

Characterising non-household contact patterns relevant to respiratory transmission in the USA: analysis of a cross-sectional survey

Juliana C Taube, Zachary Susswein, Vittoria Colizza, Shweta Bansal



Summary

Background Interpersonal contact has a crucial role in the transmission of infectious diseases. Characterising heterogeneity in contact patterns across individuals, time, and space is necessary to inform accurate estimates of transmission risk, particularly to explain superspreading, predict differences in vulnerability by age, and inform physical distancing policies. Current respiratory disease models often rely on data from the 2008 POLYMOD study conducted in Europe, which is now outdated and is potentially unrepresentative of behaviour in other geographical regions. We aimed to understand the variation in contact patterns in the USA across time, spatial scales, and demographic and social classifications during the COVID-19 pandemic, and to estimate what social behaviour looks like at baseline, in the absence of an ongoing pandemic.

Methods For this study of contact patterns relevant to respiratory transmission during a pandemic, we examined 10·7 million responses to the US COVID-19 Trends and Impact Survey between June 1, 2020, and April 30, 2021 (ie, during the COVID-19 pandemic); the survey recruited participants aged 18 years and older in the USA through Facebook. Data were post-stratified by age and gender to correct for sample representation. We used generalised additive models to characterise spatiotemporal heterogeneity in respiratory contact patterns during the pandemic at the county-week scale; we established how contact patterns vary by urbanicity, age (18–54 years, 55–64 years, 65–74 years, or ≥75 years), gender (male or female), race or ethnicity (Asian, Black or African American, Hispanic, White, or other), and contact setting (work, shopping for essentials, social gatherings, or other). We used a regression approach to estimate baseline (non-pandemic) contact patterns.

Findings Although contact patterns varied over time during the COVID-19 pandemic, the average number of daily contacts was relatively stable after controlling for the effect of incidence-mediated risk perception and disease-related policy. The mean number of non-household contacts was spatially heterogeneous, varying across urban versus rural settings, regardless of the presence of disease. Additional heterogeneity was observed across age, gender, race or ethnicity, and contact setting. Mean number of contacts decreased with age for individuals older than 55 years and was lower in women than in men. During periods of increased national incidence of disease, the contacts of White individuals and contacts at work or social gatherings showed the greatest change.

Interpretation Our findings indicate that US adult baseline contact patterns show little variability over time after controlling for disease, but high spatial variability regardless of disease, with implications for understanding the seasonality of respiratory infectious diseases. The highly structured spatiotemporal, demographic, and social heterogeneity in contact patterns reported here could inform the risk landscape of respiratory infectious disease transmission in the USA and the implementation of targeted interventions, and our county-level estimates of non-pandemic contact rates could fill gaps in parameterising future disease models.

Funding National Institutes of Health, Agence Nationale de la Recherche, and EU Horizon Europe.

Copyright © 2025 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

The transmission of respiratory infectious diseases via direct or droplet routes requires close contact. Research over the past two decades has shown that human contact patterns are highly variable between individuals and across geography,^{1,2} and highlighted the consequences of this variability for epidemic outcomes and dynamics.³ However, to date, detailed empirical data on contact patterns across the USA have not been available. This lack of data leaves

several important gaps in our understanding of the drivers of disease transmission. For example, knowing what factors (eg, age, season, or location) influence contact patterns across the USA, and how these factors contribute to variability in infection risk, is essential to design targeted interventions and to generate accurate estimates of transmission risk across individuals, space, and time.

Most infectious disease models assume homogeneous mixing among individuals—ie, all individuals have the

Lancet Digit Health 2025

Published Online
<https://doi.org/10.1016/j.lindig.2025.100888>

Department of Biology,
 Georgetown University,
 Washington, DC, USA
 (J C Taube AB, Z Susswein BS,
 V Colizza PhD, Prof S Bansal PhD);
 INSERM, Sorbonne Université,
 Institut Pierre Louis
 d'Épidémiologie et de Santé
 Publique, Paris, France (V
 Colizza)

Correspondence to:
 Prof Shweta Bansal, Department
 of Biology, Georgetown
 University, Washington,
 DC 20057, USA
shweta.bansal@georgetown.edu

Research in context**Evidence before this study**

We searched Google Scholar for social contact data in the USA covering time periods during and before the COVID-19 pandemic, published in English from database inception to Feb 1, 2024, using the search terms “contact patterns”, “social contact data”, “disease-relevant contacts”, “change in contacts pandemic”, “urban rural social contacts”, and “seasonality in contact patterns”. We reviewed the bibliographies of the retrieved articles and included known literature not found through our searches. We excluded studies that used mobility data, focused on children, or excluded the USA. Previous work has been limited to the state scale or to subsets of counties (eg, focused on a few cities, a single state, or a few counties within a state) rather than including all counties in the USA.

Added value of this study

To our knowledge, we contribute the first high-resolution contact estimates for the USA during the COVID-19 pandemic and infer non-pandemic contact patterns at fine spatial and temporal scales. Our results indicate that the number of contacts per person per day is fairly stable over time after controlling for the effect of incidence-mediated risk perception and disease-related

policy (ie, in the absence of major disease), suggesting that the number of contacts is not a primary driver of respiratory infectious disease seasonality in the USA. We also identify groups at the greatest risk of disease owing to their high number of contacts, including younger adults (aged 18–54 years), men, Hispanic individuals, and Black individuals.

Implications of all the available evidence

This study shows the importance of incorporating age-specific contact patterns and the spatial heterogeneity of contact patterns into future disease models to build accurate estimates of transmission risk. We show that temporal variability in contact patterns is insufficient to drive the seasonality of respiratory infectious disease, that adaptive behaviours in response to disease shift risk along an urban-rural gradient, and that some vulnerable groups are at increased risk of exposure owing to their increased numbers of contacts. We advocate that geographical and social heterogeneity in exposure to disease as a result of differing contact patterns be captured more comprehensively to facilitate accurate predictions of infectious disease dynamics and effective and equitable disease mitigation.

same contact rate and the same ability to transmit disease. Homogeneous mixing models have transformed the prediction and control of disease outbreaks (eg, through estimation of the time-varying effective reproduction number⁴), but produce epidemic dynamics and outcomes that differ from those of models that incorporate heterogeneities in contact patterns.^{3,5} The POLYMOD study, published in 2008, was the first extensive survey to characterise heterogeneities in routine contact patterns relevant to the transmission of respiratory infectious diseases. The data, from eight European countries, suggested that individual contact rates are not homogeneous but have a heavy-tailed distribution and are highly assortative by age.¹ POLYMOD estimates have been used to understand epidemic dynamics, design vaccine strategies, and predict intervention outcomes;^{6–10} however, they did not capture dynamic behaviour in the presence of disease. To fill this gap, the CoMix study was conducted during the COVID-19 pandemic to collect contact data contemporaneous with disease across Europe;^{11,12} estimates from the study were incorporated into forecasting efforts in the UK, with mixed results.¹³ Notably, neither survey captured fine-scale temporal variability or social heterogeneity in contact, and they might not be representative of behaviour in the USA.

Beyond individual heterogeneity, there are several meaningful dimensions across which contact patterns might vary. Spatial heterogeneity in human behaviour has a crucial role in disease dynamics;^{14–17} if contact patterns exhibit such heterogeneity, this could explain observed hotspots of disease burden or the dynamics of disease spread. Likewise, changes in contact patterns over time

could contribute to the seasonality of respiratory infectious diseases.^{18,19} Contact might also vary with surrounding disease transmission, as individuals shift their behaviour to mitigate risk. These potential heterogeneities in contact patterns profoundly affect our understanding and prediction of epidemic dynamics and our ability to target behavioural interventions; however, few empirical data exist on such heterogeneities. To identify the dimensions across which contact meaningfully varies, high-resolution contact data are needed across geography, time, demography, and social classifications and in the context of disease transmission.

Contact patterns in the USA have gained increasing attention over the past 5 years. In addition to small or indirect earlier studies,^{20,21} several more recent studies have provided insights into variations in contact patterns across multiple dimensions. Breen and colleagues²² showed that pandemic contact patterns in the USA vary between states, but they were unable to characterise spatial heterogeneity at finer scales, even though other public health-related behaviours have been found to vary at the county level.^{23,24} Dorelien and colleagues²⁵ used time-use surveys to show that baseline contact patterns in urban and rural areas might not differ, despite the perception that urban inhabitants have more contacts. Regarding temporal variation in contact, one study conducted before the COVID-19 pandemic found no variation in adult contacts over time,²⁵ whereas another observed changes over the study period of September to May.²⁰ Many studies have found that contact is higher in younger adults (typically 18–44 years old), men, and non-White populations, although which race or

ethnicity group had the highest contact rate varied by study.^{20,26–28} Because these previous studies were limited in terms of sample size or resolution, they are constrained in their ability to comprehensively characterise heterogeneities in contact patterns across space, time, and social groups.

Now that the US COVID-19 Public Health Emergency has ended, it is necessary to characterise contact patterns not only during the pandemic but also under non-pandemic, baseline conditions. In general, physical distancing during the pandemic reduced overall contact.²⁹ However, previous work has shown that adherence to physical distancing guidelines was heterogeneous across populations, driven in part by health disparities and social inequities that affect the ability of individuals to engage in behavioural interventions³⁰ and by a patchwork of non-pharmaceutical interventions (eg, school closures, working from home, gathering bans, and mask requirements) implemented at various spatial scales. For example, urban areas showed greater reductions in mobility and, therefore, probably greater reductions in contact than rural areas.³¹ Individuals of higher socioeconomic status are known to have had greater flexibility in their mitigation behaviour and could further reduce contacts. Understanding which groups and locations are at the highest risk of infection during pandemics and seasonal epidemics is crucial for targeted public health surveillance and resource allocation, but requires detailed contact data disaggregated by location, age, gender, and race or ethnicity.

Here, we aimed to address these gaps by developing fine-scale spatiotemporal estimates of mean non-household contacts in adults. We used an extensive national survey with more than 10 million responses collected at the county level to characterise several heterogeneities in contact patterns throughout the pandemic and infer non-pandemic contact patterns. We focused on four central questions: How does mean contact vary over time? How does mean contact vary across geography in urban versus rural settings? How do contact patterns vary across age, gender, and race or ethnicity classifications? What are contact patterns like under non-pandemic, baseline conditions? Our results are, to our knowledge, the most comprehensive high-resolution estimates of US contact patterns to date, and could inform future disease models in the USA, provide insight into local and temporal variation in behaviour in response to public health messaging, and contribute to our understanding of the drivers of respiratory infectious disease seasonality.

Methods

Study design

This study was designed with the aim of characterising heterogeneities in non-household contact patterns relevant to respiratory infectious disease transmission across time, spatial scales, and demographic and social classifications during a pandemic and at baseline in the USA. We analysed

data from the US COVID-19 Trends and Impact Survey (CTIS)³²—a large, cross-sectional survey conducted in all 50 states of the USA and the District of Columbia—from June 1, 2020, to April 30, 2021. We focused on non-household contacts because they have a crucial role in the dynamics of casual contact infections.³³ To explore variability in contact patterns over time, particularly during the autumn to spring period when seasonal changes are expected,³⁴ we developed two statistical models. In the first model, we address noise and representation issues in the data to estimate pandemic contact patterns. In the second model, rather than capturing all sources of heterogeneity, we aim to parsimoniously account for pandemic-related effects to estimate non-pandemic contact patterns. We hypothesised that if temporal variability in contact is fully explained by disease incidence-mediated risk perception and intervention-related changes in behaviour, then baseline behaviour should be temporally stable. This study was reviewed by the Institutional Review Board at Georgetown University (Washington, DC, USA) and was determined not to be human subjects research.

Survey data

The CTIS was created by the Delphi Group at Carnegie Mellon University (Pittsburgh, PA, USA) to monitor the spread and impact of the COVID-19 pandemic in the USA, and was distributed through a partnership with Facebook. Beginning on April 6, 2020, a random, state-stratified sample of active Facebook users aged 18 years and older were invited daily to take the survey about COVID-19 and report how many people they had direct contact with outside their household. The survey used a well established definition of contact relevant to respiratory disease transmission—"a conversation lasting more than 5 minutes with a person who is closer than 6 feet away from you, or physical contact like hand-shaking, hugging, or kissing"—and therefore offered an advantage over GPS location-based mobility data to characterise contact patterns; previous work has also shown that social contact data are more predictive of transmission than mobility data.^{35–37} Contacts were disaggregated by settings outside the home in the survey question: work, shopping for groceries or other essentials, social gatherings, or other. We analysed the sum of contacts across all settings starting on June 1, 2020 (to capture stabilised pandemic patterns beyond the acute disruption during spring 2020). We removed responses from individuals with more than 72 contacts in the past 24 h (the 95th percentile) and conducted sensitivity analyses of this truncation point (appendix pp 33–36).

Age, gender (male, female, non-binary, prefer to self-describe, or prefer not to answer), and race or ethnicity (American Indian or Alaska Native, Asian, Black or African American, Hispanic, Native Hawaiian or Pacific Islander, White, or multiple or other race) were self-reported in the survey, and the survey data are representative of the overall US population in terms of age, sex, and race or ethnicity (appendix p 45). Race or ethnicity was categorised into Asian, Black or African American, Hispanic, White, or

See Online for appendix

other for the analysis; all categories except Hispanic were considered non-Hispanic. To adjust for unrepresentative sampling at the county scale, we generated response weights to match county age and sex distributions and post-stratified the data (appendix p 1). To conduct raking with American Community Survey data, which is based on sex, we had to assume that sex and gender were equivalent and binary (which comprises 98% of the data). Owing to limited sample size and inappropriate entries, we could not consider responses with genders outside of this binary construct, which we recognise as a limitation. As data on race or ethnicity were not available throughout the full study period (these data were collected only from Sept 8, 2020 to the end of the study period), we could not include them in the raking procedure for all estimates. To test the sensitivity of our estimates to the inclusion of race or ethnicity in the raking weights, we compared mean contact estimates at the state-month level for Sept 8, 2020, to April 30, 2021 (appendix p 29). We additionally account for race or ethnicity in the raking weights for any analyses specific to these characteristics.

County urbanicity was determined using the 2013 National Center for Health Statistics (NCHS) Urban–Rural Classification Scheme for Counties, in which class 1 indicates a large central metropolitan area and class 6 represents rural non-core areas.³⁸

Spatiotemporal contact estimation

To address weekly noise and low sample sizes in the data, we estimated smoothed county-week mean non-household contacts per person per day with hierarchical generalised additive models. Such models, which fit arbitrary smooth curves to data,³⁹ are a popular form of regression, especially for time series modelling.^{40,41} The mgcv R package for fitting generalised additive models was designed for flexible and efficient estimation with structured time series data,^{39,42,43} accounting for partial observations and dropout, and literature on the application of mgcv in this area is robust.

We use penalised thin-plate splines weighted by sample size with a weekly smooth for each state and a factor smooth for each county within the state, with shared smoothness parameters across counties. Information is therefore shared within but not between states; estimates are penalised to the state average, which could be biased towards urban counties. We chose thin-plate regression splines because they have appealing theoretical properties: they have the best theoretical performance in terms of minimising mean squared error relative to the objective function and, unlike cubic regression splines, are knot-free.³⁹ Thin-plate splines do not require user-specified knot positions, instead requiring only sufficient basis dimension.

The model is structured as follows:

$$c_{i,t} = \beta_0 + f_s(t) + f_i(t) + \varepsilon_{i,t}$$

where

$$f_j(t) = \sum_{k=1}^K \beta_{j,k} b_{j,k}(t),$$

and $c_{i,t}$ represents mean contact for each county i and week t . The smooth function f_j is a penalised thin-plate regression spline, with j as state s or county i level,⁴² and $\varepsilon_{i,t}$ is a Gaussian error term. The coefficients $\beta_{j,k}$ are estimated for each of the K basis functions $b_{j,k}$.

We fit separate generalised additive models to estimate contact by age (18–54 years, 55–64 years, 65–74 years, or ≥75 years), gender (male or female), race or ethnicity (Asian, Black or African American, Hispanic, White, or other), and contact setting (work, shopping for essentials, social gatherings, or other) at the county (or state, for race or ethnicity) level during the pandemic period; for additional details, see the appendix (pp 1–3).

Baseline contact estimation

To infer contact patterns in the absence of a pandemic, we used a linear regression model of weekly contact (as estimated using the generalised additive models) predicted by national case incidence,⁴⁴ state-level policy stringency (measured via the Oxford Stringency Index⁴⁵), county-level policy stringency (calculated as the sum of various ordinal county policies^{46–52}), and percentage of the county vaccinated against COVID-19²⁴ (appendix p 26). We propose that changes in behaviour are driven by how people perceive risk, and that this perception is largely shaped by reported disease incidence. To test this, we compared a subjective measure of perceived risk (from a different CTIS question) and found that disease incidence was a stronger predictor of behaviour than self-reported risk (appendix p 46). We also considered the role of incidence information across scales (county, state, and national), finding that state and national incidence were highly collinear, and county incidence was less informative than national incidence (appendix p 46). The Oxford Stringency Index captures the extent of policy limitations on human behaviour, including school, work, and transit closures; stay-at-home requirements; and travel restrictions⁴⁵ at the state level. The county policy variable comprised county-level restaurant and bar closures, mask mandates, gathering bans, and stay-at-home orders; for additional sensitivity analyses related to the policy variables, see the appendix (pp 38–39). We hypothesised that increases in disease incidence would lead to a linear decrease in contact rates, which is supported by the survey data (appendix p 24). The model is as follows:

$$\begin{aligned} c_{i,t} = & \beta_{0,i} + \beta_{1,i} d_{n,t} + \beta_{2,i} v_{i,t} + \beta_{3,i} p_{s,t} + \beta_{4,i} p_{i,t} + \\ & \beta_{5,i} d_{n,t} \times v_{i,t} + \varepsilon_{i,t} \end{aligned}$$

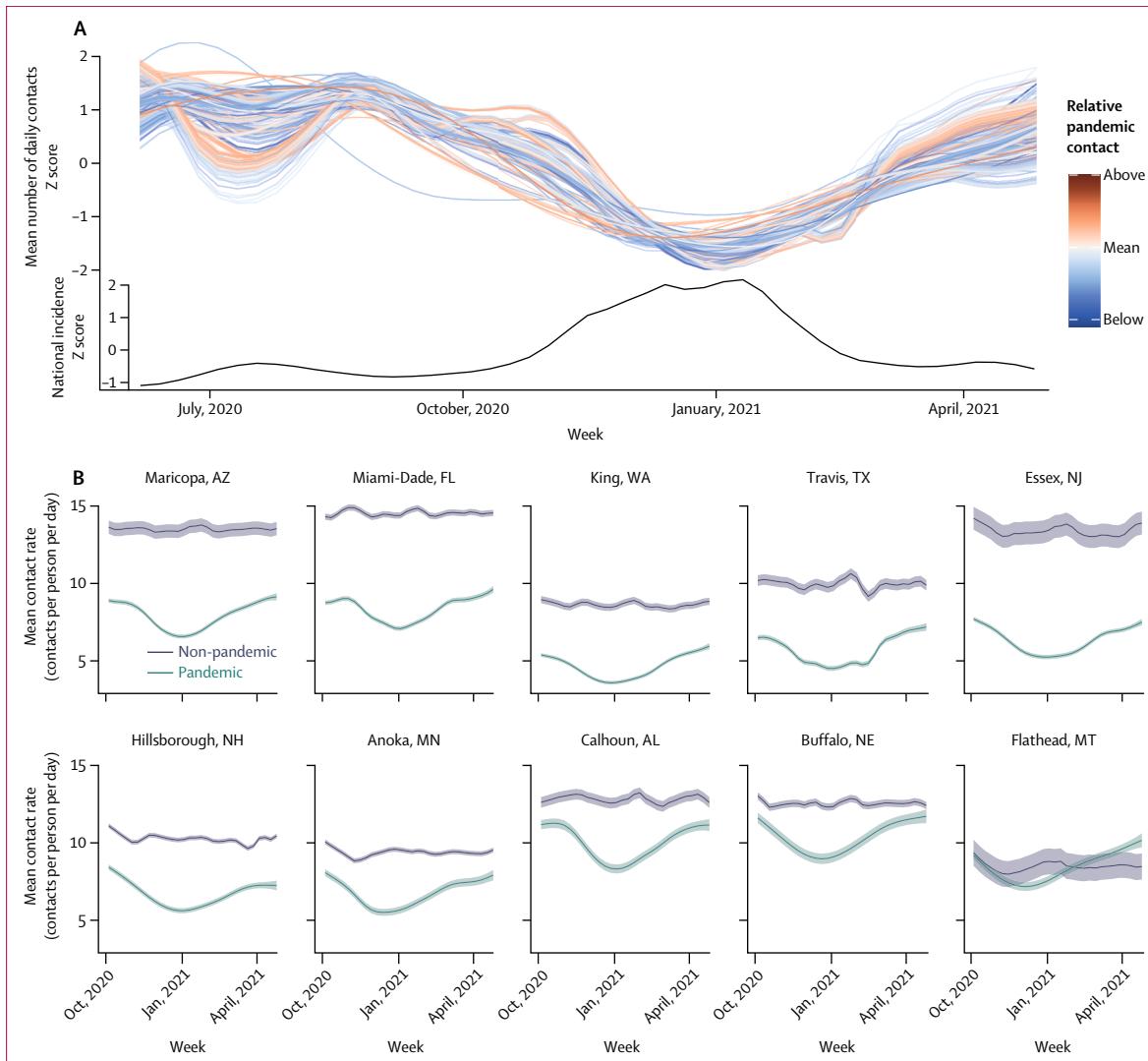


Figure 1: Contact dynamics observed over time during the COVID-19 pandemic and estimated non-pandemic contact dynamics, by county

(A) Mean number of daily non-household contacts for individual counties over time during the COVID-19 pandemic. Contact is presented as a Z score relative to each county's mean to allow comparison between time series despite the large range of mean contact values across counties. Each line represents a county and is coloured by mean contact relative to the national mean. The black line shows the Z score of the centred 3-week rolling average of national COVID-19 case incidence for context. Counties had similar contact dynamics over time: most counties had higher contact during the summer of 2020, and all had lower contact during the winter of 2020–21. Counties in which contact decreased in the summer of 2020 were typically in states that had a higher incidence of COVID-19 during that time. (B) Mean contact rate (non-household contacts) in the absence of disease (baseline; slate) was effectively constant over time, compared with observed contact during the pandemic (teal), across a diverse set of counties. We controlled for disease using a linear regression model that predicts contact from national case incidence, state and county policy data, and county vaccination coverage. This analysis is restricted to Oct 1, 2020, to April 30, 2021, to encompass a full wave of COVID-19. Shaded areas represent 1 SD above and below the fitted contact value or estimated non-pandemic value.

where c represents mean contact; i represents each county, s each state, and n the nation; t represents each week; d represents the 4-week rolling average (mean of the previous 3 weeks and the current week t) of national incident cases; v represents the cumulative percentage of the population that is vaccinated for COVID-19; p_s represents the Oxford Stringency Index centred at the minimum value observed in the regression period for each state; p_i represents the sum of county-level policy metrics centred at the minimum value observed in the regression period for

each county, and ε is normally distributed with a mean of 0 and a variance of σ^2 . Note that counties are not pooled together, which enables us to capture differences in county-level responses to disease and policy metrics. Additional model selection details can be found in the appendix (p 46).

Contact after controlling for disease ($b_{i,t}$) was defined as

$$b_{i,t} = \phi_i(\beta_{0,i} + \varepsilon_{i,t}),$$

where

$$\phi_i = \frac{\omega_{2019,i}}{\omega_{2020,i}}$$

and $\omega_{y,i}$ is defined as the mean number of trips into or within a given county i for October to December of year y . We added the residual from the linear regression model to account for temporal changes in contact not captured by the incidence and policy data. The mobility data were obtained from the SafeGraph Social Distancing dataset.⁵³ We scaled by pre-pandemic mobility data to account for the substantial decrease in contact after the introduction of SARS-CoV-2 into the USA. Although mobility data do not directly measure contact, they are a reasonable proxy for contact as they are highly correlated (appendix p 23).

We fit separate regressions to estimate contact by age, gender, race or ethnicity, and setting at the county or state level for non-pandemic conditions. These models enable us to estimate baseline contact for each demographic or social group by estimating the effect of disease-mediated risk perception and policy.

This analysis was limited to the period from Oct 1, 2020, to April 30, 2021, to encompass a full wave of COVID-19 in the USA. Diagnostics for these regression models are provided in the appendix (pp 29–31).

Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

We characterised heterogeneities in average daily contact patterns in the USA at the county-week scale from June 1, 2020, to April 30, 2021, using 10·7 million valid responses from the CTIS³² (appendix p 19). Contact varied across the pandemic period, although most counties had similar contact dynamics: increased contact during June–August, 2020 and in April, 2021, and decreased contact from September, 2020 to March, 2021 (figure 1A). In some counties—predominantly in Florida, Arizona, Texas, and other parts of the southern USA—contact was also reduced in the summer of 2020, coinciding with a COVID-19 surge in the region. This variation appears to be inversely related to SARS-CoV-2 incidence.

We explored this association using regression models incorporating national case incidence, state and county policies, and county vaccination coverage to predict contact from Oct 1, 2020, to April 30, 2021. After controlling for the effect of disease on behaviour, contact was temporally stable across counties (figure 1B), and remaining fluctuations were neither substantial nor systematic. We validated this regression approach by using disease trends from May 16, 2021, to June 25, 2022 to predict contact patterns, then compared these predictions to actual contact avoidance data from a separate CTIS question and found consistent results

(appendix pp 42–44). Therefore, our baseline contact estimates can be interpreted as conservative predictions for non-pandemic contact beyond the acute phase of the COVID-19 pandemic.

Mean non-household contact was spatially heterogeneous during the pandemic (figure 2A) and in the absence of a pandemic (figure 2B). During the pandemic, the highest degree of contact was observed across the central and southern USA, whereas the lowest contact rates were observed along the north Atlantic and western coasts. By contrast, under baseline conditions, we found a different geographical pattern of contact rates. We investigated these geographical patterns by considering the association between contact and urbanicity. During the COVID-19 pandemic, respondents in the most urban US counties (NCHS class 1) tended to have fewer contacts than those in more rural counties (figure 2C). This difference was eliminated by controlling for disease (figure 2C), suggesting that individuals in urban counties are expected to have slightly more contacts than rural residents under this contact definition in non-pandemic situations. These results are robust to truncation in the reported number of contacts (appendix pp 33–36).

Contact also varied across demographic and social classifications during the pandemic and at baseline. Older respondents tended to have fewer contacts: individuals aged between 18 years and 54 years reported similar numbers of contacts on average across the study period, whereas the contact rate decreased past the age of 55 years (figure 3A, appendix p 40). Men tended to have more contacts than women (figure 3B), whereas Hispanic respondents had the most contacts and Asian respondents had the fewest (figure 3C). The highest number of contacts occurred in work settings, followed by when shopping for essentials (figure 3D). Using separate regression models and county-specific mobility data, we inferred baseline contact rates for each social category. Although responsiveness to disease incidence varies within social classifications (eg, men vs women; appendix p 27), the trends within classifications stay the same from pandemic to baseline contact across all the different social categories (age, gender, and race; figure 3). Our baseline estimates are consistent with those from other studies (appendix pp 3–6). During the pandemic, contact estimates were more variable across studies, potentially owing to differences in contact definition, survey design, or survey period (appendix pp 7–18).

Discussion

Interpersonal contact is required for the spread of directly transmitted pathogens such as SARS-CoV-2. Nevertheless, contact patterns are poorly understood and difficult to predict. Previous contact studies have focused on European nations and measured contact at coarse spatial and temporal scales.^{1,22,28,54} These broad scopes leave open questions about how contact patterns vary subnationally, across seasons, and between demographic and social classifications. Here we estimated non-household contacts at the county-

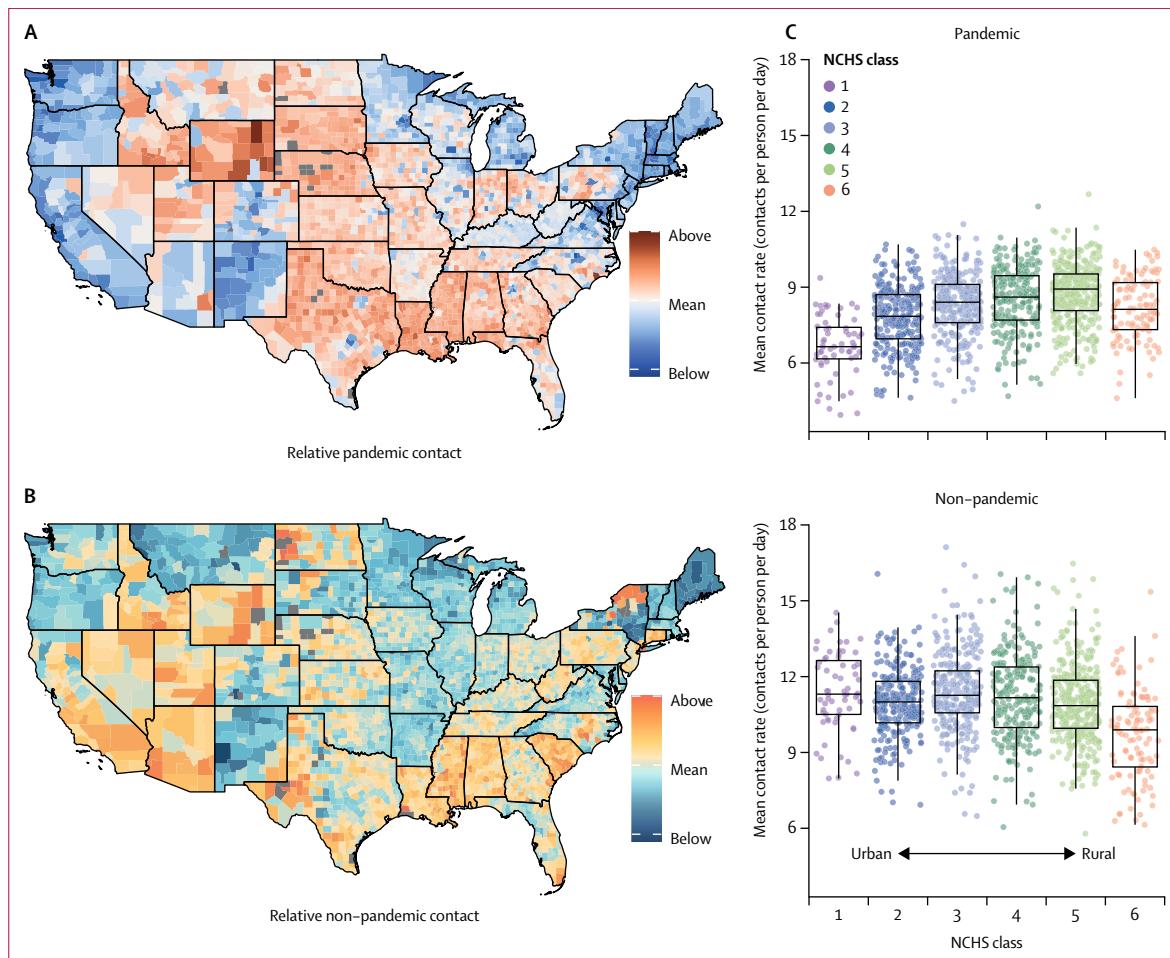


Figure 2: Spatial heterogeneity and urban-rural gradient of pandemic and estimated non-pandemic contact

(A) Mean number of non-household contacts per person per day for each county relative to the national mean (8.7 contacts per person per day) during the COVID-19 pandemic (Oct 1, 2020, to April 30, 2021). There was high spatial heterogeneity in contact, even within states, which was fairly consistent across time (appendix p 28). Counties shaded in grey did not have a sufficient sample size to estimate contact. (B) Map of inferred mean number of non-household contacts per person per day for each county relative to the national mean (10.9 contacts per person per day) in a non-pandemic scenario. Spatial heterogeneity in contact remains high, although which counties have values above and below the national mean has shifted compared with the pattern observed during the COVID-19 pandemic. (C) The mean contact rate (non-household contacts) for each county decreases with increasing urbanicity during the pandemic, but increases with urbanicity during inferred non-pandemic times. Only counties with ten or more responses per week each week (from Oct 1, 2020, to April 30, 2021) are included. NCHS class describes the urbanicity of the county, with 1 indicating a large central metropolitan area and 6 representing rural, non-core areas. NCHS=National Center for Health Statistics.

week scale in the USA using responses from a large national survey during the COVID-19 pandemic (June 1, 2020, to April 30, 2021). We used post-stratification and generalised additive models weighted by sample size to address sample representation and size issues. We compared our findings with those from several smaller previous studies and found consistent patterns. We also used a regression approach to infer non-pandemic contact patterns by controlling for the effect of disease and pandemic interventions. Our findings have several implications for public health researchers and policy makers and could facilitate the much-needed improvement of future disease models and interventions in the USA.

We found that most communities exhibited similar temporal dynamics during the early COVID-19 pandemic,

in which contact rates had an inverse relationship with disease incidence. After controlling for the effect of incidence-mediated risk perception and disease-related policy, we observed little variability in contact patterns over time, suggesting that changes in numbers of contacts cannot explain the seasonality of respiratory infectious diseases. Although contact has been shown to differ between summer and winter in other countries,^{19,55} our results show that contact does not vary meaningfully from autumn to spring, the crucial period during which respiratory diseases emerge and diminish in temperate climates. As such, the role of contact in disease seasonality warrants further study. Our analysis period was limited to 6 months, which might not capture full annual seasonality; longer-term work should investigate whether these trends remain over

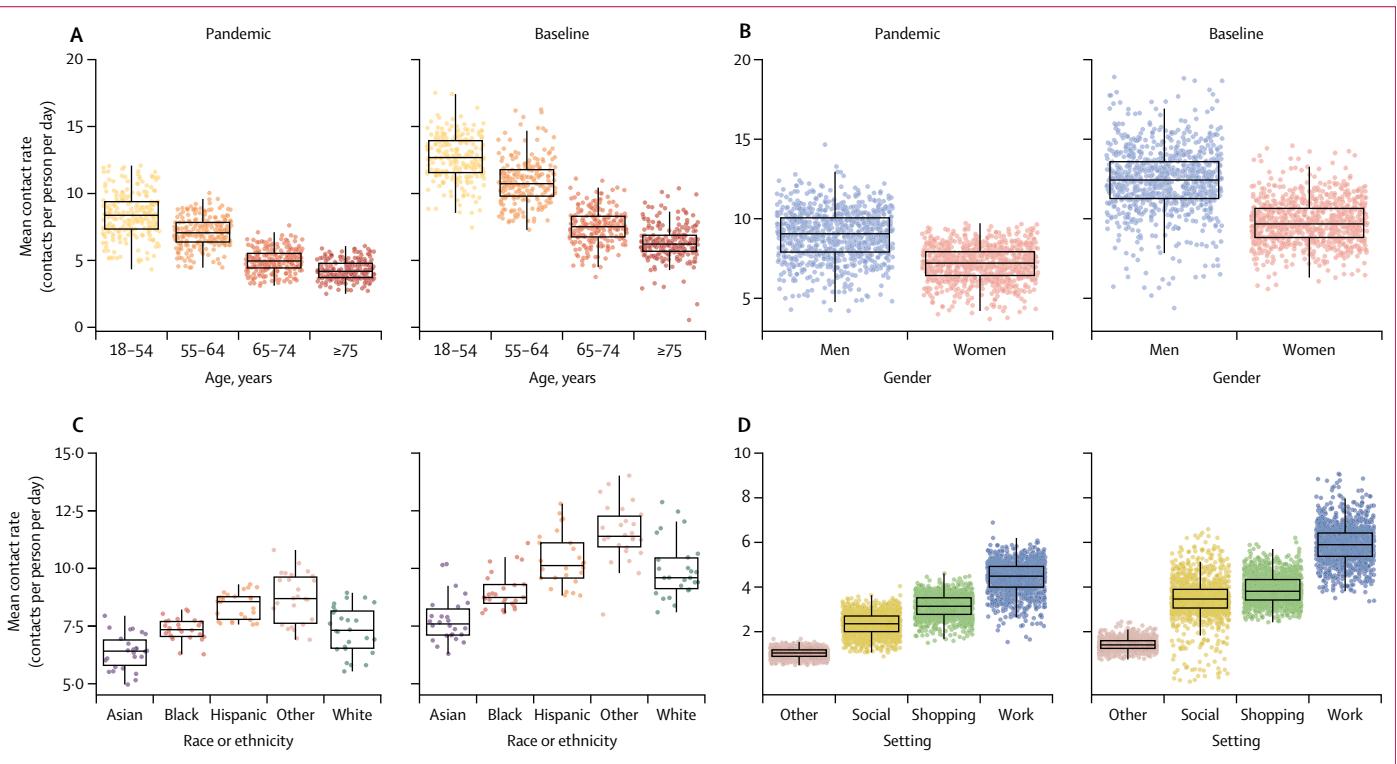


Figure 3: Contact by age, gender, race or ethnicity, and setting during the pandemic and at baseline

(A) Mean pandemic and baseline non-household contact rate by age. Each point represents a county-age category. Analysis was limited to counties with five or more responses per age category per week. (B) Mean pandemic and baseline non-household contact rate by gender. Each point represents a county-gender category. Analysis was limited to counties with five or more responses per gender category per week. (C) Mean pandemic and baseline non-household contact rate by race or ethnicity. Each point represents a state-race or ethnicity category. Analysis was limited to states with ten or more responses per race or ethnicity category per week. All racial and ethnic categories are non-Hispanic unless labelled otherwise. Other denotes individuals who reported their race as American Indian or Alaska Native, Native Hawaiian or Pacific Islander, or other, or as falling in multiple categories. (D) Mean pandemic and baseline non-household contact rate by setting. Each point represents a county-setting. Analysis was limited to counties with ten or more responses per setting per week.

multiple years and be used to validate our inferences. Notably, the survey question we used did not differentiate between indoor and outdoor contact nor whether or not individuals were wearing masks; the setting of human contact is known to affect the likelihood of transmission and has been shown to be seasonal.³⁴ Therefore, with the data presented here, we hypothesise that the setting of contact could be a larger driver of seasonality than the number of contacts. Our finding also reinforces other work showing that fine-grain temporal data might not be necessary for incorporating behaviour into infectious disease models.⁵⁶ However, our hierarchical generalised additive model makes county trajectories within a state more similar by pulling county-level trends towards shared state-level trends, potentially reducing the spatial and temporal variation in our contact estimates.

The high spatial heterogeneity in contact patterns that we observe enables us to identify areas at increased risk of respiratory transmission during seasonal epidemics and pandemics due to high contact rates, such as the southern USA. This result also highlights the importance of high-resolution spatial data: the high variability in contact within states would be obscured if data were aggregated to the state level. We found that people in urban counties had fewer

contacts on average during the pandemic than those in rural counties, which is unsurprising given evidence that urban counties were more responsive to pandemic restrictions.^{26,29,31} Under baseline conditions, however, we found that people in urban areas had more contacts on average than those in suburban or rural areas, which is consistent with behavioural heuristics based on population density. Previous research has not found a consistent relationship between population density and contact rate, both outside the USA^{57–59} and within the USA.^{25,29} This discrepancy might be explained by differences in contact definition; definitions that are tied to density are likely to be more representative of aerosol transmission. Given that geographical variation in disease risk creates challenges for disease surveillance, mitigation, and public health communication, further investigation into the role of spatial heterogeneity in behaviour is warranted.

Understanding which groups have high rates of contact is essential for the development of more precisely targeted interventions and to address public health disparities arising from structural inequities. Although the nature of the CTIS precludes any analysis of contact assortativity or clustering, we can identify demographic and social classifications at greater risk due to higher

contact rates. Indeed, degree (or number of contacts) has been shown to be the most important predictor of disease risk compared with other metrics.³ Like other studies with smaller sample sizes, both pre-pandemic and during the COVID-19 pandemic, we found that older adults had fewer contacts than younger adults.^{1,27,28,60} However, a limitation of our study (and most other contact studies) is a lack of data on children (aged <18 years). We also found that men had more contacts than women during the pandemic, as other national USA surveys have shown.^{27,28} Our non-pandemic model shows that this difference persists under baseline conditions, in contrast to findings from a US study during the 2007–08 influenza season,²⁰ which could be a result of increased contacts by women in the home, and from the POLYMOD study,¹ in which no meaningful difference in contact was found between genders. Additionally, we found that Hispanic individuals had the highest contact rates during the pandemic and Asian respondents had the lowest; these results agree with those from a national survey in 2022,²⁸ but disagree with pre-pandemic time-use data.²⁵ White respondents showed the most responsiveness to changes in disease incidence, probably reflecting an increased ability to work from home. We note, however, that a lack of demographic-specific and social group-specific mobility data could limit our inference of non-pandemic contact patterns across social categories, suggesting potential biases in our baseline contact estimates by social category (appendix p 26). Similarly, a lack of incidence data at the county level disaggregated by demographic group precludes us from analysing the effect of group-specific disease incidence on group-specific behaviour. Census data show that 93% of US counties saw only a 3% or smaller change in population size from April 1, 2020, to July 1, 2021.⁶¹ Therefore, we expect any changes in the demographic or social distribution within counties due to the pandemic to have negligible effects on the baseline contact estimates. There could be additional biases in the data that we have not addressed, such as social desirability bias, over-representation or under-representation by political party, and under-representation of rural areas. Overall, our work highlights that social heterogeneities in contact patterns could be responsible for socially structuring transmission risks for respiratory infections and could shape the landscape of response to disease.

In summary, we have developed some of the most detailed pandemic and baseline estimates of contact patterns in the USA to date, which will be key to informing accurate estimates of transmission risk that account for spatial clustering. Our results can also aid the development of more efficiently targeted interventions. Our work highlights the value of collecting fine-scale behavioural data and the need for long-term longitudinal data collection on contact patterns in the USA. We provide some of the first evidence that US adult contact patterns might not vary over time but do vary across counties, with ramifications for

understanding the seasonality of respiratory infectious diseases. Improving our understanding of contact patterns, which are such an integral component of disease transmission and key to the implementation of targeted interventions in the context of a disease outbreak, should be prioritised in future research efforts.

Contributors

JCT conducted the analyses with contributions from ZS. JCT, VC, and SB interpreted the findings. JCT drafted the manuscript, which was edited by all authors. SB conceptualised and supervised the study. JCT, ZS, and SB accessed and verified the data. VC was not part of the data agreement with Carnegie Mellon University and so did not have access to the individual survey responses. All authors had final responsibility for the decision to submit the manuscript for publication.

Declaration of interests

We declare no competing interests.

Data sharing

Contact estimates, both during the pandemic and at baseline, at the county-week scale and all code to produce and analyse these data are available on GitHub at https://github.com/bansallab/resp_contact. Individual survey responses cannot be shared by the authors, but researchers can refer to <https://cmu-delphi.github.io/delphi-epidata/symptom-survey/data-access.html> if they would like to enter an agreement for data usage with the Carnegie Mellon University Delphi Group.

Acknowledgments

We thank the Carnegie Mellon University Delphi Group, especially Alex Reinhart for sharing the US COVID-19 Trends and Impact Survey results openly and freely and Casey Zipfel for collating pandemic policy data. This research was supported by the National Institutes of Health under award numbers R01GM123007 and R35GM153478 (to SB) and the ANR grant DATAREDUX ANR-19-CE46-0008-03 and EU Horizon Europe grant ESCAPE 101095619 (to VC).

References

- 1 Mossong J, Hens N, Jit M, et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med* 2008; **5**: e74.
- 2 Hoang T, Coletti P, Melegaro A, et al. A systematic review of social contact surveys to inform transmission models of close-contact infections. *Epidemiology* 2019; **30**: 723–36.
- 3 Bansal S, Grenfell BT, Meyers LA. When individual behaviour matters: homogeneous and network models in epidemiology. *J R Soc Interface* 2007; **4**: 879–91.
- 4 Gostic KM, McGough L, Baskerville EB, et al. Practical considerations for measuring the effective reproductive number, R_t . *PLOS Comput Biol* 2020; **16**: e1008409.
- 5 Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. Superspreaders and the effect of individual variation on disease emergence. *Nature* 2005; **438**: 355–59.
- 6 Rohani P, Zhong X, King AA. Contact network structure explains the changing epidemiology of pertussis. *Science* 2010; **330**: 982–85.
- 7 Mistry D, Litvinova M, Pastore y Piontti A, et al. Inferring high-resolution human mixing patterns for disease modeling. *Nat Commun* 2021; **12**: 323.
- 8 Kretzschmar ME, Rozhnova G, Bootsma MCJ, van Boven M, van de Wijgerd JHHM, Bonten MJM. Impact of delays on effectiveness of contact tracing strategies for COVID-19: a modelling study. *Lancet Public Health* 2020; **5**: e452–59.
- 9 Aleta A, Martín-Corral D, Pastore y Piontti A, et al. Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19. *Nat Hum Behav* 2020; **4**: 964–71.
- 10 Overton CE, Stage HB, Ahmad S, et al. Using statistics and mathematical modelling to understand infectious disease outbreaks: COVID-19 as an example. *Infect Dis Model* 2020; **5**: 409–41.

- 11 Gimma A, Munday JD, Wong KLM, et al. Changes in social contacts in England during the COVID-19 pandemic between March 2020 and March 2021 as measured by the CoMix survey: a repeated cross-sectional study. *PLoS Med* 2022; **19**: e1003907.
- 12 Coletti P, Wambua J, Gimma A, et al. CoMix: comparing mixing patterns in the Belgian population during and after lockdown. *Sci Rep* 2020; **10**: 21885.
- 13 Munday JD, Abbott S, Meakin S, Funk S. Evaluating the use of social contact data to produce age-specific short-term forecasts of SARS-CoV-2 incidence in England. *PLoS Comput Biol* 2023; **19**: e1011453.
- 14 Viboud C, Bjørnstad ON, Smith DL, Simonsen L, Miller MA, Grenfell BT. Synchrony, waves, and spatial hierarchies in the spread of influenza. *Science* 2006; **312**: 447–51.
- 15 Favier C, Schmit D, Müller-Graf CDM, et al. Influence of spatial heterogeneity on an emerging infectious disease: the case of dengue epidemics. *Proc Biol Sci* 2005; **272**: 1171–77.
- 16 Susswein Z, Valdano E, Brett T, Rohani P, Colizza V, Bansal S. Ignoring spatial heterogeneity in drivers of SARS-CoV-2 transmission in the US will impede sustained elimination. *medRxiv* 2021; published online Aug 10. <https://doi.org/10.1101/2021.08.09.21261807> (preprint).
- 17 Truelove SA, Graham M, Moss WJ, Metcalf CJE, Ferrari MJ, Lessler J. Characterizing the impact of spatial clustering of susceptibility for measles elimination. *Vaccine* 2019; **37**: 732–41.
- 18 Bharti N, Tatem AJ, Ferrari MJ, Grais RF, Djibo A, Grenfell BT. Explaining seasonal fluctuations of measles in Niger using nighttime lights imagery. *Science* 2011; **334**: 1424–27.
- 19 Kummer AG, Zhang J, Jiang C, et al. Evaluating seasonal variations in human contact patterns and their impact on the transmission of respiratory infectious diseases. *Influenza Other Respir Viruses* 2024; **18**: e13301.
- 20 DeStefano F, Haber M, Curriyan D, et al. Factors associated with social contacts in four communities during the 2007–2008 influenza season. *Epidemiol Infect* 2011; **139**: 1181–90.
- 21 Zagheni E, Billari FC, Manfredi P, Melegaro A, Mossong J, Edmunds WJ. Using time-use data to parameterize models for the spread of close-contact infectious diseases. *Am J Epidemiol* 2008; **168**: 1082–90.
- 22 Breen CF, Mahmud AS, Feehan DM. Novel estimates reveal subnational heterogeneities in disease-relevant contact patterns in the United States. *PLoS Comput Biol* 2022; **18**: e1010742.
- 23 Taube JC, Susswein Z, Bansal S. Spatiotemporal trends in self-reported mask-wearing behavior in the United States: analysis of a large cross-sectional survey. *JMIR Public Health Surveill* 2023; **9**: e42128.
- 24 Tiu A, Susswein Z, Merritt A, Bansal S. Characterizing the spatiotemporal heterogeneity of the COVID-19 vaccination landscape. *Am J Epidemiol* 2022; **191**: 1792–802.
- 25 Dorélien AM, Ramen A, Swanson I, Hill R. Analyzing the demographic, spatial, and temporal factors influencing social contact patterns in U.S. and implications for infectious disease spread. *BMC Infect Dis* 2021; **21**: 1009.
- 26 Dorélien AM, Venkateswaran N, Deng J, et al. Quantifying social contact patterns in Minnesota during stay-at-home social distancing order. *BMC Infect Dis* 2023; **23**: 324.
- 27 Feehan DM, Mahmud AS. Quantifying population contact patterns in the United States during the COVID-19 pandemic. *Nat Commun* 2021; **12**: 893.
- 28 Nelson KN, Siegler AJ, Sullivan PS, et al. Nationally representative social contact patterns among U.S. adults, August 2020–April 2021. *Epidemics* 2022; **40**: 100605.
- 29 Klein B, LaRock T, McCabe S, et al. Characterizing collective physical distancing in the U.S. during the first nine months of the COVID-19 pandemic. *PLoS Digit Health* 2024; **3**: e0000430.
- 30 Garnier R, Benetka JR, Kraemer J, Bansal S. Socioeconomic disparities in social distancing during the COVID-19 pandemic in the United States: observational study. *J Med Internet Res* 2021; **23**: e24591.
- 31 Zang E, West J, Kim N, Pao C. U.S. regional differences in physical distancing: evaluating racial and socioeconomic divides during the COVID-19 pandemic. *PLoS One* 2021; **16**: e0259665.
- 32 Salomon JA, Reinhart A, Bilinski A, et al. The US COVID-19 Trends and Impact Survey: continuous real-time measurement of COVID-19 symptoms, risks, protective behaviors, testing, and vaccination. *Proc Natl Acad Sci USA* 2021; **118**: e2111454118.
- 33 Read JM, Eames KT, Edmunds WJ. Dynamic social networks and the implications for the spread of infectious disease. *J R Soc Interface* 2008; **5**: 1001–07.
- 34 Susswein Z, Rest EC, Bansal S. Disentangling the rhythms of human activity in the built environment for airborne transmission risk: an analysis of large-scale mobility data. *eLife* 2023; **12**: e80466.
- 35 Crawford FW, Jones SA, Carter M, et al. Impact of close interpersonal contact on COVID-19 incidence: evidence from 1 year of mobile device data. *Sci Adv* 2022; **8**: eabi5499.
- 36 Tomori DV, Rübsamen N, Berger T, et al. Individual social contact data and population mobility data as early markers of SARS-CoV-2 transmission dynamics during the first wave in Germany—an analysis based on the COVIMOD study. *BMC Med* 2021; **19**: 271.
- 37 Delusu F, Tizzoni M, Gauvin L. The limits of human mobility traces to predict the spread of COVID-19: a transfer entropy approach. *PNAS Nexus* 2023; **2**: pgad302.
- 38 Ingram DD, Franco SJ. 2013 NCHS Urban–Rural Classification Scheme for Counties. Vital and Health Statistics Series 2, Number 166. April, 2014. <https://www.cdc.gov/nchs/data-analysis-tools/urban-rural.html> (accessed Sept 7, 2021).
- 39 Wood SN. Generalized additive models: an introduction with R, 2nd edn. Chapman & Hall/CRC, 2017.
- 40 Simpson GL. Modelling palaeoecological time series using generalized additive models. *Front Ecol Evol* 2018; **6**: 149.
- 41 Mellor J, Christie R, Overton CE, et al. Forecasting influenza hospital admissions within English sub-regions using hierarchical generalized additive models. *Commun Med (Lond)* 2023; **3**: 190.
- 42 Pedersen EJ, Miller DL, Simpson GL, Ross N. Hierarchical generalized additive models in ecology: an introduction with mgcv. *PeerJ* 2019; **7**: e6876.
- 43 Wood SN, Zheyuan L, Gavlin S, Augustin NH. Generalized additive models for gigadata: modeling the U.K. Black Smoke Network daily data. *J Am Stat Assoc* 2017; **112**: 1199–210.
- 44 The New York Times. Coronavirus (Covid-19) data in the United States. 2021. <https://github.com/nytimes/covid-19-data> (accessed March 8, 2023).
- 45 Hale T, Angrist N, Goldszmidt R, et al. A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). *Nat Hum Behav* 2021; **5**: 529–38.
- 46 Zipfel C. The interplay between human behavior and infectious disease dynamics. PhD thesis, Georgetown University, 2021.
- 47 Katz R, Too K, Robertson H, et al. Open data for COVID-19 policy analysis and mapping. *Sci Data* 2023; **10**: 491.
- 48 Centers for Disease Control and Prevention. U.S. state and territorial gathering bans: March 11, 2020–August 15, 2021 by county by day. https://data.cdc.gov/Policy-Surveillance/U-S-State-and-Territorial-Gathering-Bans-March-11-/7xvh-y5vh/about_data (accessed Aug 29, 2021).
- 49 Centers for Disease Control and Prevention. U.S. state and territorial orders closing and reopening bars issued from March 11, 2020 through August 15, 2021 by county by day. https://data.cdc.gov/Policy-Surveillance/U-S-State-and-Territorial-Orders-Closing-and-Reope/9kjw-3miq/about_data (accessed Aug 29, 2021).
- 50 Centers for Disease Control and Prevention. U.S. state and territorial orders closing and reopening restaurants issued from March 11, 2020 through May 31, 2021 by county by day. https://data.cdc.gov/Policy-Surveillance/U-S-State-and-Territorial-Orders-Closing-and-Reope/647a-wjd2/about_data (accessed Aug 29, 2021).
- 51 Centers for Disease Control and Prevention. U.S. state and territorial public mask mandates from April 10, 2020 through August 15, 2021 by county by day. https://data.cdc.gov/Policy-Surveillance/U-S-State-and-Territorial-Public-Mask-Mandates-Fro/62d6-pm5i/about_data (accessed Aug 29, 2021).
- 52 Centers for Disease Control and Prevention. U.S. state and territorial stay-at-home orders: March 15, 2020 – August 15, 2021 by county by day. https://data.cdc.gov/Policy-Surveillance/U-S-State-and-Territorial-Stay-At-Home-Orders-Marc/y2iy-8irm/about_data (accessed Aug 29, 2021).

-
- 53 SafeGraph. Social distancing metrics. <https://docs.safegraph.com/docs/social-distancing-metrics> (accessed Dec 19, 2023).
- 54 Jarvis CI, Van Zandvoort K, Gimma A, et al. Quantifying the impact of physical distance measures on the transmission of COVID-19 in the UK. *BMC Med* 2020; **18**: 124.
- 55 Nagpal S, Kumar R, Noronha RF, et al. Seasonal variations in social contact patterns in a rural population in north India: implications for pandemic control. *PLoS One* 2024; **19**: e0296483.
- 56 Pullano G, Alvarez-Zuzek LG, Colizza V, Bansal S. Characterizing US spatial connectivity and implications for geographical disease dynamics and metapopulation modeling: longitudinal observational study. *JMIR Public Health Surveill* 2025; **11**: e64914.
- 57 Read JM, Lessler J, Riley S, et al. Social mixing patterns in rural and urban areas of southern China. *Proc Biol Sci* 2014; **281**: 20140268.
- 58 Kiti MC, Kinyanjui TM, Koech DC, Munywoki PK, Medley GF, Nokes DJ. Quantifying age-related rates of social contact using diaries in a rural coastal population of Kenya. *PLoS One* 2014; **9**: e104786.
- 59 Kleynhans J, Tempia S, McMorrow ML, et al. A cross-sectional study measuring contact patterns using diaries in an urban and a rural community in South Africa, 2018. *BMC Public Health* 2021; **21**: 1055.
- 60 Mousa A, Winskill P, Watson OJ, et al. Social contact patterns and implications for infectious disease transmission—a systematic review and meta-analysis of contact surveys. *eLife* 2021; **10**: e70294.
- 61 United States Census Bureau. County population totals and components of change: 2020–2024. <https://www.census.gov/data/tables/time-series/demo/popest/2020s-counties-total.html> (accessed Dec 6, 2024).