medRxiv preprint doi: https://doi.org/10.64898/2026.08.18.26360623; this version posted August 21, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Quantifying infection-relevant contact patterns among young children in childcare settings in the United States

Sara L Loo*[1], Anjalika Nande*[2,3], Shaun Truelove[1], and Alison L Hill**[2,4]

[1]Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA
[2]Institute for Computational Medicine, Johns Hopkins University, Baltimore, MD, USA
[3]Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK
[4]Department of Ecology & Evolutionary Biology, University of Toronto, Toronto, ON, Canada

*These authors contributed equally to this work
**Corresponding author, 25 Willcocks Street, Toronto, Ontario, Canada M5S 3B2.
alison.hill@utoronto.ca

Acknowledgments

We thank members of the Johns Hopkins Infectious Disease Dynamics Group and the Insight Net Social Mixing Working group for helpful feedback on this project.

Funding statement

SLL, AN, ST and ALH were supported by the CDC-CFA award for the Atlantic Coast Center for Infectious Disease Dynamics and Analytics (ACCIDDA): NU38FT000012.

Data availability

Code for generating the contact matrices and simulations conducted in this study is open source and available on Github: https://github.com/HopkinsIDD/childcare_contacts

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.64898/2026.08.18.26360623; this version posted August 21, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Quantifying infection-relevant contact patterns among young children in childcare settings in the United States

Abstract

Age is a primary determinant of symptom severity and transmission patterns for many infectious diseases, motivating the use of age-stratified models parameterized by contact matrices. In the United States, the absence of direct contact surveys has required estimating synthetic contact matrices from demographic data on household size, school attendance, and workforce participation. However, this likely underestimates contacts among children under age 5, who often attend group childcare missing from censuses. The goal of this study was to use nationally-representative data on childcare arrangements (the Early Childhood Program Participation Survey) to reconstruct daily contacts occurring in childcare settings, and augment existing all-age contact matrices. For infants under 1 year of age, we estimated 0.2 daily contacts with other infants, increasing to 0.7 daily contacts with same-age peers for 1- or 2-year-olds, 1.3 for 3-year-olds, and 3.5 for 4-year-olds. Including childcare settings increases estimated contacts among young children by up to six fold. Using simulations of measles outbreaks in inadequately vaccinated populations, we show that prior contact matrices significantly underestimated the outbreak frequency, size, and impact on preschool age groups. Our findings highlight the need for targeted data collection on childcare contacts to improve model-based evaluation of interventions particularly for young children.

Introduction

For many infectious diseases, age is a primary determinant of symptom severity and of the patterns of physical contact that can lead to transmission. Thus, classical compartmental mathematical models of infectious disease spread often stratify individuals by age, and describe the rate of disease transmission between different age groups with a contact matrix. Contact matrices and person-to-person interactions are notoriously difficult to measure, but typically describe the daily number (or frequency) of contacts meeting certain spatial and temporal criteria and have been estimated in the past using physical proximity sensors (1–3) or mobile phone data (4, 5), with detailed individual-level surveys or contact "diaries" (6–11), approximated with data from existing demographic censuses (12–14), or inferred from patterns of disease spread (15, 16). These matrices are often composed of several setting-specific matrices, for example contacts in households, workplaces, schools, or the general community.

The burden of many infectious diseases is concentrated in young children. American children under the age of 5 on average have $4-7$ upper respiratory infections a year (17). Pediatric infections have significant direct and indirect economic costs for households, governments, and other payers (18). For example, in the United States, hospitalizations due to respiratory syncytial virus cost approximately $124 per all births per year (19), and before vaccines were introduced, hospitalization with rotavirus infection was estimated to lead to 3.4 lost days of work for parents and a median direct cost of $4565 per hospitalization (20). The more severe outcomes of infections in young children are often attributed to relative lack of immunity (due to waning maternally-acquired antibodies, lack of prior exposure, and vaccination schedules) or other physiological characteristics. The role of increased transmission due to high rates of close physical contact is a common anecdotal explanation but has rarely been formally evaluated. Contacts among children may be especially poorly captured in existing contact matrices. Even with detailed contact diary surveys, responses for younger children are recorded by parent proxies who are likely not present with children in school, and problems of "recall" are likely to be especially salient.

In the United States, while there have been some recent state-level surveys and studies investigating contacts, and particularly changes in contacts during the COVID-19 pandemic (e.g. 7, 21), there is a lack of comprehensive (geographically and age-specific) survey data capturing age-by-age specific contacts relevant to infections spread by close contact, of the type collected in the POLYMOD study covering 8 European countries (6) or in the GlobalMix study covering 4 LMIC countries (22). Instead, two highly cited papers

medRxiv preprint doi: https://doi.org/10.64898/2026.08.18.26360623; this version posted August 21, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

have attempted to impute a plausible contact matrix by using high-level census data and survey derived contact matrices from other countries (12, 13). These contact matrices are the primary sources used in US-based mathematical models (for e.g., 23–29), but they suffer from a serious missing data problem for child-to-child contacts among the youngest children. In the vast majority of US localities, public schooling begins between ages 5 and 6, at "kindergarten". Before that, however, most young children and even many infants are in childcare while their guardians work (30), split roughly between home-based care (which is often informal and unlicensed) and more formal "center-based" care. In these childcare settings, children are often in very close contact with other non-sibling, or non-household, children. However, for the most part, information on childcare arrangements is not recorded in the census or other datasets (which are used to inform synthetic contact matrices (13)). Respondents can indicate "preschool or nursery school" as an option for individuals age 3 and older. However, these terms are not clearly defined, and do not reflect the most common American terms for childcare; e.g., an unlicensed neighbor caring for three unrelated children, would typically be considered home daycare, not a "preschool". Similarly, UNESCO data used in Prem et al. (12) does not include informal care. This limited information misses children under three years old, and even for ages 3-5, fails to capture the diversity of childcare settings and contact patterns, resulting in lower than expected average daily contacts; e.g., the average number of daily contacts between 3 year olds is approximately 1.175, while between 2 year olds it is even lower, at 0.178.

In this study we use data from an in-depth national survey of childcare arrangements for children in the United States to reconstruct realistic age-stratified contacts for young children, and augment existing matrices to be used in infectious disease models (Figure 1). The Early Childhood Program Participation (ECPP) Survey is administered by the National Center for Education Statistics (NCES) every few years to a random subset of guardians of young children (identified via census responses) (31). It collects detailed data on the type(s) of childcare used including the setting, provider type, and hours of attendance, in addition to characteristics of the child and their household. We use the results from this survey along with other sources such as federal guidelines governing licensed childcare facilities to generate synthetic age-structured childcare settings.

From this, we estimate age-by-age average daily contacts for children from infancy through 5 years at a monthly age resolution to produce updated contact matrices to input into transmission models. Our findings suggest that contacts among young children have been dramatically underestimated by prior synthetic contact matrices for the US. Consequently, prior modeling studies using these matrices have likely underestimated the predicted contribution of young children to population-level disease transmission and the importance of disease control in these groups. To highlight the relevance of these results, we demonstrate the effect of our augmented contact matrix on the predicted impact of measles introductions into four scenarios of changing vaccination coverage in children. Our results stress the need to develop more targeted surveys to quantify contacts in young children, especially given the diverse childcare arrangements prevalent in the United States.

Methods

Modeling contacts in childcare between young children

Our goal is to calculate the contact matrix $\mathbf{C}$ whose values $C_{i,j}$ are defined as the average number of contacts per day that a child of age i (*index*) has with children of age j (*contact*) due to interactions in childcare environments. Following the methodology of the ECPP survey (31), we divide childcare setting types k into i) in-home care by a relative caregiver, ii) in-home care by a non-relative caregiver, and iii) center-based care. The overall contact matrix can be decomposed into contributions from each care type,

$$\mathbf{C}_{i,j} = \sum_k p_i^k C_{i,j}^k, \tag{1}$$

where p_i^k is the proportion of children in age group i who receive care in care type k on any given day, $C_{i,j}^k$ is the average number of contacts per day that a child of age i has with children of age j due to interactions

medRxiv preprint doi: https://doi.org/10.64898/2026.08.18.26360623; this version posted August 21, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

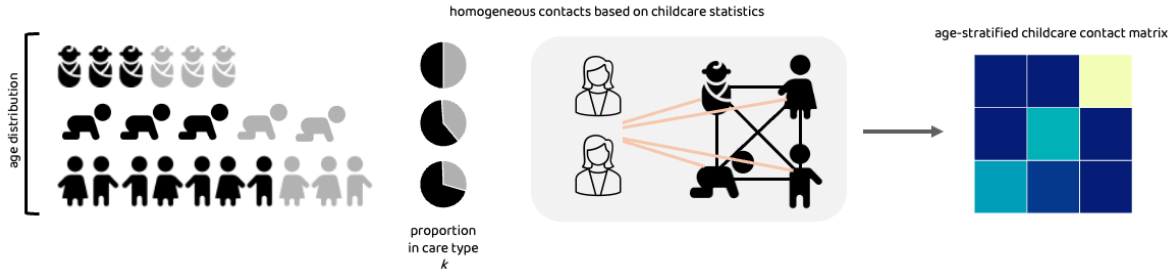


Figure 1: Overview of the construction of contact matrices in childcare settings. Our approach combines information on the age distribution of children under 5, the proportion in each age group in various care types and settings, the frequency of attendance in care, and characteristics of the care setting including class sizes, age groups, and caregiver ratios. We create synthetic childcare settings with statistics matching reported values, and assuming homogeneous mixing within a childcare setting on a given day, create matrices estimating daily contacts per age-by-age group pair.

in childcare type k. Note that children can attend more than one type of care so we can have $\sum_k p_i^k > 1$. Our notation is summarized in Table S1.

Considering the simplest scenario where we assume that children mix homogeneously with all other children receiving care in the same care type:

$$C_{i,j}^k = \left(\frac{p_j^k a_j}{\sum_m p_m^k a_m} \right) \cdot (n^k - 1), \tag{2}$$

where a_j is the fraction of the entire population in age group j and n_k is the average number of children being cared for together in care type k.

We note that $C_{i,j}^k$ is independent of i, and is given by the age distribution of children in that childcare type in general. This is a reasonable assumption for home care arrangements where there are fewer children and a smaller likelihood of being divided into "classes". For those in center-based care, we assume that children are instead divided into "classes" based on ages and that children mainly interact with other children in the same class. We expand care types to include a class setting denoted by l and take into account the probability a child is in care type k and class l based on their age. This is described in further detail in Supplemental Methods.

Based on standards established by large scale contact surveys (6), contact matrices typically count average 'daily contacts'. Therefore, the proportion of children in age group i who receive care of type k on any given day can be decomposed into $p_i^k = f_i^k \cdot \rho_i^k$, where ρ_i^k is the proportion of children aged i who regularly receive care in k (at least once a week), and f_i^k is the average fraction of days of the week they are in care of that type.

Modeling contacts with caregivers

We also estimate contacts between young children and their caregivers in each type of care. This follows suit from other methods (13) where teachers' contacts are incorporated as contacts that occur in 'school' settings. We follow the same logic as above: contacts are dependent on the probability that a teacher in care setting k is of index age i, and encounters a contact individual of age j on an average day. The formulae for estimating these contacts are summarised in Table S2, and described in further detail in Supplemental Methods.

medRxiv preprint doi: https://doi.org/10.64898/2026.08.18.26360623; this version posted August 21, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Augmenting existing contact matrices

After estimating the average daily contacts between young children, we augment an existing well-cited contact matrix (Mistry et al. (13)) to include these missing childcare contacts. To avoid double-counting, we remove all school-setting contacts for individuals under 5 from this matrix—these are either 0 for those between 0-3 or describe contacts for preschoolers as defined by the American Community Survey (ACS) census—and add our estimated childcare contacts. Note, this procedure reduces the overall number of contacts between 3–4-year-olds and older children relative to the matrix of Mistry et al., which includes these cross-age school contacts, whereas ours does not. The origin of these contacts is unclear, as school interactions between preschoolers and children attending elementary school or higher grades would generally not be expected if contacts were assigned strictly based on the ACS classification.

Data

To estimate the childcare utilization statistics and characteristics needed to calculate age-specific contacts, we use the 2019 NCES ECPP Study (31) (the most recent version with raw data publicly available). This is a comprehensive National Household Education Survey that estimates early childhood care and education arrangements for children in the United States under the age of five who were not yet enrolled in kindergarten. For each care type respondents are asked a range of questions including the number of hours and days per week on average their children are in care, and the number of other children cared for in each type. The survey questions and responses used in our analysis are provided in Table S5. There were 7092 survey respondents in total.

For children who receive care in formal childcare centers, the ECPP lacks information about other children participating in these settings. Instead, we assumed that children receiving center-based care are arranged into age-based classes. To choose a reasonable, nationally-representative scheme for arranging children into these classes in our model, we reviewed state-specific regulations on minimum caregiver ratios and maximum class sizes by age (reported on state websites and in "Child Care and Development Fund (CCDF) Plan" documents (32)), as well as guidelines for the same from the federal government's Office of Child Care (33).

To estimate the age and distribution of caregivers, we used summary statistics from the National Survey of Early Care and Education (NSECE) (34, 35), a series of surveys targeted at households, home-based care providers, and classroom staff in early care and education classrooms as a proxy for the age distribution of carers.

Parameter estimates

We extract the monthly age of the index individual and the proportion of children in age group i who routinely receive childcare of type k, ρ_i^k from the ECPP survey.

For each care type k, we calculate the average fraction of days per week children spend in these care types, f_i^k. We used a generalised linear model to fit a natural spline with three degrees of freedom to this fraction of days as a function of age, and found that there was no significant trend by age, but the difference between care types was significant (Figure S2). Thus, we used a fixed value for the average fraction of days a child attends each care type, independent of age.

To estimate the average number of children being cared for together in each setting ($n^{k,l}$), we use a different approach for home-based and center-based care. For at-home care by a relative or non-relative, we use the mean number of other children present in their child's care setting based on the ECPP survey. For children in center-based care, we assumed that centers were operating with age-specific groups of children (classes) and that all contacts occurred within these classes. We assumed that the age bands defining each class, as well as the number of students and caregivers per class, were informed by federal regulations on minimum caregiver ratios and maximum class sizes by state-defined age groups. Additional details are provided in Supplemental Methods.

medRxiv preprint doi: https://doi.org/10.64898/2026.08.18.26360623; this version posted August 21, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

We estimated the conditional age distribution of caregivers given they are in a care setting k. For children taken care of at home by a relative, the median age of the caregivers for at-home relative care was 60 years (interquartile range IQR: 52-65), with a range of 11 to 86 (31).

For non-relative home care and center-based care, we used summary statistics from the NSECE (34, 35). Using these data, we fit a smoothing spline and normalise to ensure a probability distribution (see Figure S1). Further, for the probability that an adult in age group i routinely works as a childcare provider of any type, we use data from the Bureau of Labor Statistics to obtain the age distribution of persons employed providing "child day care services" (36).

Finally, we approximated the US age distribution by month by taking US Census data (37) and assuming that age is distributed equally over monthly age groups.

Case study: Measles simulation

We hypothesized that the underestimation of contacts among young children in existing contact matrices would lead to systematic underestimates of the burden of emerging infectious diseases in this age group, and create biases in estimates of the impact of control measures. Motivated by the resurgence of measles in the US and Canada in 2025, we simulated the distribution of outbreak sizes after the introduction of a single infected case of infection into a population of 100,000. We assumed an R_0 value of 12 (38) in the absence of childhood measles-mumps-rubella (MMR) vaccination, and examined two levels (low: 70%, high: 95% coverage) of vaccination coverage for both younger (1-5 years) and older children (6-13 years) leading to four scenarios (Table 1). These values were chosen to represent recent patterns observed in county-level 2-dose MMR vaccination coverage for kindergarteners in the US (39, 40), where low coverage in younger children and higher coverage in older children indicate declining coverage rates (e.g. Portsmouth county, Virginia), and high coverage in younger children and low in older children indicate increasing coverage rates (e.g. Lassen county, California). For all scenarios we assumed 95% vaccination coverage for ages 14-64 since historically national coverage levels have been high in the US (41) and assumed older adults (aged 65+) are fully immune due to exposure and recovery from natural infection. Vaccine efficacy against infection for one dose of MMR was assumed to be 93% and 97% for two doses (42) and was modeled as an all-or-nothing vaccine, that is, a fraction (given by the efficacy) of the vaccinated individuals are fully protected from infection, while the rest remain fully susceptible. Children aged 1-3 years were assumed to have received one dose of MMR and all other age groups had both doses. We assumed an average infectious period of 8 days (43).

In each scenario we stochastically simulated the spread of measles starting from one infected young child (≤ 5 years) using age-specific contact matrices produced by Mistry et al. (13), and the augmented version including childcare contacts we present here. For each scenario, we ran 10,000 replicates of the simulation and investigated the distribution of total outbreak size (total measles cases) and distribution of cases by age-groups. Model details are provided in Supplemental Methods.

medRxiv preprint doi: https://doi.org/10.64898/2026.08.18.26360623; this version posted August 21, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Table 1: Four scenarios of MMR coverage in children under 13, representing consistently high or low coverage since 2017, or declining or increasing coverage since 2017. Example county-level 2-dose MMR vaccination coverage rates for kindergartners from 2017 to 2023 were extracted from (40).

	High vaccine coverage in older kids (95%)	**Low vaccine coverage in older kids (70%)**
High vaccine coverage in young children (95%)	Consistently high vaccine coverage (95%) in all children under 13, e.g., Northumberland, PA: average 96% since 2017	Increasing vaccine coverage since 2017, e.g., Lassen county, CA: 74% in 2017, 96% in 2023
Low vaccine coverage in young children (70%)	Declining vaccine coverage since 2017, e.g., Portsmouth county, VA: 97% in 2017 to 70% in 2023	Consistently low coverage in all children under 13, e.g., Yates county, NY: average 73% since 2017

medRxiv preprint doi: https://doi.org/10.64898/2026.08.18.26360623; this version posted August 21, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Results

Frequency and characteristics of childcare attendance in the US

On average, 64% of children under 5 were receiving some type of childcare by someone other than their parent or guardian on a regular basis at the time of survey completion; 28% receive home-based care from a relative, 14% are cared for home by a non-relative, and 35% attend a childcare center (Table 2, Figure 2). The proportion of children receiving any type of childcare increased with age - from 40% for infants less than 1 year old to 73% of 4 year olds. For infants, in-home relative care was the most common care type (26% of infants), but for children aged 3-4 center-based care dominated (52%). The proportion of children receiving home-based care increased slightly from birth to around 2 years of age and then decreased for older ages, while the proportion receiving center-based care increased continually (Figure 2). Across all ages, children receiving care spent on average 3.7 days per week in home-based relative care, 3.9 days in home-based non-relative care, and 4.2 days at childcare centers (Table 2, Figure S2A). The average "day" of childcare corresponded to 6.35 hours of care averaging over all care types.

			Age group		
Childcare setting		All (0-4 years)	Infant (<1 year)	Toddler (1-2 years)	Preschool (3-4 years)
None					
	% attending	36.5	59.7	40.5	27.2
Any					
	% attending	63.5	40.3	59.5	72.8
	days/week	3.97	4.22	3.94	3.96
Relative home					
	% attending	28.0	25.6	30.0	26.7
	days/week	3.66	4.12	3.64	3.57
	care group size	1.75	1.68	1.67	1.83
Non-relative home					
	% attending	14.0	8.7	16.1	13.3
	days/week	3.93	4.06	3.94	3.89
	care group size	3.89	3.36	3.73	4.09
Childcare center					
	% attending	35.4	10.6	24.4	51.6
	days/week	4.21	4.58	4.28	4.16
	care group size	Not reported			

Table 2: Childcare use summary from ECPP. Ages indicated are inclusive (i.e., 3-4 includes all 3 and 4 year olds).

Estimated contacts occurring in childcare settings

We used the above data on childcare attendance and characteristics of each childcare setting to construct age-stratified contact matrices in each childcare setting (Figure S3). We estimated that children cared for in-home by relatives contacted an average of 0.8 other children per day compared to 2.7 for children cared for in-home by non-relatives. For children attending childcare centers, estimated contacts varied from 4.6 for infants to 12.9 for 4 year olds. For children receiving home-based care, the contact patterns were determined mainly by the age distribution of children receiving care in this setting, since we assumed home-care settings do not divide children up by age. Children who regularly received home based care from a non-relative had on average 0.3, 0.6, 0.7, 0.6, 0.5 contacts with children ages 0,1,2,3, and 4. For care from relative, contacts were substantially lower, at 0.1, 0.2, 0.2, 0.2 and 0.1 respectively. For center-based care, our inferred contact matrices showed strong associative mixing between ages, due to the division into classes by

medRxiv preprint doi: https://doi.org/10.64898/2026.08.18.26360623; this version posted August 21, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

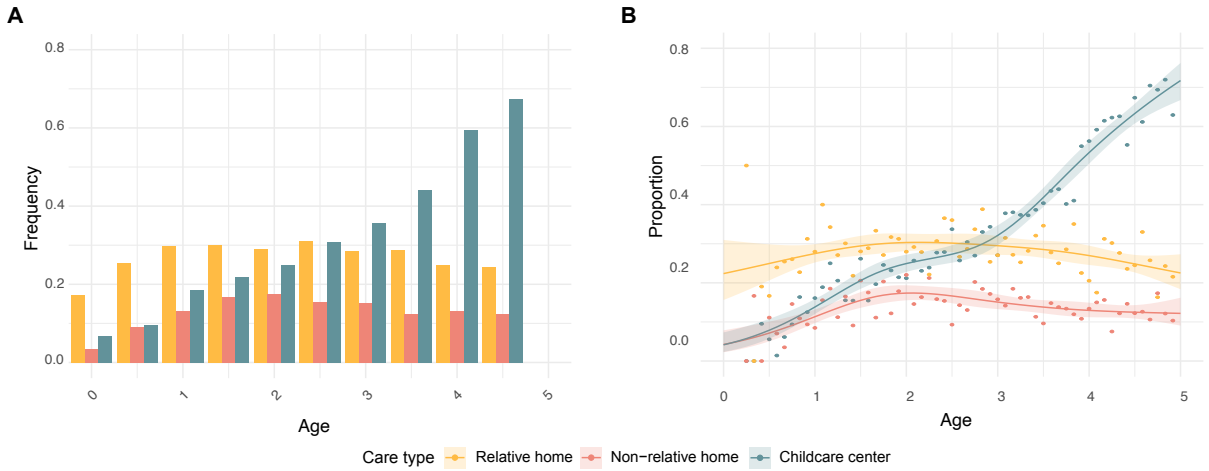


Figure 2: Frequency of childcare attendance in the US by age from ECPP (31) data. (A) The frequency of children accessing any type of care (at-home by a relative (yellow), at-home by a non-relative (red), and Childcare center (blue) settings), by six month age bins. (B) Generalised linear model fit (with 4 degrees of freedom) and 95% confidence intervals for the proportion of children accessing each childcare type by age in months.

age (or developmental stage). For example, 2 year olds in childcare centers were inferred to have 3.4 contacts with other 2 year olds, but only 1.4 with 1 year olds (see Supplemental Figure S3).

Combining contacts across settings and taking into account the proportion of US children in each age group who use any (and each) form of childcare, we constructed an overall age-stratified matrix for contacts occurring in childcare settings (Figure 3A). We observed age-based clustering of contacts, increasing for older children, driven by increased prevalence of center-based care as children approach school age. Overall, we estimated that the average daily contacts occurring in childcare settings was 0.4 for infants, 1.1 for 1 year olds, 1.5 for 2 year olds, 2.9 for 3 year olds, and 5.3 for 4 year olds. Infants on average had 0.2 contacts with other infants, 0.07 with 1 year olds, and only 0.04 with 4 year olds. 4 year olds had on average 3.5 contacts with other 4 year olds, indicating the large relative number of contacts in program care. These values are summarised in Figure 3A, and in the associated files in the provided GitHub repository.

We augmented the contact matrices estimated by Mistry et al. (13) for young children with our childcare matrix (Figure 3B, see Figure S4 for all age-groups). The total number of contacts in young children increased substantially, with the largest increase seen within age-groups; for e.g., contacts between infants, 1-year olds, and 2-year olds increased 3.5, 6.1 and 4.8 fold respectively (Figure 3C). Overall, contacts among children aged 5 years or younger increased by $\sim 20 - 500\%$.

Effect of increased contacts in children in a measles case study

In the measles outbreak simulation we found that across scenarios, the updated matrix with higher contacts among younger children increased the probability and size of outbreaks (Figure 4). Note, since the goal was to compare the effects of the two matrices on measles transmission, we have not included any control measures in the model which would be enacted during real-world outbreaks. The largest difference between the updated with-childcare and the original no-childcare contact matrix occurred under the scenario with low vaccination coverage among younger children and high coverage among older children, where the no-childcare matrix is most likely to substantially underestimate the outbreak size. Differences in age-specific contributions to the epidemic were driven primarily by younger children, whose contribution is increased as expected under a matrix including childcare. For both vaccination coverage levels, the updated matrix—by increasing contacts among young children and reducing between young and older children—results in smaller spillover into older

medRxiv preprint doi: https://doi.org/10.64898/2026.08.18.26360623; this version posted August 21, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

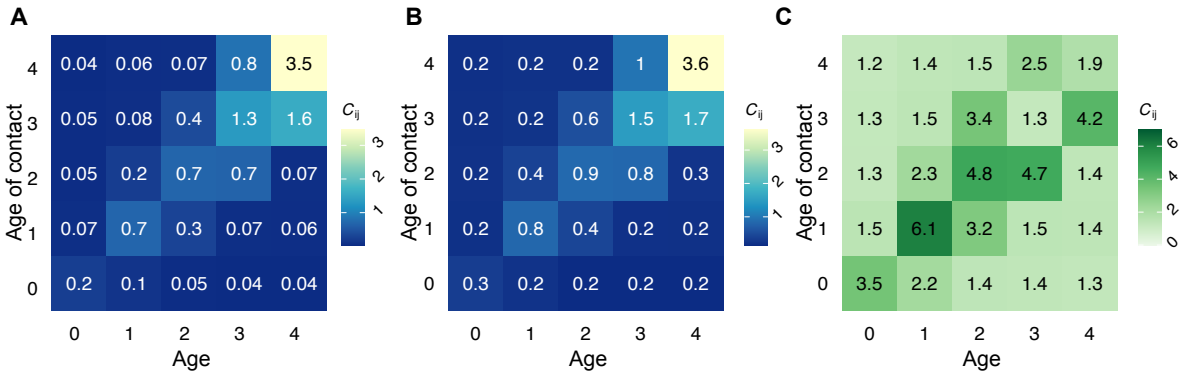


Figure 3: Average daily number of contacts for children under 5 (by yearly age group). (A) Daily number of contacts in childcare settings; and (B) Daily number of contacts in all settings by augmenting existing synthetic matrices in Mistry et al. (13) with childcare matrix. (C) Fold increase in contacts in all settings after augmenting with childcare matrix. The age of the index individual is on the x-axis and the age of the corresponding contact is given on the y-axis.

children. The contribution of individuals over 14 years to the epidemic is largely unchanged since the added adult-to-young children contacts in childcare settings don't affect total adult contacts in a meaningful way (Figure S5). Full distributions of the outbreak sizes broken down by age-groups and mean age-specific attack rates are provided in Supplementary Figures S6, S7.

Discussion

In this study we used a national survey of childcare use to update estimates of the frequency of contact among children under age 5 in the United States, and demonstrated the potential impact on disease transmission patterns. Reliable models of age-stratified infection risk and age-targeted interventions require estimates of the frequency of contact between different age groups. However, comprehensive national surveys of disease-relevant contacts have not been conducted in the United States. Existing US contact matrices are inferred from census data - which lacks information on childcare use - and are calibrated using data from contact surveys in European countries (6) - which differ significantly from the US, where there is no federally-mandated paid parental leave, resulting in very early placement of children in non-parental care arrangements. We argue that prior approaches have led to substantial underestimation of contacts among children not yet in formal schooling, a gap we address by creating updated contact matrices that include our estimates of contacts occurring in childcare settings.

We inferred a significantly higher number of average daily contacts between young children compared to values in commonly used synthetic matrices for the United States (13), with sixfold increase among children aged 1 and a 1.2—5-fold increase among all other age groups under age 5. These results align with expected childcare attendance patterns, and studies of levels of infections in young children, especially those in care (44–46). Simulating outbreaks of measles, we found that our updated estimates of contacts among young children substantially increased i) the probability of an outbreak occurring, ii) the overall outbreak size, and iii) the relative contribution of children ages 0-5 vs 6-13, compared to prior estimates. These differences are most pronounced in scenarios with lower vaccination coverage among younger children versus older children, a scenario that may be becoming increasingly common given recent declines in vaccination coverage. These findings highlight the need for accurate representation of contact patterns to inform outbreak planning efforts.

An advantage of the Early Childhood Program Participation survey, compared with census data, is that by reporting children's age at a monthly resolution, it allows us to estimate contacts at a monthly - vs yearly

medRxiv preprint doi: https://doi.org/10.64898/2026.08.18.26360623; this version posted August 21, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

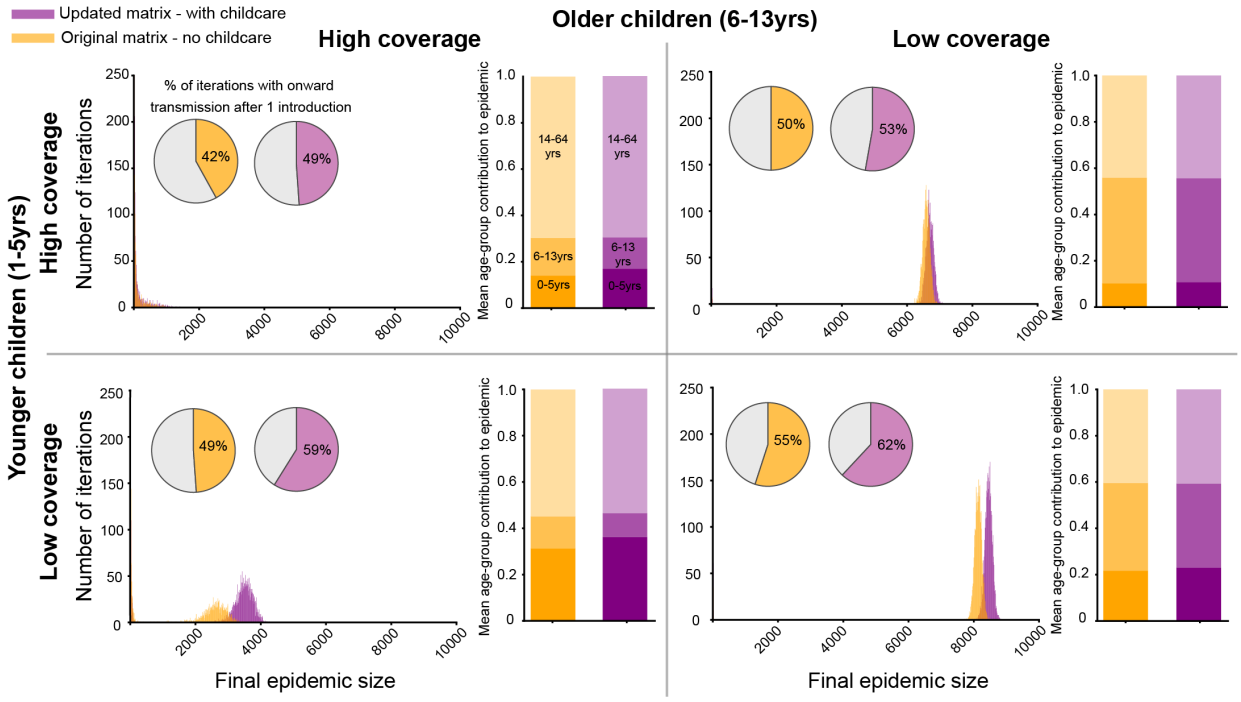


Figure 4: Simulation results of the measles case study across vaccination scenarios. Each of the four panels contains histograms of the final epidemic sizes for iterations where onward transmission occurred after the introduction if a single infection, and bar graphs of the mean contribution per age-group (0-5yrs, 6-13yrs, 14-64yrs) to the total epidemic for both the original - no childcare (orange) and updated - with childcare (purple) contact matrices. Simulation results are for 10,000 replicate introductions each in a population of size 100,000.

medRxiv preprint doi: https://doi.org/10.64898/2026.08.18.26360623; this version posted August 21, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

- age scale. This is especially important for young children, where developmental milestones and fine age scales determine changes in patterns of care (in moving up classes or federally mandated class sizes) and thus exposure to infections. Further, this age scale is important in the context of childhood vaccination policies, where recommendations are defined monthly. For example, a three dose schedule for diphtheria, tetanus, and acellular pertussis is recommended for administration at 2, 4 and 6 months of age, and an RSV monoclonal antibody is seasonally recommended in a child's first year and thus strongly tied to the monthly age of the infant (47). It is therefore useful to have contact matrices at this finer scale to enable simulation of scenarios for optimizing new and existing interventions for childhood infections.

The results of the ECPP survey and the methods we present here provide important insight into the type of data that should be collected in future purpose-designed contact surveys to reliably estimate average daily contacts in childcare settings. Contact surveys should ask specifically about formal and informal childcare arrangements, with clear definitions of any terms like 'daycare' used, and researchers should develop standardized survey methods to ensure coverage of all types of childcare, for all relevant ages. Across the US, there are likely strong geographical differences in contact frequencies for young children, as regulations around childcare group size vary greatly; e.g., Texas and North Carolina allow 20+ 2 year olds per class, Florida allows 40 pre-K students, and Georgia allows 6 infants per caregiver. Conducting contact surveys across different states would help better characterize the different disease dynamics across the country. Globally, any best-practice guidance for conducting new contact surveys should explicitly address how to capture contacts among the youngest age groups, taking location-specific variations in childcare structures into account. Finally, future studies should attempt to directly link contact patterns with epidemiological data to better validate the relationship between contact type, setting, duration, frequency, etc with disease risk.

References

[1] Vanhems P, Barrat A, Cattuto C, Pinton JF, Khanafer N, Régis C, et al. Estimating Potential Infection Transmission Routes in Hospital Wards Using Wearable Proximity Sensors. PLOS ONE. 2013 Sep;8(9):e73970. Publisher: Public Library of Science. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0073970.

[2] Cattuto C, Broeck WVd, Barrat A, Colizza V, Pinton JF, Vespignani A. Dynamics of Person-to-Person Interactions from Distributed RFID Sensor Networks. PLOS ONE. 2010 Jul;5(7):e11596. Publisher: Public Library of Science. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0011596.

[3] Salathé M, Kazandjieva M, Lee JW, Levis P, Feldman MW, Jones JH. A High-Resolution Human Contact Network for Infectious Disease Transmission. Proceedings of the National Academy of Sciences. 2010 Dec:201009094.

[4] Klein B, LaRock T, McCabe S, Torres L, Friedland L, Kos M, et al. Characterizing collective physical distancing in the U.S. during the first nine months of the COVID-19 pandemic. PLOS Digital Health. 2024 Feb;3(2):e0000430. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC10846712/.

[5] Wymant C, Ferretti L, Tsallis D, Charalambides M, Abeler-Dörner L, Bonsall D, et al. The Epidemiological Impact of the NHS COVID-19 App. Nature. 2021 Jun;594(7863):408-12.

[6] Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. PLoS medicine. 2008;5(3):e74.

[7] Feehan DM, Mahmud AS. Quantifying population contact patterns in the United States during the COVID-19 pandemic. Nature communications. 2021;12(1):893.

medRxiv preprint doi: https://doi.org/10.64898/2026.08.18.26360623; this version posted August 21, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

[8] van Hoek AJ, Andrews N, Campbell H, Amirthalingam G, Edmunds WJ, Miller E. The social life of infants in the context of infectious disease transmission; social contacts and mixing patterns of the very young. PloS one. 2013;8(10):e76180.

[9] Kiti MC, Kinyanjui TM, Koech DC, Munywoki PK, Medley GF, Nokes DJ. Quantifying Age-Related Rates of Social Contact Using Diaries in a Rural Coastal Population of Kenya. PLOS ONE. 2014 Aug;9(8):e104786. Publisher: Public Library of Science. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0104786.

[10] Johnstone-Robertson SP, Mark D, Morrow C, Middelkoop K, Chiswell M, Aquino LDH, et al. Social Mixing Patterns Within a South African Township Community: Implications for Respiratory Disease Transmission and Control. American Journal of Epidemiology. 2011 Dec;174(11):1246-55. Available from: https://doi.org/10.1093/aje/kwr251.

[11] Datta S, Mercer CH, Keeling MJ. Capturing sexual contact patterns in modelling the spread of sexually transmitted infections: Evidence using Natsal-3. PLOS ONE. 2018 Nov;13(11):e0206501. Publisher: Public Library of Science. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0206501.

[12] Prem K, Cook AR, Jit M. Projecting social contact matrices in 152 countries using contact surveys and demographic data. PLOS Computational Biology. 2017 Sep;13(9):e1005697. Available from: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005697.

[13] Mistry D, Litvinova M, Pastore y Piontti A, Chinazzi M, Fumanelli L, Gomes MFC, et al. Inferring high-resolution human mixing patterns for disease modeling. Nature Communications. 2021 Jan;12(1):323. Available from: https://www.nature.com/articles/s41467-020-20544-y.

[14] Arregui S, Aleta A, Sanz J, Moreno Y. Projecting social contact matrices to different demographic structures. PLOS Computational Biology. 2018 Dec;14(12):e1006638. Publisher: Public Library of Science. Available from: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1006638.

[15] Ratmann O, Grabowski MK, Hall M, Golubchik T, Wymant C, Abeler-Dörner L, et al. Inferring HIV-1 transmission networks and sources of epidemic spread in Africa with deep-sequence phylogenetic analysis. Nature Communications. 2019 Mar;10(1):1411. Publisher: Nature Publishing Group.

[16] Jombart T, Cori A, Didelot X, Cauchemez S, Fraser C, Ferguson N. Bayesian Reconstruction of Disease Outbreaks by Combining Epidemiologic and Genomic Data. PLOS Computational Biology. 2014 Jan;10(1):e1003457. Publisher: Public Library of Science.

[17] Heikkinen T, Järvinen A. The common cold. The Lancet. 2003 Jan;361(9351):51-9.

[18] de Broucker G, Sim SY, Brenzel L, Gross M, Patenaude B, Constenla DO. Cost of nine pediatric infectious illnesses in low-and middle-income countries: a systematic review of cost-of-illness studies. Pharmacoeconomics. 2020;38(10):1071-94.

[19] Simões EA. The burden of respiratory syncytial virus lower respiratory tract disease in infants in the United States: a synthesis. The Journal of Infectious Diseases. 2022;226(Supplement_2):S143-7.

[20] Mast TC, Walter EB, Bulotsky M, Khawaja SS, DiStefano DJ, Sandquist MK, et al. Burden of childhood rotavirus disease on health systems in the United States: results from active surveillance before rotavirus vaccine introduction. The Pediatric Infectious Disease Journal. 2010;29(2):e19-25.

[21] Dorélien AM, Venkateswaran N, Deng J, Searle K, Enns E, Alarcon Espinoza G, et al. Quantifying social contact patterns in Minnesota during stay-at-home social distancing order. BMC Infectious Diseases. 2023 May;23(1):324. Available from: https://doi.org/10.1186/s12879-022-07968-1.

medRxiv preprint doi: https://doi.org/10.64898/2026.08.18.26360623; this version posted August 21, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

[22] Nelson KN, Kiti MC, Shiiba M, Sacoor C, Bardají A, Macicame I, et al. Characterizing social behavior relevant for infectious disease transmission in four low- and middle-income countries, 2021-2023. Nature Communications. 2025 Oct;16:9586.

[23] Moghadas SM, Shoukat A, Fitzpatrick MC, Wells CR, Sah P, Pandey A, et al. Projecting hospital utilization during the COVID-19 outbreaks in the United States. Proceedings of the National Academy of Sciences. 2020 Apr;117(16):9122-6.

[24] Bubar KM, Reinholt K, Kissler SM, Lipsitch M, Cobey S, Grad YH, et al. Model-Informed COVID-19 Vaccine Prioritization Strategies by Age and Serostatus. Science. 2021 Jan.

[25] Matrajt L, Eaton J, Leung T, Dimitrov D, Schiffer JT, Swan DA, et al. Optimizing Vaccine Allocation for COVID-19 Vaccines Shows the Potential Role of Single-Dose Vaccination. Nature Communications. 2021 Jun;12(1):3449.

[26] Lu FS, Nguyen AT, Link NB, Molina M, Davis JT, Chinazzi M, et al. Estimating the cumulative incidence of COVID-19 in the United States using influenza surveillance, virologic testing, and mortality data: Four complementary approaches. PLOS Computational Biology. 2021 Jun;17(6):e1008994.

[27] Zheng Z, Pitzer VE, Shapiro ED, Bont LJ, Weinberger DM. Estimation of the Timing and Intensity of Reemergence of Respiratory Syncytial Virus Following the COVID-19 Pandemic in the US. JAMA Network Open. 2021 Dec;4(12):e2141779.

[28] Monod M, Blenkinsop A, Xi X, Hebert D, Bershan S, Tietze S, et al. Age groups that sustain resurging COVID-19 epidemics in the United States. Science. 2021 Mar;371(6536):eabe8372.

[29] Aleta A, Martín-Corral D, Pastore y Piontti A, Ajelli M, Litvinova M, Chinazzi M, et al. Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19. Nature Human Behaviour. 2020 Sep;4(9):964-71.

[30] Hanson R, Bobrowski J. Early Childhood Program Participation: 2023. Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education;. NCES 2024-112. Available from: http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2024112.

[31] Cui J, Natzke L. Early Childhood Program Participation: 2019 | IES. National Center for Education statistics; 2021. Available from: https://nces.ed.gov/use-work/resource-library/report/first-look-ed-tab/early-childhood-program-participation-2019?pubid=2020075REV.

[32] Office of Child Care. Approved CCDF Plans (FY 2022-2024). United States Department of Health and Human Services; 2024. Available from: https://acf.gov/occ/form/approved-ccdf-plans-fy-2022-2024.

[33] Office of Child Care. Supervision: Ratios and Group Sizes | Childcare.gov;. Available from: https://web.archive.org/web/20250930222735/https://childcare.gov/consumer-education/regulated-child-care/supervision-ratios-and-group-sizes.

[34] United States Department of Health and Human Services Administration for Children and Families Office of Planning RaE, NSECE Project Team (National Opinion Research Center). National Survey of Early Care and Education (NSECE) Public-Use Files, [United States], 2019. Inter-university Consortium for Political and Social Research [distributor]; 2024. Available from: https://www.childandfamilydataarchive.org/cfda/archives/cfda/studies/37941/versions/V8.

[35] Kim Y, Lea JEA. The Early Care and Education Workforce: A National Snapshot from the NSECE Data. Center for the Study of Child Care Employment, University of California, Berkeley; 2025. Available from: https://cscce.berkeley.edu/publications/data-snapshot/national-snapshot-ece-workforce-nsece.

medRxiv preprint doi: https://doi.org/10.64898/2026.08.18.26360623; this version posted August 21, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

[36] United States Bureau of Labor Statistics. Employed persons by detailed industry and age;. Available from: https://www.bls.gov/cps/cpsaat18b.htm.

[37] Bureau USC. SEX BY SINGLE-YEAR AGE; 2020. Accessed on 29 October 2025. U.S. Census Bureau. Available from: https://data.census.gov/table/DECENNIALDHC2020.PCT12?q=PCT12&d=DEC+Demographic+and+Housing+Characteristics.

[38] Guerra FM, Bolotin S, Lim G, Heffernan J, Deeks SL, Li Y, et al. The basic reproduction number (R0) of measles: a systematic review. The Lancet Infectious Diseases. 2017 Dec;17(12):e420-8.

[39] Saiyed S, Dong E, Gardner LM. The relationship between state vaccination exemption policies and MMR vaccination trends in the U.S. Vaccine. 2025 Oct;64:127773. Available from: https://www.sciencedirect.com/science/article/pii/S0264410X25010709.

[40] Dong E, Saiyed S, Nearchou A, Okura Y, Gardner LM. Trends in County-Level MMR Vaccination Coverage in Children in the United States. JAMA. 2025 Aug;334(8):730-2. Available from: https://doi.org/10.1001/jama.2025.8952.

[41] National Center for Immunization and Respiratory Diseases; Immunization Services Division. Vaccination Coverage among Young Children (0 – 35 Months). Centers for Disease Control and Prevention;. [Accessed: 2025-12-22]. Available from: https://www.cdc.gov/childvaxview/about/interactive-reports.html.

[42] National Center for Immunization and Respiratory Diseases; Division of Viral Diseases. Measles Vaccination. Centers for Disease Control and Prevention; 2025. [Accessed: 2025-12-22]. Available from: https://www.cdc.gov/measles/vaccines/index.html.

[43] National Center for Immunization and Respiratory Diseases; Division of Viral Diseases. Measles Clinical Diagnosis Fact Sheet; 2025. [Accessed: 2025-12-22]. Available from: https://www.cdc.gov/measles/hcp/communication-resources/clinical-diagnosis-fact-sheet.html.

[44] Brady MT. Infectious disease in pediatric out-of-home child care. American Journal of Infection Control. 2005 Jun;33(5):276-85. Available from: https://www.sciencedirect.com/science/article/pii/S0196655304006455.

[45] Søegaard SH, Spanggaard M, Rostgaard K, Kamper-Jørgensen M, Stensballe LG, Schmiegelow K, et al. Childcare attendance and risk of infections in childhood and adolescence. International Journal of Epidemiology. 2023 Apr;52(2):466-75. Available from: https://doi.org/10.1093/ije/dyac219.

[46] Collins JP, Shane AL. Infections Associated With Group Childcare. Principles and Practice of Pediatric Infectious Diseases. 2018:25-32.e3. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC7152033/.

[47] American Academy of Pediatrics. AAP Immunization Schedule; 2026. Available from: https://publications.aap.org/redbook/resources/15585/AAP-Immunization-Schedule.

medRxiv preprint doi: https://doi.org/10.64898/2026.08.18.26360623; this version posted August 21, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Supplementary Methods

Interpreting contacts among young children in published contact matrices

By definition, contact matrices used in mathematical models of infectious disease should describe the number, frequency, or duration of contacts that have the possibility of spreading the disease of interest, between two defined groups of individuals. For diseases spread by close spatial or physical contact, it's usually impossible to measure contacts direction, though some studies have used purpose-deployed physical proximity sensors (1–3) or co-location of mobile devices (4, 5) to approximate contact patterns. More commonly, models used contact matrices derived from "contact diaries", where individuals record the number of unique individuals with whom they have made contact (of a certain duration, intensity, and physical proximity) in a particular 24-hour period. The most well-known example of this method is the POLYMOD survey covering 8 European countries. No such similarly comprehensive contact survey has been conducted in the US, and so instead, two highly cited papers have attempted to impute a plausible contact matrix by using high-level census data and survey derived contact matrices from other countries (12, 13). However, both studies have limitations with regard to the treatment of contacts between young children.

In Mistry et al (13), the average number of contacts occurring in schools ("school" component of matrix) is populated by combining estimates of the frequency of school attendance with estimates of school sizes. School enrollment is taken from the American Community Survey (the detailed "long-form" census administered to a subset of the population every year) (48). For children ages 3 and older, census respondents can choose "nursery school, preschool" as an option, in addition to "kindergarten", "grades 1 - 12", "college undergraduate years" or "graduate or professional school". For older children, it appears that separate data on size distribution of elementary and high schools (and potentially also the grades they serve, which can vary by region) is used to assign contacts.

There are several major limitations to this approach. The resulting contact matrices have zeros for all contacts between age groups less than 3. It is unlikely – even for individuals aged 3-5 – that all attendance in childcare would be recorded with this answer, as many types of childcare are not colloquially referred to as "preschool or nursery school" in the US. It is unclear how the average number of contacts (assumed equal to total school size) is assigned for students in "preschool or nursery school". Since students are of course unlikely to contact every single student in their school, total contact number in the school setting are re-scaled using weights the authors estimated. By repeating this procedure for imputing "synthetic" contact matrices for the eight countries that *did* have empirically-derived contact surveys, and then comparing the synthetic matrices to those from the surveys, the authors derived scaling factors for each setting relating the two. However, differences in the way childcare is encoded in the European-based POLYMOD vs the US mean the weights were likely unreliable in a new setting, and they don't address the issue that children are also organized into age-based classes.

A similar issue occurs in Prem et al (12). The UNESCO data they use specifically "excludes purely family-based arrangements that may be purposeful but are not organized in a 'programme' (e.g. informal learning by children from their parents, other relatives or friends is not included)", which therefore also excludes the most common form of childcare for young children in the US.

Although we focus here on young children, because of their near-absence from existing synthetic matrices, for the reasons mentioned above, "school" contact matrices are likely unreliable for older students too. Students in school spend the majority of time in individual classrooms that are stratified by grade, especially in elementary school, and friendships tend to be highly assortative by age. Confirming this common sense, previous studies that used wearable or mobile-phone based physical proximity sensors in schools found strong evidence of assortativity by class, grade, and age (49). The same was found in detailed contact surveys (children could answer for themselves beginning at age 9 or 13 depending on country in POLYMOD).

The synthetic contact matrices developed in Mistry et al and Prem et al represent an innovative use of existing demographic data and have enabled important mathematical modeling work, and we believe our additional

medRxiv preprint doi: https://doi.org/10.64898/2026.08.18.26360623; this version posted August 21, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

analysis of childcare enrollment data improves these valuable estimates. However, the need persists for more reliable and direct measurements of contacts.

Modeling contacts in center-based childcare between young children

To calculate the average number of daily contacts for children who receive center-based childcare (i.e. care at a "daycare" or "preschool"), we assume that in most cases children in that care type would be divided up into "classes" based on ages, and that children mainly interact with other children in the same class. In this case, we expand each of these care types to include a setting (such as a class or group) indexed by l, and then define $q_i^{k,l}$ as the *conditional* probability an individual of age i who receives childcare of care type k is in class/setting l. Then, $C_{i,j}^k$ the average number of contacts per day that a child of age i has with children of age j in care type k becomes

$$C_{i,j}^k = \sum_l q_i^{k,l} C_{i,j}^{k,l} = \sum_l q_i^{k,l} \left(\frac{q_j^{k,l} p_j^k a_j}{\sum_m q_m^{k,l} p_m^k a_m} \right) \cdot (n^{k,l} - 1), \tag{S1}$$

where $n^{k,l}$ is now the average number of children per class setting l in childcare type k. For example, imagine we are interested in estimating the contacts for children 0-11 months old in childcare centers, and assume that all centers have infant classes with 6 children total ($n_{\text{childcare center, infant class}} = 6$) and 100% of children under age 1 are in this class ($q_{0\text{yr}}^{\text{childcare center, infant class}} = 1.0$). If we assume that childcare center rules are such that children are eligible to move up to the toddler class any time after they reach age 12 months, but they may not move immediately on their birthday due to space constraints or developmental milestones (e.g., not yet walking), then our calculation would additionally need to use - for example - the fact that 10% of children in the 1 year old age group are still in the infant class ($q_{1\text{ yr}}^{\text{childcare center, infant class}} = 0.1$), along with the overall distribution of age groups within childcare center settings ($p_{0yr}^{\text{childcare center}} a_{0yr}$, $p_{1yr}^{\text{childcare center}} a_{1yr}$).

The specific scheme we use for assigning age-based classes, and the data that informs it, is described in the subsequent 'Parameter Estimates' section.

Modeling contacts with caregivers

In addition to the child-to-child contacts, we also include contacts between these young children and their caregivers in each type of care. This follows suit from calculations undertaken by others (13) where teachers were removed from the working population and their contacts made at work were instead incorporated into the school setting matrices. First, for care types where all individuals are in a single group (e.g., at-home care where children are *not* separated into classes), we assume that that there is only a single caregiver. That is, there are no adult-to-adult contacts (note that this can be relaxed). In these care types k, if the index individual is a young child of age i, the number of contacts with a caregiver of age j in setting k is the probability that a caregiver in care-type k is in age group j multiplied by number of caregivers in that setting, that is

$$C_{i,j}^k = r_j^k \cdot n_c^k, \tag{S2}$$

where $n_c^k = 1$ is the number of caregivers in these specific settings of care-type k and r_j^k is the probability that a caregiver in care type k is of age j.

Conversely, if the index individual is a caregiver of age i and the contact is a young child of age j, this becomes the age distribution of children in this care type of a given size

$$C_{i,j}^k = p_j^k \cdot n^k. \tag{S3}$$

Second, for center-based childcare where children are likely divided into classes, the number of contacts between young children of age i with adult contacts of age j is dependent on the the conditional probability

medRxiv preprint doi: https://doi.org/10.64898/2026.08.18.26360623; this version posted August 21, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

that the child is in that class, the probability that the caregiver is of age j and the number of caregivers in that class l. This assumes that children interact with each caregiver in their class equally. This gives

$$C_{i,j}^{k} = \sum_{l} q_i^{k,l} C_{i,j}^{k,l} = \sum_{l} q_i^{k,l} \cdot r_j^{k} \cdot n_c^{k,l}, \tag{S4}$$

where $n_c^{k,l} = n^{k,l}\sigma^{k,l}$ is the number of caregivers per class, decomposed into the number of children per class and the caregiver:student ratio $\sigma^{k,l}$. For the corollary contacts of adult caregivers of age i with young children of age j this becomes,

$$C_{i,j}^{k} = \sum_{l} C_{i,j}^{k,l} = \sum_{l} \left(\frac{q_j^{k,l} p_j^{k} a_j}{\sum_m q_m^{k,l} p_m^{k} a_m} \right) \cdot n^{k,l}. \tag{S5}$$

We also here model contacts between caregivers. We assume that caregivers in at-home care by a relative or non-relative are sole-carers and there are therefore no contacts between adult caregivers. For adults who work in childcare centers, we assume for simplicity that caregivers only contacts others in the same class (given a lack of data regarding overall 'school' size of childcare centers). Then, the number of contacts between caregivers in class l is given by

$$C_{i,j}^{k} = \sum_{l} C_{i,j}^{k,l} = r_j^{k} \sum_{l} (n_c^{k,l} - 1). \tag{S6}$$

The formulae used to construct these contact matrix entries are summarized in Table S2.

Parameter estimates

This section is a more detailed version of the 'Parameter estimates' section from the main text.

To impute age-specific contact matrices from ECPP data, we need the age of the respondent's child at the time the survey was completed. The publicly-available data includes month and year of birth, but no the survey completion date. We estimated age using the reported birth month and year, assuming every survey was completed in August 2019. In reality, survey responses were returned anywhere between January and August 2019. This method likely overestimates ages, on average, but avoids having any negative ages. Note that individuals with children up to age 6 by Dec 2018 who were not yet in kindergarten were asked to participate. However, we focus on individuals age 0-4 in this analysis, since these are children most likely to be in care (vs formal schooling) and since they represent 95% of the data

We constructed contact matrices using monthly age group bins, in contrast to prior work that reported only yearly age groups. For disease applications where there are sharp age-specific differences in immunity (due to, for example, infant vaccination schedules), narrower age bands are commonly used in models and childcare is likely the setting where contacts vary most dramatically with age for younger children. For modeling applications where wider age groups can be used, the resulting matrices can be aggregated. Although the survey includes children up to age 7, we focus on individuals age 0-4 in this analysis, since these are children most likely to be in care (vs formal schooling) and since they represent 95% of the data. Note that individuals with children up to age 6 who were not yet in kindergarten were asked to participate in early 2019, so children could have been up to age 7 by the time the survey was returned)

To determine the proportion of children in age group i who routinely receive childcare of type k, ρ_i^k, we calculate the frequency of responses regarding program attendance in the ECPP. The survey question is a binary response to whether children are receiving care at each of the following childcare types: at home from a relative ("relative home"), at home from a non-relative ("non-relative home"), or attending a daycare center, preschool or pre-kindergarten not in a private home ("childcare center"). The proportion is calculated

medRxiv preprint doi: https://doi.org/10.64898/2026.08.18.26360623; this version posted August 21, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

as the fraction of 'yes' responses over the total respondents, as a function of age i. Since children may receive childcare in more than one setting, these responses may sum to more than 1 for any particular age i.

For each of these care types, respondents are asked how many days and hours each week the child receives care in this care type. We use the responses to this question to calculate the average fraction of days per week children spend in these care types, f_i^k. We used a generalised linear model to fit a natural spline with three degrees of freedom to this fraction of days as a function of age (in monthly bins), and found that there was no significant trend by age, but the difference between care types was significant (Figure S2). Thus, we used a fixed value for the average fraction of days a child attends each care type, independent of age (Table S3). Although we found vidence of some bimodality in hours of care per day of care (Figure S2C), suggestive of half-day care for some children, for consistency with contact surveys which count daily contacts regardless of their duration over a threshold (e.g. $>$ 15 minutes) we ignored this in our calculations.

To estimate the average number of children being cared for together in each setting ($n^{k,l}$), which determines an index child's total contacts with other young children while in that care setting, we use a different approach for home-based care and center-based care. For at-home care by a relative or non-relative, respondents in the ECPP were asked how many *other* children were present in their child's care setting. Survey responses were categorized into [None, 1-2, 3-5, 6 or more]. To obtain the average of the responses, we assume these responses correspond to average class sizes of [1, 2.5, 5, 7].

For children in center-based care, we assumed that centers were operating with age-specific groups of children (classes) and that all contacts occurred within these classes. We assumed that the age bands defining each class, as well as the number of students and caregivers per class, were informed by the state-specific regulations on minimum caregiver ratios and maximum class sizes by state-defined age groups ((32)) as well as federal recommendations (33).

Although there was substantial variation across states, we were able to choose a representative center structure that we believe represents well the approximate situation in centers across the country. We divided centers into fives classes: $l \in$ { 0-11 months (infant), 1 years old (younger toddler), 2 years old (older toddler), 3 years old (preschool), 4+ years old (pre-kindergarten) } (Table S4). For the infant and toddler classes, we assumed that children could move up classes after their birthdays but not before, and that there was some variation in the exact age they advance classes (e.g. due to space constraints at centers, use of developmental milestones like walking or potty training for advancement, and parental preferences), so that 25% move up during their first month in the new age group, another 25% in the second month, another 25% in the third month, and the final 25% in the fourth month. For the preschool and pre-kindergarten classes, we assume that centers switch to moving children up as a cohort based on their age by a defined start of the school year (e.g. Sept 1), so that the exact age at which children move up classes is evenly distributed across the months of that age group. In reality, the cut-off between age groups varies somewhat by state (for example, NY has separate ratios for infants <6 weeks; in IL, the infant designation continues until 15 months; in CA all children 2 and over are governed by the same rules). Within each age group, we chose to use the federal guidelines for the class sizes and caregiver-to-student ratios: 6 children per class for infants, 8 per class for toddlers (younger and older), 14 for preschoolers, and 16 for pre-kindergarten, with 2 teachers per class in each age group (Table S4). While some state-specific regulations allow much higher class sizes and caregiver:student ratios (e.g., Texas and North Carolina allow 20+ 2 year olds per class, Florida allows 40 pre-K students, and Georgia allows 6 infants per caregiver), we use the federal guidelines to provide a minimum bound for the number of contacts young children may have on average across the United States.

We next estimated the conditional age distribution of caregivers given they are in a care setting k, r_i^k. For children taken care of at home by a relative, ECPP provided the age of the caregiver (in years). The median age of the caregivers for at-home relative care was 60 years (interquartile range IQR: 52-65), with a range of 11 to 86. However, ECPP does not ask the specific age of at-home non-relative caregivers, instead providing a binary >18 or under 18 option: 98.9% are $\geq$ 18. For center-based care, no information on caregiver age is collected in the ECPP. Thus for both non-relative home care and center-based care we estimated the age distribution of caregivers using the NSECE. From this, we used summary statistics from the NSECE

medRxiv preprint doi: https://doi.org/10.64898/2026.08.18.26360623; this version posted August 21, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

(34, 35), where age is categorized for unlisted and listed home-care, and center-based care settings. We took the mean of unlisted and listed home-care as a representative value for our non-relative care setting, and the centre-based care setting. Using these data, we fit a smoothing spline (see Figure S1). Further, for the probability that an adult in age group i routinely works as a childcare provider of any type, we use data from the Bureau of Labor Statistics to obtain the age distribution of persons employed providing "child day care services".

We approximated the US age distribution by month by taking US Census data (37) and assuming that age is distributed equally over monthly age groups.

Additional assumptions of contact model

There are several additional assumptions and limitations to our approach to translating data from the ECPP survey into age-specific contact matrices for young children.

Based on standards established by large scale contact surveys (6) - which asked index individuals to name contacts they encountered in the prior day that met a particular threshold level of interaction - contact matrices typically count average 'daily contacts'. Thus, if two individuals are in contact only 3 out of 7 days per week, this counts as 0.43 daily contacts. Most children only attend childcare arrangements a subset of days of the week, which we assume reduces not only their contacts, but also the total number of other children they may encounter when with their caregiver. In essence, this is equivalent to assuming that children are equally likely to be attending childcare any day of the week, which has limitations. It's likely that the majority of children attend childcare only on the five weekdays, not weekends, and thus their days of attendance are correlated with those of other children, increasing their likelihood of receiving care on the same day and thus being in contact. Moreover, for licensed childcare centers and perhaps also some home caregivers, children accessing part-time care are enrolled in an organized way such that the center always has their maximum number of children. For example, if one child only comes Monday, Wednesday, and Friday, they may accept another child that only comes Tuesdays and Thursdays.

For children who use multiple types of care in a typical week (e.g., 1 day of childcare center and 1 day with a family member), we sum these days to get the total days. However, we may be overestimating total days in care, as an alternative scenario is that both types occur on the same day (i.e., a family member picks up a child from a childcare center to provide after-hours care at their home).

For home-based childcare, respondents in the ECPP were asked how many other children were present in their child's care setting. Survey responses were categorized into [None, 1-2, 3-5, 6 or more]. To obtain the average of the responses, we assume these responses correspond to average class sizes of [1, 2.5, 5, 7], which is imprecise. We could have alternatively estimated a parametric uni-modal distribution of class sizes and fit it to this data, but in the absence of more data went with the simplest approach.

In the creation of contact matrices for children in center-based childcare, we assumed all interactions occurred within the class structure we created. However, children may also be able to "contact" children from other classes - either by direct interactions in common areas like hallways, playgrounds, etc, or indirect interactions like shared airflow or use of contaminated objects. In addition, we have not taken into account the tremendous regional variation in regulations around the class size in childcare centers.

For home-based childcare, we have assumed that there is always a single caregiver for the entire group of children accessing care in that location, though in reality there could be multiple; for example, with cohabiting family members sharing in caregiving or with a home daycare center that has hired additional employees.

Finally, an unfortunate aspect of the design of the ECPP survey is that respondents did not actually report the exact age of the child at the time the survey was completed on their behalf. Instead, they reported birth month and year, but their individual completion date is not shared with any of the publicly-available data. For the analysis reported here, we assumed every survey was completed in August 2019 (in reality, the

medRxiv preprint doi: https://doi.org/10.64898/2026.08.18.26360623; this version posted August 21, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

survey was conducted between January and August 2019), the last possible date. This means our estimates of age are likely slightly overestimated, and that our estimates of total contacts for a given age are likely underestimated, since contacts tend to increase with age. This will have a larger effect on the assumed monthly age (and thus monthly contacts) of children less than one year of age. This is a major limitation of the current data. While survey receipt date is recorded in restricted-use datafiles, there is currently no way for researchers to request access to those.

Our method for calculating the average number of contacts between age groups in each childcare setting ($C^k_{i,j}$, (S3)) assumes that, on average, the index child i contacts others j according to the full age distribution of children in that care setting, which includes the index child themselves. This is a slight simplification, because in reality the age distribution of contacts should exclude the index individual. If each care setting has a small number of children and if children tend to be non-randomly sharing caregivers (e.g. if same age children vs age-spaced children are more likely that random to share a caregiver), then this assumption could be more important. However, properly accounting for this would require a much more complicated calculation in which the age distribution of contacts is calculated conditional on both index individual age and total number of children in the care setting, and the dataset is likely not large enough to expand the calculation in this manner and maintain reasonable numbers of observations.

To facilitate use in mathematical models, we combine our estimates for contacts in childcare settings with pre-existing contact matrices estimated for other settings (including household) (Figure S4). However, its possible we may be double-counting contacts among some siblings if they attend childcare together, as ECPP does specifically ask about whether the other children who share the care-giver are siblings or non-siblings. Future surveys designed for the purposes of quantifying total infection-relevant contacts should make sure to count only non-sibling peers in care.

Finally, in the analysis of ECPP conducted internally by the National Center for Education Statistics (NCES), individual data points are weighted in an attempt to account for biases in responses to the survey. We did not employ those weights in our analysis. Overall, NCES found a slight bias towards responses by individuals identifying as white, more highly educated, and with children over 1 year of age, though the absolute effect was small (see Ch. 7, 8, (50)).

Measles simulation model

We use a density-dependent age-stratified SIR model for the measles simulations,

$$\dot{S}_i = -\beta \sum_j S_i C_{ij} I_j \tag{S7}$$

$$\dot{I}_i = \beta \sum_j S_i C_{ij} I_j - \gamma I_i \tag{S8}$$

$$\dot{R}_i = \gamma I_i, \tag{S9}$$

where, $i = 1, \ldots, n$ are the number of age-groups, and S_i, I_i and R_i are the number of susceptible, infected and recovered individuals per-age group in the population. We assume the population size per-age group $N_i = S_i + I_i + R_i$ is fixed and the total population is $N = \sum_i N_i$. C is the age-mixing contact matrix with the element C_{ij} corresponding to the daily number of contacts between an individual of age i with an individual of age j. For simplicity we assume that the infectivity β (probability of transmission per infectious contact per day) and rate of recovery γ (probability per day an infected individual recovers) are not age-dependent. Using the next-generation matrix method (51), the basic reproduction number R_0 for the model is the leading eigenvalue of the matrix,

medRxiv preprint doi: https://doi.org/10.64898/2026.08.18.26360623; this version posted August 21, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

$$(FV^{-1})_{ij} = \frac{\beta}{\gamma} \sum_k S_i(t=0) C_{ij} \delta_{kj}, \tag{S10}$$

where δ is the Kronecker delta function. We fix R_0 and γ, and calculate β using Eq. S10. Note, this model can be written in a fully vectorized form as,

$$\dot{\mathbf{S}} = -\beta(\mathbf{S}.\mathbf{E})\mathbf{C}\mathbf{I} \tag{S11}$$

$$\dot{\mathbf{I}} = \beta(\mathbf{S}\cdot\mathbf{E})\mathbf{C}\mathbf{I} - \gamma\mathbf{I} \tag{S12}$$

$$\dot{\mathbf{R}} = \gamma\mathbf{I}, \tag{S13}$$

where, $\mathbf{S}, \mathbf{I}, \mathbf{R}$ are $(n \times 1)$ vectors, $\mathbf{C}$ is the $n \times n$ contact matrix, and $\mathbf{E}$ is a $(n \times 1)$ vector where each entry is one of the basis matrices of the identity matrix; for example if $n = 2$,

$$\mathbf{E} = \begin{pmatrix} \mathbf{E_1} \\ \mathbf{E_2} \end{pmatrix}$$

$$\mathbf{E_1} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \mathbf{E_2} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Re-writing the model in this form makes it easy to generalize it for any number of age-groups. We implement a stochastic version of this model using the Tau-Gillespie algorithm (52) in Python using JAX-NumPy (53) to optimize the speed of the simulations.

Links to Early Childhood Participation Program survey 2019

- Questionnaire
- Reports on data
- Data File Users' Manual
- Public-Use Data (login required)
- Restricted-Use Data

medRxiv preprint doi: https://doi.org/10.64898/2026.08.18.26360623; this version posted August 21, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Supplementary Figures

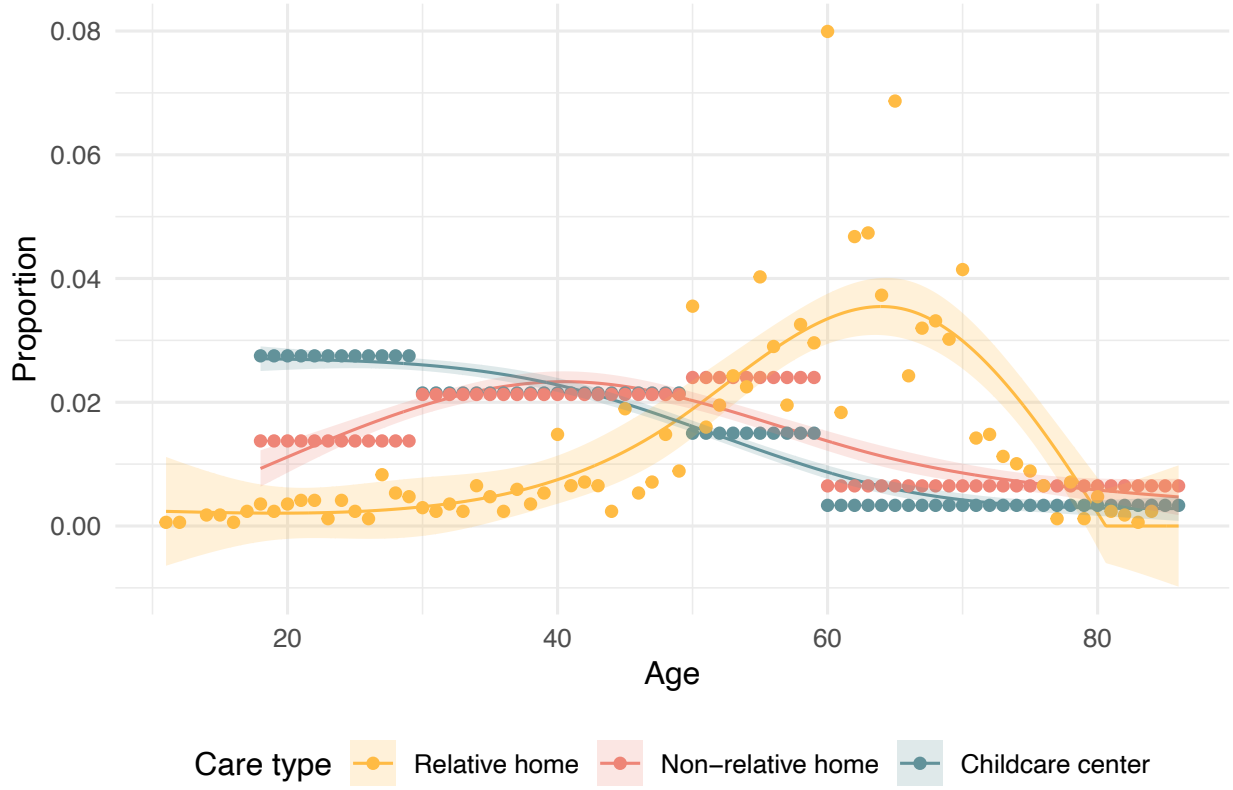


Figure S1: Age distribution of caregivers of young children. The age distribution of caregivers in each care type. The age of at-home relative caregivers is taken from the ECPP (31), and that of at-home non-relative caregivers and center-based caregivers are taken from the NSECE (34, 35).

medRxiv preprint doi: https://doi.org/10.64898/2026.08.18.26360623; this version posted August 21, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

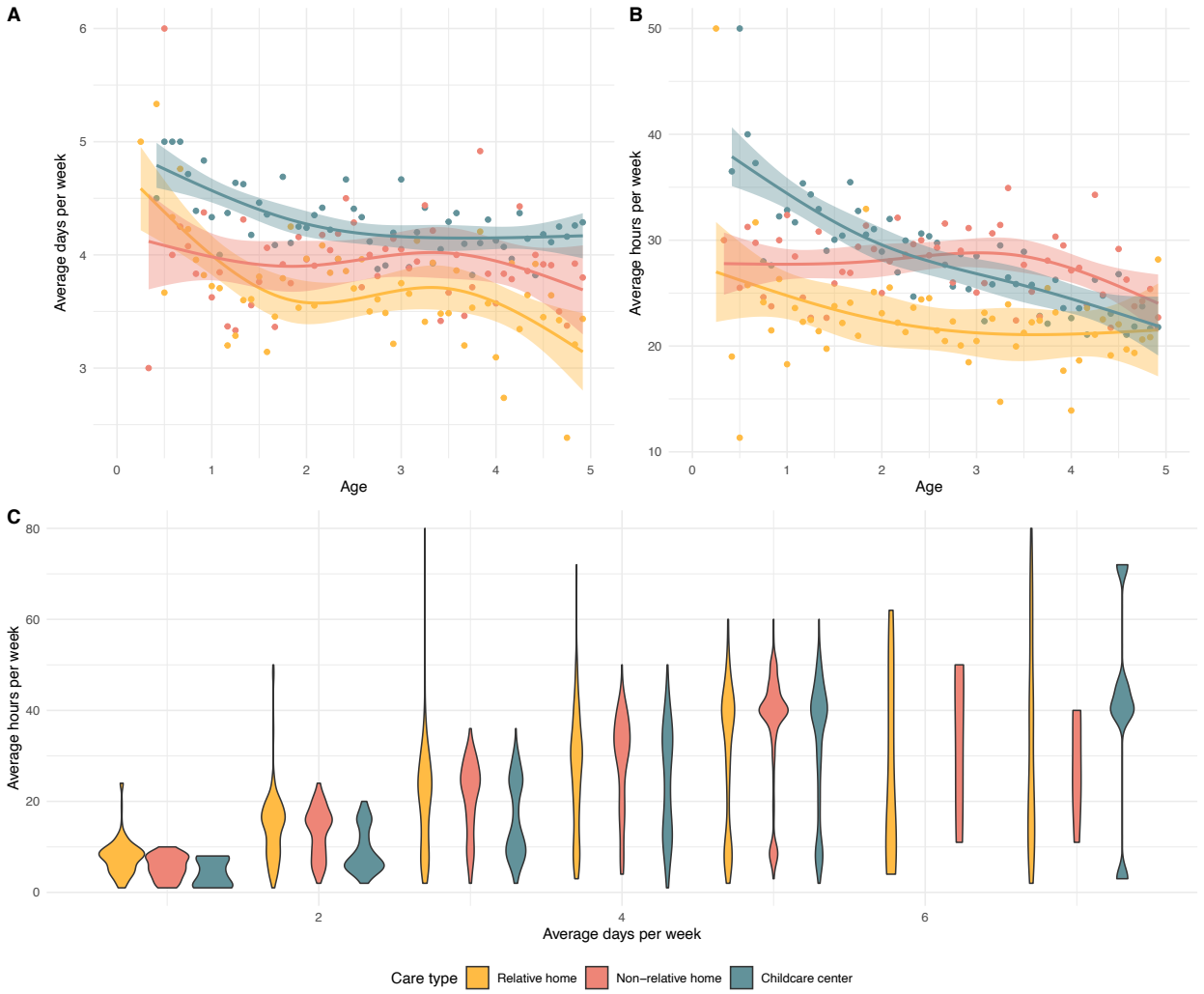


Figure S2: Time spent by young children in different childcare settings. From Early Childcare Participation Program survey, the (A) average days per week, and (B) average hours per week by age, with a generalised linear model fit with three degrees of freedom for each care type. (C) The average hours per week in care as a function of the average days per week, showing bimodal distribution for some days, indicating half-days of care.

medRxiv preprint doi: https://doi.org/10.64898/2026.08.18.26360623; this version posted August 21, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

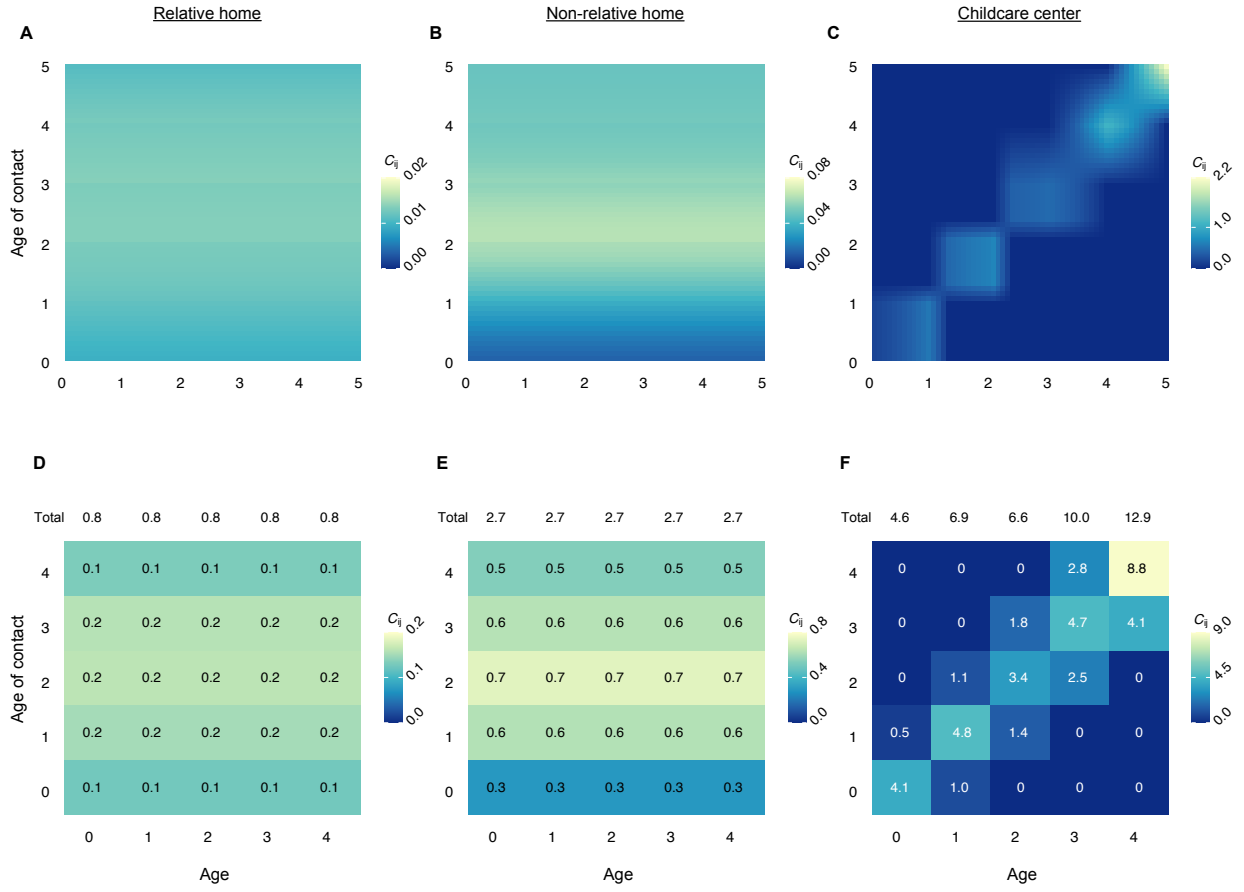


Figure S3: Setting-specific contact matrix for young children in childcare. The average daily number of contacts for index individual of age i (columns) with contact of age j (rows), for (A, D) In-home care by a relative, (B, E) In-home care by a non-relative, and (C, F) Childcare center. The first row of panels (A-C) are setting-specific contact matrices at a monthly age-scale, and the second row of panels (D-F) show the corresponding yearly matrices, with the total contacts each index individual has with all others under 5 in that setting shown in the top row of each panel.

medRxiv preprint doi: https://doi.org/10.64898/2026.08.18.26360623; this version posted August 21, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

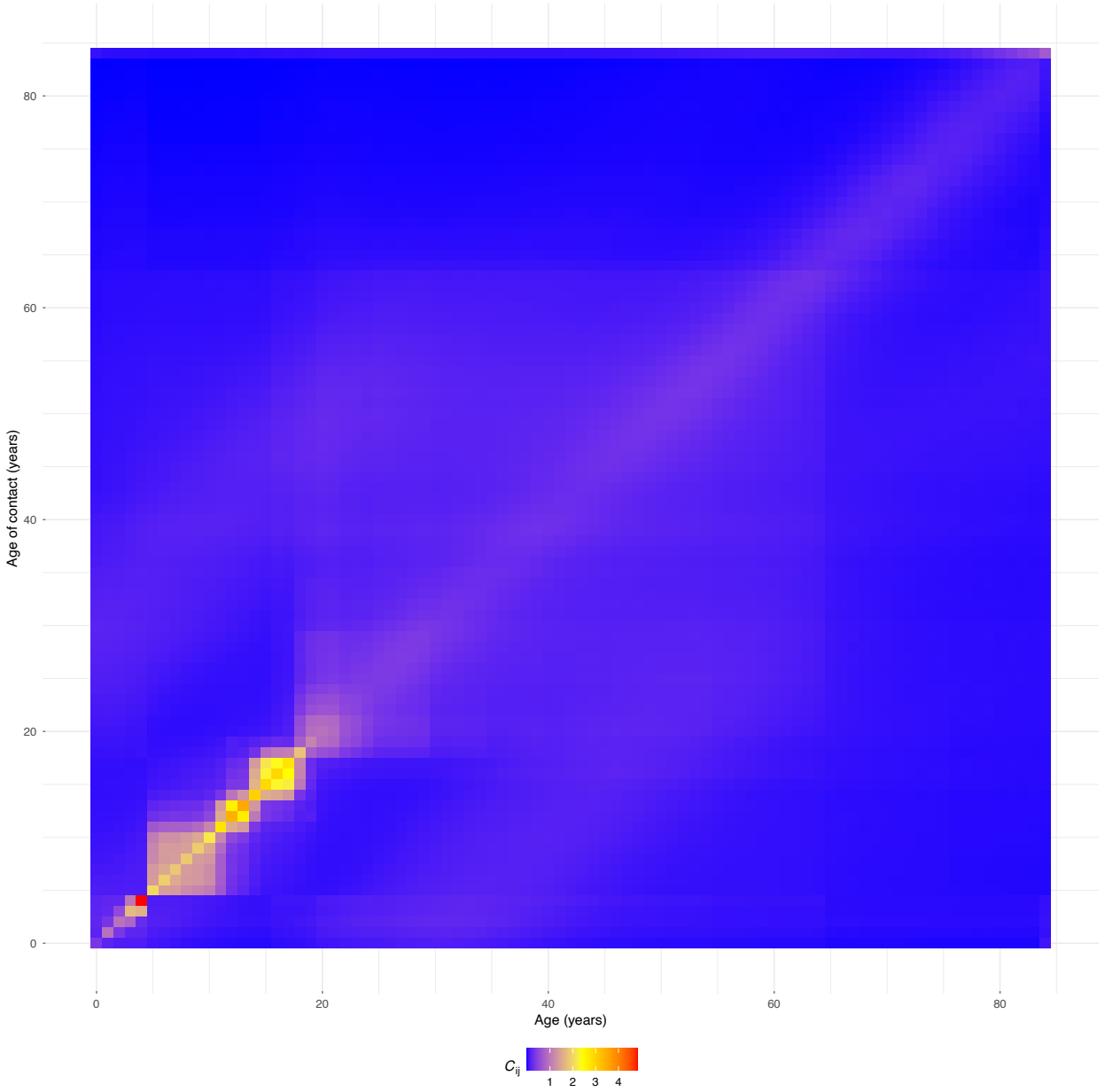


Figure S4: All-age all-setting contact matrix augmented with childcare contacts. The all-setting contact matrix for the US presented by Mistry et al. (13), which includes imputed contacts occurring at home, work, school, and community, was augmented by first removing "school" setting contacts for all individuals under 5, and then adding the childcare setting contacts estimated in this paper. This augmentation includes contacts occurring between children aged 0-4 and childcare providers.

medRxiv preprint doi: https://doi.org/10.64898/2026.08.18.26360623; this version posted August 21, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

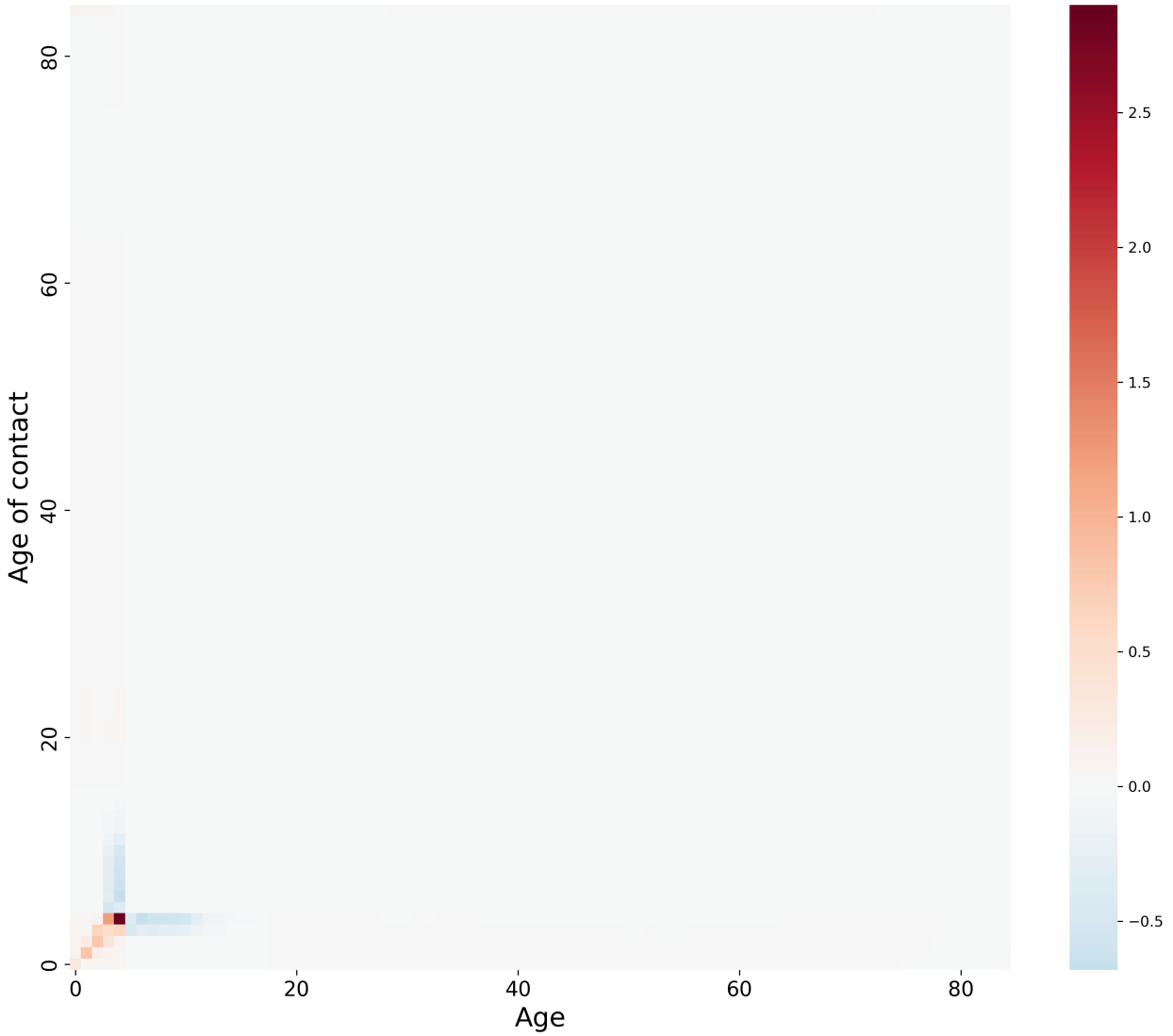


Figure S5: Difference between the augmented all-age all-setting contact matrix with the one presented in Mistry et al. Augmented contact matrix minus the one presented in Mistry et al. Red and blue cells correspond to increase and decrease in contacts in the augmented matrix compared to Mistry et al. respectively.

medRxiv preprint doi: https://doi.org/10.64898/2026.08.18.26360623; this version posted August 21, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

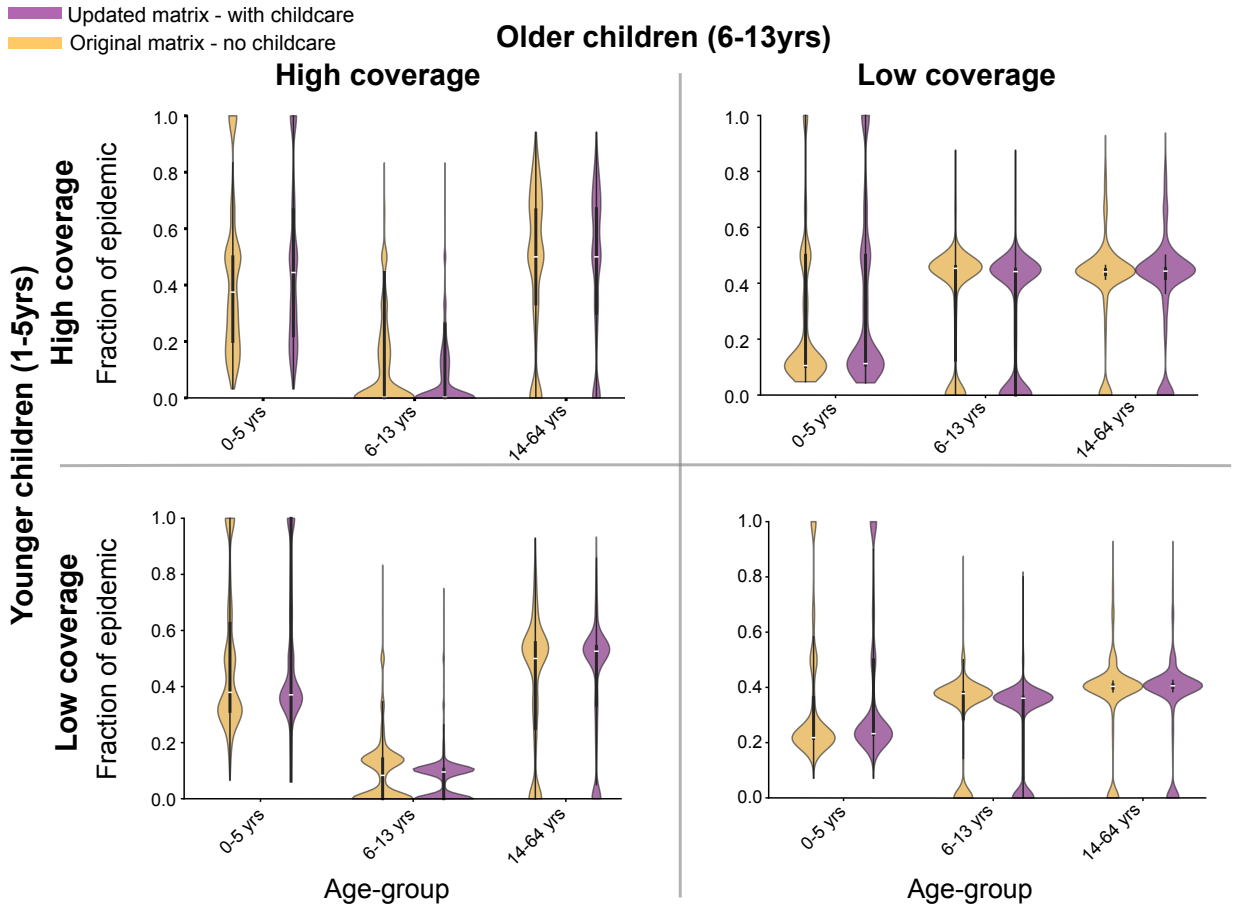


Figure S6: Age-group contribution towards the epidemic for the measles simulation case study across vaccination scenarios. Each of the four panels contains violin plots of the fraction of the total epidemic size belonging to each age-group (0-5yrs, 6-13yrs, 14-64yrs) for iterations where onward transmission occurred after 1 infection was introduced for both the original Mistry et al (orange) and updated (purple) contact matrices. Simulation results are for 10,000 iterations in a population of size 100,000.

medRxiv preprint doi: https://doi.org/10.64898/2026.08.18.26360623; this version posted August 21, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

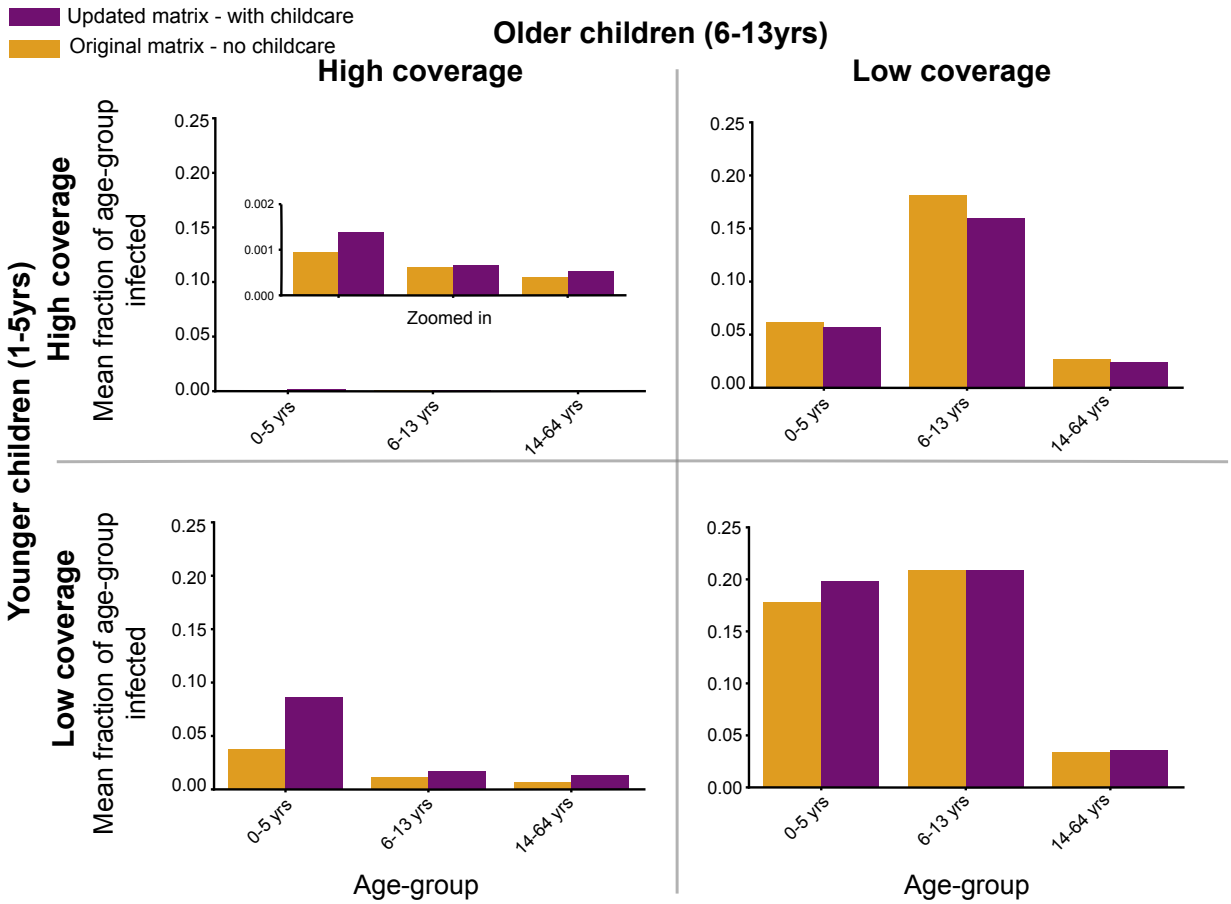


Figure S7: Mean age specific attack rates in the measles simulation case study across vaccination scenarios. Each of the four panels contains bar graphs whose height is the mean of the fraction of the age-groups (0-5yrs, 6-13yrs, 14-64yrs) infected for iterations where onward transmission occurred after 1 infection was introduced for both the original Mistry et al (orange) and updated (purple) contact matrices. Simulation results are for 10,000 iterations in a population of size 100,000.

medRxiv preprint doi: https://doi.org/10.64898/2026.08.18.26360623; this version posted August 21, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Supplementary Tables

Table S1: Parameter definitions, units and data sources for calculating average number of contacts on a given day of care.

Parameter	Description	Units	Source
$C_{i,j}$	Average contacts that each child of age i has with children of age j	# of contacts per day	Calculated
$C^k_{i,j}$	Average contacts that each child of age i has with children of age j in childcare type k	# of contacts per day	Calculated
p^k_i	Probability child in age group i routinely receives childcare of type k on any given day	[0,1]	Early Childhood Program Participation (ECPP) Survey 2019 (31)
f^k_i	Average fraction of days per week child of age i spends in childcare type k	[0,1]	Early Childhood Program Participation (ECPP) Survey 2019 (31)
ρ^k_i	Proportion of children in age group i who receive childcare of type k regularly	[0,1]	Early Childhood Program Participation (ECPP) Survey 2019 (31)
n^k	Average number of children in childcare type k	# kids	Childcare.gov Supervision: Ratios and Group Sizes (Recommendations), 2025 (33)
$q^{k,l}_i$	Conditional probability child in age group i who receives childcare of type k is in class l	[0,1]	
a_i	Fraction of the entire population in age group i	[0,1]	United States Census Bureau, National Population by Characteristics, 2025 (54)
$n^{k,l}$	Average number of children per class l in childcare type k	# kids	Childcare.gov Supervision: Ratios and Group Sizes (Recommendations), 2025 (33)
$\sigma^{k,l}$	Caregiver:Student ratio in class l in care type k	# adults / # kids	Childcare.gov Supervision: Ratios and Group Sizes (Recommendations), 2025
r^k_i	Probability that an adult who works as carer of type k is age i	[0,1]	National Survey of Early Care and Education (NSECE) 2019 (34, 35)
n^k_c	Average number of caregivers per childcare type k	# adults	Childcare.gov Supervision: Ratios and Group Sizes (Recommendations), 2025 (33)
$n^{k,l}_c$	Average number of caregivers per class l in childcare type k	# adults	Childcare.gov Supervision: Ratios and Group Sizes (Recommendations), 2025 (33)

medRxiv preprint doi: https://doi.org/10.64898/2026.08.18.26360623; this version posted August 21, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Table S2: Summary of contact matrix entries. Formulae used for calculating average number of contacts between index individual of age i with contact of age j on a given day of care in care type k and setting l (if relevant, for program care). Parameters are defined in Table S1.

Childcare type, k	Index	Contact	$C^k_{i,j}$	$C^{k,l}_{i,j}$
None	-	-	0	0
Home care	Young child	Young child	$\left(\frac{p^k_j a_j}{\sum_m p^k_m a_m}\right) \cdot (n^k - 1)$	-
Home care	Young child	Adult	$r^k_j \cdot n^k_c$	-
Home care	Adult	Young child	$p^k_j n^k$	-
Home care	Adult	Adult	0	-
Program care	Young child	Young child	$\sum_l q^{k,l}_i C^{k,l}_{i,j}$	$\left(\frac{q^{k,l}_j p^k_j a_j}{\sum_m q^{k,l}_m p^k_m a_m}\right) \cdot (n^{k,l} - 1)$
Childcare center	Young child	Adult	$\sum_l q^{k,l}_i C^{k,l}_{i,j}$	$r^k_j n^{k,l}_c$
Childcare center	Adult	Young child	$\sum_l C^{k,l}_{i,j}$	$\left(\frac{q^{k,l}_j p^k_j a_j}{\sum_m q^{k,l}_m p^k_m a_m}\right) \cdot n^{k,l}$
Childcare center	Adult	Adult	$\sum_l C^{k,l}_{i,j}$	$r^k_j (n^{k,l}_c - 1)$

Table S3: Characteristics of childcare settings. Attendance statistics, class sizes and caregiver:student ratios from Early Childcare Program Participation study (31) and federal guidelines (33). These values are used to construct contact matrices, $C^{k,l}_{i,j}$.

Childcare type k	Attendance (days/week) (fraction, f^k_i)	Class l	Class size $n^{k,l}$	Caregiver:Student $\sigma^{k,l}$
None	-	-	0	0
Relative home	3.7 (0.53)	-	1.8	1:1.8
Non-relative home	3.9 (0.56)	-	3.9	1:3.9
Childcare center	4.3 (0.62)	Infants, 0-11 months	6	1:3
		Younger toddlers, 1 year	8	1:4
		Older toddlers, 2 years	8	1:4
		Preschool, 3 years	14	1:7
		Pre-Kindergarten, 4 years	16	1:8

medRxiv preprint doi: https://doi.org/10.64898/2026.08.18.26360623; this version posted August 21, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Table S4: Recommended child size from Office of Child Care (33).

Care type	Class	Recommended Size	Caregiver:Student ratio
Child Care Center	Infants <12 months	6	1:3
	13-35 months	8	1:4
	3 years	14	1:7
	4 years	16	1:8
	5 years	16	1:8
Large family child care homes	Infants <12 months	6	1:2
	13-23 months	8	1:2
	24-35 months	12	1:3
	3 years	12	1:7
	4-5 years	12	1:8

Table S5: Questions used from ECPP survey (31). Questions relevant for each care type, and the survey responses allowed within the survey.

Question	Care type	Responses	ECPP Code
Is this child now receiving care from a relative other than a parent or guardian on a regular basis, for example, from grandparents, brothers or sisters, or any other relatives?	Relative home	Yes/No	RCNOW
About how many days each week does this child receive care from this relative?	Relative home	Integers, 1-7	RCDAYS
How many other children does this relative care for while caring for this child?	Relative home	None, 1-2, 3-5, ≥ 6	RCOTCH
Is this child now receiving care in your home or another home on a regular basis from someone who is not related to him or her?	Non-relative home	Yes/No	NCNOW
About how many days each week does this child receive care from this person?	Non-relative home	Integers, 1-7	NCDAYS
How many other children does this provider care for while caring for this child?	Relative home	None, 1-2, 3-5, ≥ 6	NCOTCH
Is this child now attending a day care center, preschool, or prekindergarten not in a private home?	Childcare center	Yes/No	CPNNOWX
How many days each week does this child go to this program?	Childcare center	Integers, 1-7	CPDAYS

medRxiv preprint doi: https://doi.org/10.64898/2026.08.18.26360623; this version posted August 21, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Supplementary References

[48] Bureau USC. American Community Survey (ACS). United States Census Bureau; 2026. https://www.census.gov/programs-surveys/acs.html.

[49] Stehlé J, Voirin N, Barrat A, Cattuto C, Isella L, Pinton JF, et al. High-Resolution Measurements of Face-to-Face Contact Patterns in a Primary School. PLOS ONE. 2011 Aug;6(8):e23176. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0023176.

[50] Jackson M, Kaiser A, Battle D, Wan C, Quenneville G, Kincel B, et al. National Household Education Surveys Program of 2019: Data File User's Manual. U.S. Department of Education, Washington, DC: National Center for Education Statistics; 2021.

[51] Heffernan JM, Smith RJ, Wahl LM. Perspectives on the basic reproductive ratio. Journal of The Royal Society Interface. 2005 Sep;2(4):281-93.

[52] Gillespie D. Approximate Accelerated Stochastic Simulation of Chemically Reacting Systems. J Chem Phys. 2001 07;115:1716-33.

[53] Bradbury J, Frostig R, Hawkins P, Johnson MJ, Katariya Y, Leary C, et al.. JAX: composable transformations of Python+NumPy programs; 2018. Available from: http://github.com/jax-ml/jax.

[54] Bureau USC. National Population by Characteristics: 2020-2025; 2026. https://www.census.gov/data/datasets/time-series/demo/popest/2020s-national-detail.html.