Contacts (IQR) Rate Ratio (95% CI)
Age Group
0 to <15yrs 2479 10.6 (4 − 12) Ref RR = 1
15 to <65yrs 5271 16.5 (5 − 19) 0.99 (0.95 − 1.03)
65+yrs 579 11.4 (4 − 14) 0.67 (0.63 − 0.71)
Gender
Female 4123 13.9 (5 − 15) Ref RR = 1
Male 3703 14.3 (5 − 16) 1.01 (0.98 − 1.04)
Weekday
No 377 27.4 (4 − 28) Ref RR = 1
Yes 5481 12 (4 − 12) 1.09 (1.02 − 1.17)
Household Size
1 362 15.5 (3 − 18) Ref RR = 1
2 629 16.2 (3 − 26) 1.19 (1.1 − 1.29)
3 776 12.4 (3 − 15) 1.24 (1.14 − 1.34)
4 1157 12.8 (3 − 13) 1.43 (1.32 − 1.56)
5 857 10.5 (4 − 10) 1.59 (1.46 − 1.74)
6+ 2118 12.8 (6 − 11) 2.58 (2.37 − 2.8)
Method
Diary 1828 18.2 (8 − 24) Ref RR = 1
Interview 6502 13.3 (4 − 13) 0.73 (0.7 − 0.77)
Student?
No 3810 8 (4 − 9) Ref RR = 1
Yes 882 15.8 (7 − 21) 1.18 (1.03 − 1.35)
Employment?
No 3779 8.7 (4 − 10) Ref RR = 1
Yes 1722 15 (5 − 18.8) 1.07 (1.03 − 1.13)
Contact Rate Ratio
F
Study N Average Contacts (IQR) Rate Ratio (95% CI)
Age Group
0 to <15yrs 2371 14.8 (6 − 20) Ref RR = 1
15 to <65yrs 6945 14.1 (5 − 17) 0.95 (0.92 − 0.98)
65+yrs 877 8.2 (3 − 10) 0.57 (0.54 − 0.6)
Gender
Female 5507 13.8 (5 − 17) Ref RR = 1
Male 4707 13.7 (5 − 17) 0.99 (0.97 − 1.02)
Weekday
No 2482 11.2 (5 − 14) Ref RR = 1
Yes 7649 14.6 (5 − 18) 1.28 (1.24 − 1.32)
Household Size
1 958 9.3 (3 − 11) Ref RR = 1
2 2147 11.7 (4 − 14) 1.15 (1.09 − 1.22)
3 2506 13.4 (5 − 16) 1.21 (1.15 − 1.28)
4 2986 15.4 (6 − 20) 1.33 (1.25 − 1.4)
5 1165 16.3 (7 − 21) 1.41 (1.33 − 1.5)
6+ 505 16.8 (7 − 23) 1.51 (1.4 − 1.63)
Method
Diary 7720 13.8 (6 − 18) Ref RR = 1
Interview 2547 13.7 (4 − 14) 1.01 (0.97 − 1.04)
Student?
No 6523 12.9 (5 − 15) Ref RR = 1
Yes 2821 16.7 (7 − 24) 1.54 (1.42 − 1.66)
Employment?
No 5238 12.7 (5 − 16) Ref RR = 1
Yes 4259 15.8 (6 − 19) 1.6 (1.54 − 1.65)
Contact Rate Ratio


A)
LIC/LMIC
UMIC
HIC
Total contacts
Household size
B)
Mean proportion of daily contacts
Location
Home
School
Work
Other
C)
Mean daily contacts
Participant age group
D)
Total contacts made at home
Participant age
E)
Average household size
GDP (PPP)
Review data
UN global data
UN data (review sites)


A
LIC/LMIC
Proportion Physical
Participant age
Female
Male
B
UMIC
C
HIC
D
E
F
Study
N
Mean Proportion (IQR)
Rate Ratio (95% CI)
Age Group
0 to <15yrs 2220 0.8 (0.7 − 1) Ref RR = 1
15 to <65yrs 2966 0.6 (0.3 − 0.8) 0.32 (0.31 − 0.33)
65+yrs 354 0.5 (0.2 − 0.7) 0.22 (0.21 − 0.23)
Gender
Female 2677 0.6 (0.4 − 0.9) Ref RR = 1
Male 2738 0.6 (0.4 − 0.9) 0.81 (0.79 − 0.83)
Weekday
No 1061 0.7 (0.5 − 0.9) Ref RR = 1
Yes 4475 0.6 (0.4 − 0.9) 1.04 (1.01 − 1.08)
Household Size
1 75 0.3 (0 − 0.7) Ref RR = 1
2 249 0.5 (0.2 − 0.8) 1.48 (1.25 − 1.75)
3 509 0.5 (0.2 − 1) 1.4 (1.19 − 1.64)
4 690 0.6 (0.4 − 1) 1.74 (1.48 − 2.03)
5 665 0.6 (0.4 − 1) 1.78 (1.52 − 2.08)
6+ 3332 0.7 (0.5 − 0.9) 1.73 (1.48 − 2.02)
Method
Diary 2034 0.6 (0.2 − 0.9) Ref RR = 1
Interview 3507 0.7 (0.5 − 0.9) 1.88 (1.83 − 1.94)
Student?
No 2423 0.6 (0.3 − 0.9) Ref RR = 1
Yes 1955 0.7 (0.6 − 1) 1.02 (0.94 − 1.11)
Employment?
No 2648 0.7 (0.6 − 1) Ref RR = 1
Yes 1770 0.5 (0.2 − 0.7) 0.83 (0.79 − 0.87)
Odds Ratio
Age Group
0 to <15yrs 2263 0.8 (0.7 − 1) Ref RR = 1
15 to <65yrs 3506 0.4 (0.1 − 0.7) 0.42 (0.4 − 0.43)
65+yrs 502 0.2 (0 − 0.3) 0.23 (0.22 − 0.25)
Gender
Female 3318 0.5 (0.2 − 1) Ref RR = 1
Male 2953 0.5 (0.1 − 1) 1.13 (1.1 − 1.16)
Weekday
No 279 0.4 (0 − 0.7) Ref RR = 1
Yes 3591 0.7 (0.4 − 1) 1.15 (1.05 − 1.25)
Household Size
1 235 0.2 (0 − 0.2) Ref RR = 1
2 421 0.4 (0 − 0.7) 1.22 (1.06 − 1.42)
3 378 0.7 (0.2 − 1) 1.96 (1.69 − 2.28)
4 592 0.8 (0.5 − 1) 1.87 (1.61 − 2.19)
5 537 0.8 (0.6 − 1) 2.38 (2.03 − 2.8)
6+ 1684 0.7 (0.5 − 1) 1.3 (1.12 − 1.52)
Method
Diary 1303 0.5 (0.1 − 0.9) Ref RR = 1
Interview 4968 0.6 (0.2 − 1) 0.7 (0.67 − 0.72)
Student?
No 2409 0.8 (0.6 − 1) Ref RR = 1
Yes 496 0.7 (0.5 − 1) 0.98 (0.83 − 1.15)
Employment?
No 2747 0.8 (0.6 − 1) Ref RR = 1
Yes 954 0.4 (0.1 − 0.8) 1.1 (1.03 − 1.19)
Odds Ratio
Age Group
0 to <15yrs 2318 0.7 (0.5 − 1) Ref RR = 1
15 to <65yrs 6012 0.5 (0.2 − 0.8) 0.48 (0.47 − 0.5)
65+yrs 708 0.5 (0.1 − 0.8) 0.45 (0.43 − 0.47)
Gender
Female 4805 0.5 (0.2 − 0.8) Ref RR = 1
Male 4245 0.5 (0.2 − 0.8) 1.09 (1.07 − 1.12)
Weekday
No 2169 0.6 (0.3 − 0.9) Ref RR = 1
Yes 6794 0.5 (0.2 − 0.8) 0.75 (0.73 − 0.77)
Household Size
1 784 0.5 (0.1 − 0.8) Ref RR = 1
2 1937 0.5 (0.2 − 0.8) 1.02 (0.97 − 1.07)
3 2184 0.5 (0.2 − 0.8) 1.21 (1.15 − 1.27)
4 2660 0.5 (0.3 − 0.8) 1.3 (1.24 − 1.36)
5 1072 0.5 (0.3 − 0.8) 1.28 (1.22 − 1.35)
6+ 461 0.6 (0.3 − 0.9) 1.57 (1.48 − 1.67)
Method
Diary 7620 0.6 (0.3 − 0.9) Ref RR = 1
Interview 1478 0.3 (0 − 0.5) 0.37 (0.36 − 0.38)
Student?
No 5549 0.5 (0.2 − 0.8) Ref RR = 1
Yes 2682 0.6 (0.4 − 0.9) 0.78 (0.72 − 0.85)
Employment?
No 4598 0.6 (0.3 − 0.9) Ref RR = 1
Yes 3787 0.5 (0.2 − 0.7) 0.71 (0.69 − 0.73)
Odds Ratio


A
LIC/LMIC
Proportion ≥1 hour
Participant age
Female
Male
B
UMIC
Proportion ≥1 hour
Participant age
Female
Male
C
HIC
Proportion ≥1 hour
Participant age
Female
Male
D
Study N Mean Proportion (95% CI) Rate Ratio (95% CI)
Age Group
0 to <15yrs 1598 0.3 (0.1 − 0.5) Ref RR = 1
15 to <65yrs 4468 0.4 (0 − 0.6) 0.65 (0.63 − 0.67)
65+yrs 472 0.4 (0.1 − 0.7) 0.61 (0.57 − 0.64)
Gender
Female 3375 0.4 (0 − 0.6) Ref RR = 1
Male 3171 0.4 (0 − 0.6) 0.92 (0.9 − 0.95)
Weekday
No 932 0.3 (0 − 0.6) Ref RR = 1
Yes 5613 0.4 (0 − 0.6) 0.91 (0.88 − 0.95)
Household Size
1 118 0.5 (0 − 1) Ref RR = 1
2 325 0.3 (0 − 0.7) 1.34 (1.14 − 1.57)
3 613 0.4 (0 − 0.7) 1.76 (1.51 − 2.04)
4 765 0.3 (0 − 0.5) 2.75 (2.36 − 3.19)
5 749 0.3 (0 − 0.5) 2.92 (2.51 − 3.39)
6+ 3960 0.4 (0.2 − 0.6) 2.88 (2.49 − 3.33)
Method
Diary 858 0.4 (0 − 0.6) Ref RR = 1
Interview 5693 0.4 (0 − 0.6) 0.79 (0.76 − 0.83)
Student?
No 3646 0.4 (0 − 0.6) Ref RR = 1
Yes 1733 0.4 (0.1 − 0.5) 0.97 (0.89 − 1.06)
Employment?
No 3335 0.4 (0.1 − 0.6) Ref RR = 1
Yes 3028 0.4 (0 − 0.6) 1.03 (0.99 − 1.08)
Odds Ratio
E
Study N Mean Proportion (95% CI) Rate Ratio (95% CI)
Age Group
0 to <15yrs 2262 0.1 (0 − 0) Ref RR = 1
15 to <65yrs 4348 0.3 (0 − 0.5) 0.67 (0.65 − 0.7)
65+yrs 570 0.4 (0 − 0.8) 0.61 (0.58 − 0.65)
Gender
Female 3765 0.2 (0 − 0.4) Ref RR = 1
Male 3414 0.3 (0 − 0.5) 1.11 (1.08 − 1.14)
Weekday
No 278 0.4 (0 − 0.7) Ref RR = 1
Yes 4511 0.1 (0 − 0.2) 1.12 (1.03 − 1.21)
Household Size
1 303 0.4 (0 − 0.8) Ref RR = 1
2 563 0.3 (0 − 0.5) 1.47 (1.32 − 1.63)
3 536 0.2 (0 − 0.3) 1.3 (1.15 − 1.46)
4 789 0.2 (0 − 0.2) 1.53 (1.36 − 1.71)
5 698 0.1 (0 − 0.2) 1.72 (1.51 − 1.96)
6+ 1876 0.1 (0 − 0) 2.26 (2 − 2.56)
Method
Diary 945 0.3 (0 − 0.5) Ref RR = 1
Interview 6235 0.2 (0 − 0.4) 0.93 (0.9 − 0.97)
Student?
No 3560 0.1 (0 − 0) Ref RR = 1
Yes 602 0.2 (0 − 0.3) 0.96 (0.78 − 1.18)
Employment?
No 3444 0.1 (0 − 0.1) Ref RR = 1
Yes 1151 0.2 (0 − 0.4) 0.75 (0.7 − 0.8)
Odds Ratio
F
Study N Mean Proportion (95% CI) Rate Ratio (95% CI)
Age Group
0 to <15yrs 2321 0.3 (0 − 0.5) Ref RR = 1
15 to <65yrs 6001 0.5 (0.3 − 0.8) 0.48 (0.47 − 0.49)
65+yrs 708 0.6 (0.3 − 0.8) 0.35 (0.33 − 0.37)
Gender
Female 4797 0.5 (0.2 − 0.7) Ref RR = 1
Male 4245 0.5 (0.2 − 0.7) 0.98 (0.97 − 1)
Weekday
No 2169 0.4 (0.1 − 0.7) Ref RR = 1
Yes 6790 0.5 (0.2 − 0.7) 0.95 (0.92 − 0.97)
Household Size
1 788 0.6 (0.3 − 0.9) Ref RR = 1
2 1942 0.5 (0.2 − 0.8) 1.07 (1.02 − 1.12)
3 2184 0.4 (0.2 − 0.7) 1.42 (1.36 − 1.49)
4 2647 0.4 (0.2 − 0.7) 1.42 (1.36 − 1.49)
5 1068 0.4 (0.2 − 0.7) 1.44 (1.37 − 1.51)
6+ 462 0.4 (0.1 − 0.6) 1.6 (1.51 − 1.69)
Method
Diary 7636 0.5 (0.2 − 0.7) Ref RR = 1
Interview 1455 0.6 (0.3 − 0.8) 0.52 (0.51 − 0.53)
Student?
No 5546 0.5 (0.2 − 0.8) Ref RR = 1
Yes 2684 0.4 (0.1 − 0.6) 1.18 (1.09 − 1.27)
Employment?
No 4593 0.5 (0.2 − 0.7) Ref RR = 1
Yes 3779 0.5 (0.2 − 0.8) 0.92 (0.9 − 0.95)
Odds Ratio