Respiratory disease contact patterns in the US are stable but heterogeneous

Juliana C. Taube[1,], Zachary Susswein[1], Vittoria Colizza[2], and Shweta Bansal[1,*]

[1]Department of Biology, Georgetown University, Washington, DC, USA
[2]INSERM, Paris, France
[*]Corresponding Author, shweta.bansal@georgetown.edu

April 2024

Abstract

Background: Contact plays a critical role in infectious disease transmission. Characterizing heterogeneity in contact patterns across individuals, time, and space is necessary to inform accurate estimates of transmission risk, particularly to explain superspreading, predict age differences in vulnerability, and inform social distancing policies. Current respiratory disease models often rely on data from the 2008 POLYMOD study conducted in Europe, which is now outdated and potentially unrepresentative of behavior in the US. We seek to understand the variation in contact patterns across spatial scales and demographic and social classifications, whether there is seasonality to contact patterns, and what social behavior looks like at baseline in the absence of an ongoing pandemic.

Methods: We analyze spatiotemporal non-household contact patterns across 11 million survey responses from June 2020 - April 2021 post-stratified on age and gender to correct for sample representation. To characterize spatiotemporal heterogeneity in respiratory contact patterns at the county-week scale, we use generalized additive models. In the absence of pre-pandemic data on contact in the US, we also use a regression approach to produce baseline contact estimates to fill this gap.

Findings: Although contact patterns varied over time during the pandemic, contact is relatively stable after controlling for disease. We find that the mean number of non-household contacts is spatially heterogeneous regardless of disease. There is additional heterogeneity across age, gender, race/ethnicity, and contact setting, with mean contact decreasing with age and lower in women. The contacts of white individuals and contacts at work or social events change the most under increased national incidence.

Interpretation: We develop the first county-level estimates of non-pandemic contact rates for the US that can fill critical gaps in parameterizing disease models. Our results identify that spatiotemporal, demographic, and social heterogeneity in contact patterns is highly structured, informing the risk landscape of respiratory disease transmission in the US.

Funding: Research reported in this publication was supported by the National Institutes of Health under award number R01GM123007 (SB).

Research in Context

Evidence before this study: We searched Google Scholar for contact data in the US both during and prior to the pandemic published by February 1, 2024 with 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 these articles and included known literature not found via these search criteria. We excluded studies using mobility data, focusing on children, or excluding the US. Previous work has been limited to the state scale or subsets of counties (e.g., focused on a few cities, a single state, or a few counties within a state) rather than all counties in the US.

Added value of this study: We contribute the first high-resolution pandemic contact estimates for the US and infer non-pandemic contact patterns at fine spatial and temporal scales. Our results indicate that the number of contacts is fairly stable over time in the absence of major disease, suggesting that the number of contacts is not driving respiratory disease seasonality in the US. We also identify groups at greatest disease risk due to higher contacts, including younger adults, men, and Hispanic and Black individuals.

Implications of all the available evidence: This study demonstrates the importance of incorporating age-specific and spatial heterogeneity of contact patterns into future disease models to build accurate estimates of transmission risk. We demonstrate that temporal variability in contact patterns is unlikely to drive respiratory disease seasonality, that adaptive behaviors in response to disease shift risk along an urban-rural gradient, and that some vulnerable groups are at increased risk of exposure due to contact. We advocate that geographic and social heterogeneity in exposure to disease due to contact patterns be captured more comprehensively for accurate infectious disease predictions and effective and equitable disease mitigation.

Introduction

Respiratory disease transmission via direct or droplet routes requires close contact. Research over the last two decades has demonstrated that human contact patterns are highly variable across individuals and locations [1] and highlighted the consequences of this variability for epidemic outcomes and dynamics [2]. Yet, to date, there is little detailed empirical data on contact patterns in the United States. This lack of data leaves several important gaps in our understanding of the drivers of disease transmission. For example, we'd like to know what factors (e.g., age, season, or location) influence contact patterns across the US and how these factors contribute to variability in infection risk. This information is essential to design targeted interventions and generate accurate estimates of transmission risk across individuals, space, and time. With a new dataset composed of over 11 million survey responses from June 2020 through April 2021, we developed the first detailed characterization of heterogeneities in human contact patterns across the US.

Most infectious disease models assume homogeneous mixing amongst individuals. That is, all individuals have the same contact rate and ability to transmit disease. Homogeneous mixing models have transformed the prediction and control of disease outbreaks (e.g., through R_t estimation [3]) but produce different epidemic dynamics and outcomes than models that incorporate heterogeneities in contact patterns [2, 4]. The POLYMOD study published in 2008 was the first extensive survey to characterize heterogeneities in routine contact patterns relevant to respiratory disease transmission. The data, from eight European countries, suggested that individual contact rates are not homogeneous but rather heavy-tailed and highly assortative by age [1]. These estimates have been used to understand epidemic dynamics, design vaccine strategies, and predict intervention outcomes (e.g., [5–9]). During the COVID-19 pandemic, researchers conducted the CoMix study to collect contact data across Europe [10, 11]; estimates from the survey were incorporated into forecasting efforts in the UK with mixed results [12]. Notably, both surveys fail to capture fine-scale spatial and temporal variability in contact, are becoming outdated, and may not be representative of behavior in the United States.

Beyond individual heterogeneity, there are several meaningful dimensions across which contact patterns may vary. Spatial heterogeneity in human behavior plays a vital role in disease dynamics (e.g., [13–16]); if contact patterns exhibit spatial heterogeneity, that might explain observed hotspots of disease burden and/or dynamics of disease spread. Likewise, changes in contact patterns over time may contribute to respiratory disease seasonality (e.g., [17, 18]). Differences in contact across demographic and social classifications, such as age, gender, or race/ethnicity, could also drive disparities in attack rates. Contact may also vary based on surrounding disease transmission as individuals shift their behavior to mitigate risk. These potential heterogeneities in contact patterns profoundly affect our understanding and prediction of epidemic dynamics and our ability to target behavioral interventions. Yet little empirical data exists on them. To identify the dimensions across which contact meaningfully varies, we need high-resolution contact data across geography, time, demography, social classifications, and in the context of disease transmission.

There has been recent attention on contact patterns in the US; several studies have been produced in the last three years (with the exception of earlier studies [19, 20]) which shed light on variation in contact patterns across multiple dimensions. Breen et al. [21] demonstrate that contact varies between states but are unable to characterize spatial heterogeneity at finer scales in the US, even though other public health-related behaviors have been demonstrated to vary at the county-level [22, 23]. Dorelien et al. use time-use surveys to demonstrate that contact between urban and rural areas may not differ [24] despite the perception that urban inhabitants have more contacts. On temporal variation in contact, one pre-pandemic study found no variation in adult contacts over time [24], while another observed changes from September to May [19]. Many studies find that contact is higher in younger adults, men, and non-white populations, although which race/ethnicity group has the highest contact depends on the study [19, 25–27]. Because these past studies are limited in sample size or resolution, they are constrained in their ability to comprehensively characterize heterogeneities in contact patterns across space, time, and social groups.

Now that the COVID-19 public health emergency has ended, it is vital to characterize contact patterns during the pandemic but also under non-pandemic baseline conditions. In general, pandemic social distancing

reduced overall contact (e.g., [28]). However, past work has demonstrated that adherence to pandemic social distancing was heterogeneous across populations, driven in part by health disparities and social inequities that affect the ability of individuals to engage in behavioral interventions [29]. For example, urban areas exhibited more significant reductions in mobility and, therefore, likely contact [30]. Individuals of higher socioeconomic status are known to have had greater flexibility in their mitigation behavior and could further reduce contacts. Understanding which groups are at most significant infection risk during pandemics and seasonal epidemics is critical for targeted public health surveillance and resource allocation. Detailed contact data disaggregated by location, age, gender, and race/ethnicity are necessary to clarify who is at greatest risk during pandemics and seasonal epidemics in the US.

Here, we address these pressing gaps by developing fine-scale spatiotemporal estimates of mean non-household contacts in adults. We use an extensive national survey with over 11 million responses collected at the county level and correct for issues of representativeness and small sample size via post-stratification and generalized additive models. We characterize several heterogeneities in contact patterns throughout the pandemic and infer non-pandemic contact patterns by controlling for the effect of disease. We focus on four central questions: (1) Is contact seasonal? (2) How does mean contact vary across geography in urban versus rural settings? (3) How do contact patterns vary across age, gender, and race/ethnicity classifications? (4) What are contact patterns like at baseline? Our results are the most comprehensive high-resolution estimates of US contact patterns to date. Our findings can inform future disease models in the US, provide insight into local and temporal variation in behavior in response to public health messaging, and contribute to our understanding of drivers of respiratory disease seasonality.

Results

To characterize heterogeneities in contact patterns in the United States we used data from the COVID-19 Trends and Impact Survey (CTIS) from June 2020 through April 2021 [31]. Respondents self-reported the number of direct non-household contacts they had in the last 24 hours (11.2 million valid responses). We used post-stratification on age and gender and generalized additive models (GAMs) weighted by sample size to address issues of representation and noise, respectively. From the GAMs, we estimated mean non-household contact at the county-week level to examine spatial and temporal heterogeneities in contact behavior. We also inferred non-pandemic contact patterns by controlling for the impact of disease on contact behavior. Lastly, we analyzed contact pattern variability across demographic and social classifications to identify individuals at highest infection risk based on their contact patterns.

Contact is aseasonal under baseline conditions

Contact varies across the pandemic period, although most counties exhibit similar contact dynamics: increased contact during summer 2020 and spring 2021 and decreased contact during winter 2020-21 (Figure 1A). Some counties, predominantly in Florida, Arizona, Texas, and other parts of the southern US, also reduced their contact in the summer of 2020, coinciding with a COVID-19 surge in the region. This variation appears to be inversely related to national incidence.

We explore this association using regression models with county and national disease incidence and state policies to predict contact from October 2020 to April 2021. After controlling for the effect of disease on behavior, contact is seasonally stable across counties (Figure 1B). Remaining fluctuations are not substantial nor systematic.

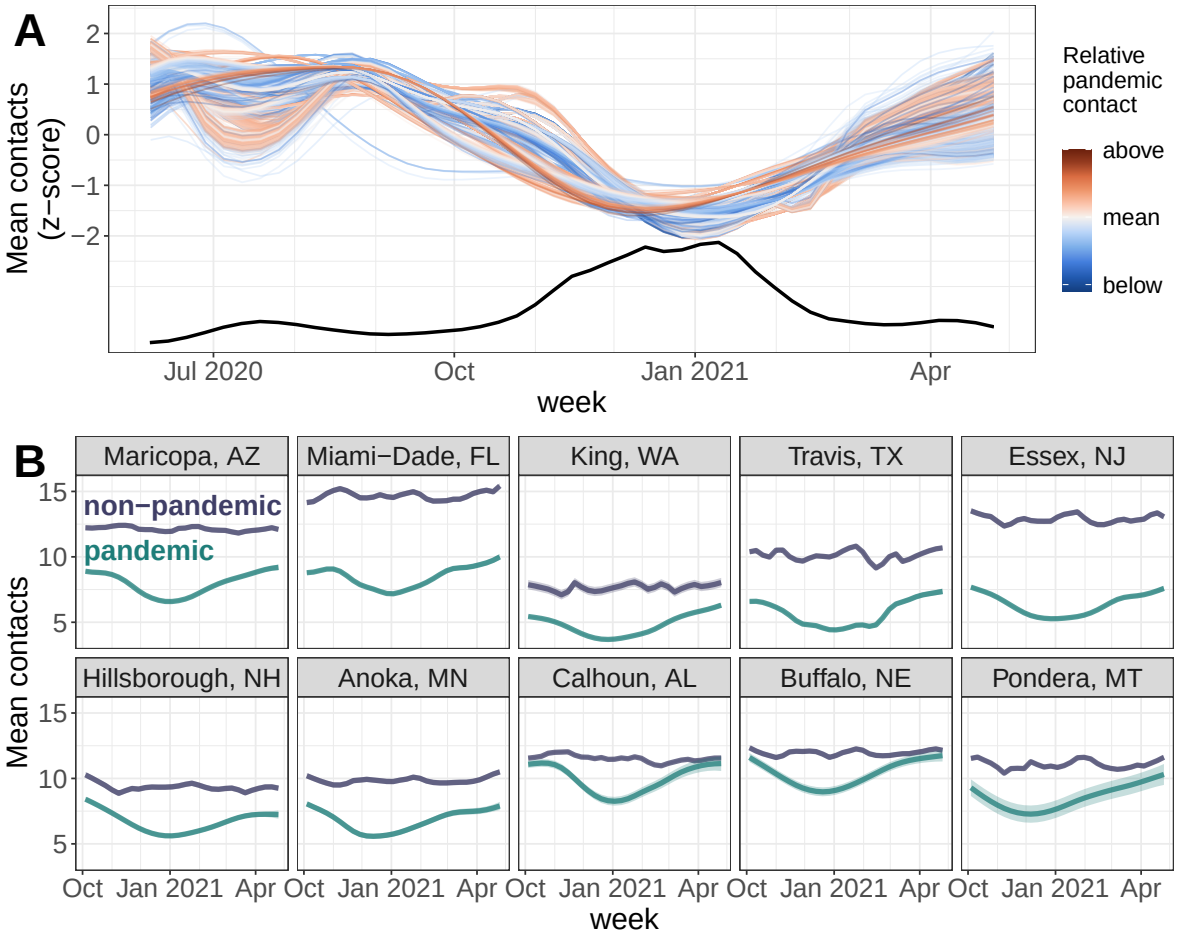

Figure 1. Counties have similar contact dynamics over time and relatively stable contact after controlling for disease. (A) Most counties had higher contact during the summer of 2020, and all had lower contact during the winter of 2020-21. Counties that experienced a dip in contact in the summer of 2020 were typically in states that exhibited higher incidence during that time. Each line represents a county colored by mean contact relative to the national mean (above or below); zscored contact relative to each county's mean is shown to allow comparison between time series despite the large range of mean contact values across counties. Black line shows the centered 3-week rolling average of national incidence for context. (B) Contact in the absence of disease (blue) is effectively constant over time compared to observed contact during the pandemic (teal). We controlled for disease using a linear regression predicting contact from county and national disease incidence and state policy data. This analysis is restricted to October 2020 to April 2021 to encompass a full wave of COVID-19. Shaded areas represent one standard error above and below the fitted contact value or estimated non-pandemic value.

Contact is geographically heterogeneous but structured by urbanicity

Mean non-household contact is spatially heterogeneous during the pandemic (Figure 2A) and in the absence of a pandemic (Figure 2B). During the pandemic, the highest degree of contact is observed across the central and southern US while the lowest contact rates are observed along the north Atlantic and western coasts. In contrast, under baseline conditions, we estimate a change in the geographic pattern in contact rates.

We investigate these geographic patterns by considering the association with urbanicity. During the COVID-19 pandemic, respondents in the most urban US counties (NCHS class 1) tended to have fewer contacts compared to more rural counties (Figure 2C). This difference is 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.

We demonstrate that these results are robust to truncation in the reported number of contacts (Figures S29, S30).

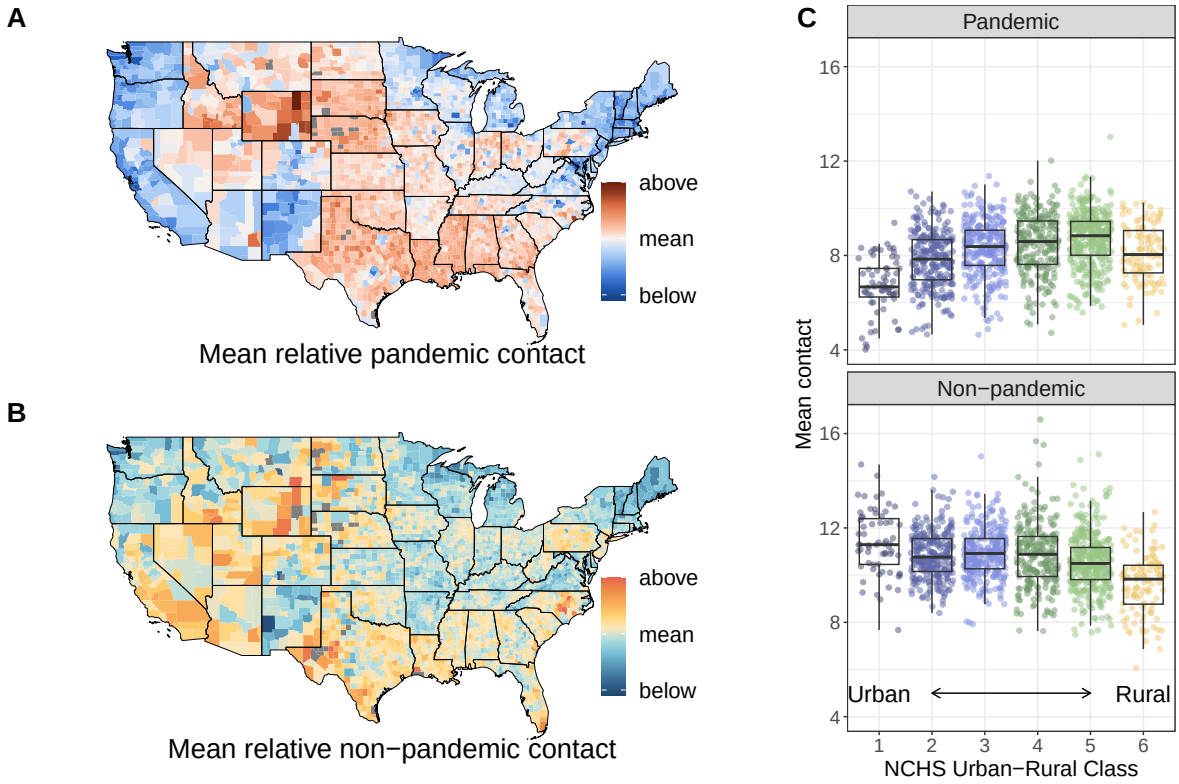

Figure 2. Contact is spatially heterogeneous regardless of disease incidence, but the urban-rural gradient reverses after controlling for disease incidence. (A) Map of mean number of non-household contacts per county relative to the national mean (8.7 contacts) during the pandemic (October 2020 - April 2021). There is high spatial heterogeneity in contact, even within states, which is fairly consistent across time (Figure S25). Gray counties did not have sufficient sample size to estimate contact. (B) Map of inferred mean number of non-household contacts per county relative to the national mean (10.6 contacts) in a non-pandemic scenario. Spatial heterogeneity in contact remains high, though which counties are above and below the national mean has shifted from the pattern observed during the pandemic. (C) Mean number of contacts for each county decreases with increasing urbanicity during the pandemic but increases with urbanicity during inferred non-pandemic times. Only counties with 10 or more responses per week each week (Oct 2020 - Apr 2021) are included.

Contact varies demographically and socially

Contact also varies across demographic and social classifications during the pandemic and at baseline. Older respondents tend to have fewer contacts, with individuals between 18 and 54 reporting about the same numbers of contacts on average across the study period (Figures 3A, S31). Men tend to have more contacts than women (Figure 3B), whereas Hispanic respondents have the most contacts compared to Asian respondents who have the fewest (Figure 3C). Most adult contacts occur in work settings, followed by shopping for essentials (Figure 3D). Using separate regression models and county-specific mobility data, we infer baseline contact for each social category. While responsiveness to disease incidence varies within social classifications

(Figure S24), the relationship between baseline contact and each social classification remains the same as during the pandemic (Figure 3). Our baseline results are consistent with those from other studies (Figures S1 to S4). During the pandemic, contact estimates are more variable across studies, potentially due to differences in contact definition, survey design, or survey period (Figures S5 to S18).

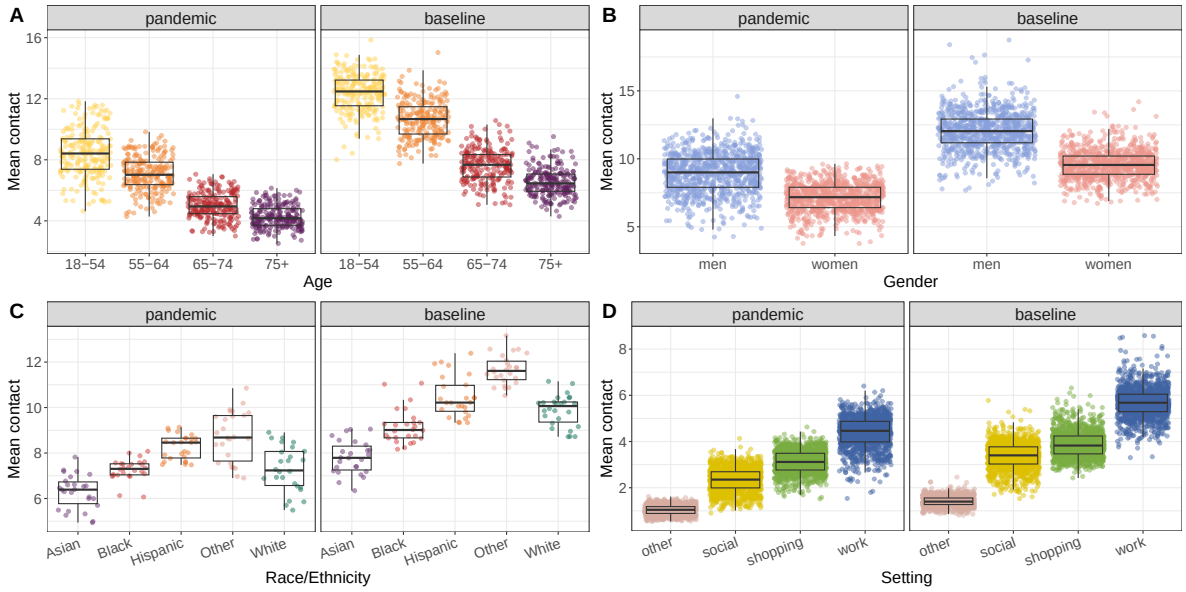

Figure 3. Contact varies across age, gender, race/ethnicity, and setting during the pandemic and at baseline. (A) Mean pandemic and baseline contact by age. Each point represents a county-age category. Analysis is limited to counties with 5+ responses per age category per week. (B) Mean pandemic and baseline contact by gender. Each point represents a county-gender category. Analysis is limited to counties with 5+ responses per gender category per week. (C) Mean pandemic and baseline contact by race/ethnicity. Each point represents a state-race/ethnicity category. Analysis is limited to states with 10+ responses per race/ethnicity category per week. All racial/ethnic categories are non-Hispanic unless labeled otherwise. Other denotes individuals who listed their race as American Indian, Alaska Native, Native Hawaiian, Pacific Islander, or "some other race". (D) Mean pandemic and baseline contact by setting. Each point represents a county-setting. Analysis is limited to counties with 10+ responses per setting per week.

Discussion

Interpersonal contact is required for the spread of directly-transmitted pathogens such as SARS-CoV-2. Nevertheless, contact patterns remain poorly understood and difficult to predict. Previous contact studies have focused on European nations and measured contact at coarse spatial and temporal scales (e.g., [1,21,27, 32]). These broad scopes leave open questions about how contact patterns vary subnationally, across seasons, and between demographic and social classifications. Here, we estimate non-household contacts at the county-week scale in the US using responses from a large national survey during the COVID-19 pandemic (June 2020 - April 2021). We used post-stratification and generalized additive models weighted by sample size to address sample representation and size issues. Despite these efforts, it is possible that we may not address all biases in the data (e.g., social desirability bias). Thus, we compare our findings to a number of smaller past studies and show consistent patterns. We also use a regression approach to infer pre-pandemic contact patterns by controlling for the effect of disease. Our findings have several implications for public health researchers and policymakers and can facilitate the much-needed improvement of future disease models and interventions in the US.

We find that most counties exhibited similar temporal dynamics during the early COVID-19 pandemic and

demonstrate an inverse relationship with disease trends. Indeed, after controlling for the effect of disease, we observe little variability in contact patterns over time suggesting that contact may be less of a contributor to respiratory disease seasonality than originally thought. As such, the role of contact in disease seasonality warrants further study; longer-term work should investigate whether these trends remain over multiple years, including when primary and secondary schools are back in session. Importantly, the survey question we use did not differentiate between indoor and outdoor contact or whether individuals were wearing masks; the setting of human contact certainly affects the likelihood of transmission and has been shown to be seasonal [33]. Thus, with the data presented here, we hypothesize that the setting of contact may be a larger driver of seasonality than the number of contacts. Our finding also reinforces other work that fine-grain temporal data may not be necessary for incorporating behavior into infectious disease models [34].

The high spatial heterogeneity in contact patterns that we observe allows us to identify areas at increased pandemic risk of transmission due to high contact rates, such as the central and southern US. This result also highlights the importance of high-resolution spatial data: there is high variability in contact within states that would be obscured if data are aggregated to the state level. We find that urban counties had fewer contacts on average during the pandemic; this is unsurprising given evidence that urban counties were more responsive to pandemic restrictions [25, 28, 30]. Under baseline conditions, however, we find that urban areas have more contacts on average than suburban or rural areas, which is consistent with behavioral heuristics based on population density. Earlier research has not found a consistent relationship between population density and contact rate, both outside the US [35–37] and in the US [24, 28]. This discrepancy may be explained by differences in contact definition; definitions that are tied to density are likely more representative of aerosol transmission. Given that geographic variation in disease risk creates challenges for disease surveillance, mitigation, and public health communication, further investigation on the role of spatial heterogeneity in behavior is warranted.

Understanding which groups have higher rates of contact is essential for the development of more targeted interventions and to address public health disparities stemming from structural inequities. While 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 has been shown to be the most important predictor of disease risk compared to other metrics [2]. Like other studies of smaller sample size, both pre-pandemic and pandemic, we find that older adults have fewer contacts than younger adults [1, 26, 27, 38]. However, a limitation of our study (and most other contact studies) is a lack of data on children (individuals aged < 18). We also find that men have more contacts than women during the pandemic, as other national US surveys have shown [26, 27]. Our non-pandemic model shows that this difference persists in baseline conditions, in contrast with [19] (which may be a result of increased contacts by women in the home) and [1] (which finds no meaningful difference in contact 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 a national survey in 2022 [27] but disagree with pre-pandemic time use data [24]. White respondents showed the most responsiveness to changes in disease incidence, likely reflecting increased ability to work from home. We note, however, that a lack of demographic/social group-specific mobility data may limit our inference of non-pandemic contact patterns across social categories, suggesting potential biases in our baseline contact estimates by social category (Figure S23). Overall, our work highlights that social heterogeneities in contact patterns may be responsible for socially structuring transmission risks for respiratory infections and may shape the landscape of response to disease.

In summary, we have developed some of the most detailed estimates of contact patterns in the US 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 behavioral data and the need for long-term longitudinal data collection on contact patterns in the US. We provide some of the first evidence that US adult contact patterns may not be seasonal but do vary across counties, with ramifications for understanding respiratory disease seasonality. Improving our understanding of contact patterns, which are such an integral component of disease transmission, should be further prioritized in research efforts going forward.

Methods

In this study, we seek to characterize heterogeneities in non-household contact patterns in the United States using survey data from June 2020 to April 2021. We use survey raking and reweighting of responses to correct for unrepresentative survey sampling. Due to the low sample size in some counties, we use generalized additive models (GAMs) to develop smoothed county-week estimates of the mean number of non-household contacts, disaggregated by geographic, demographic, and social categories. To interrogate whether there is seasonality in US contact patterns in the absence of a pandemic, we account for the effect of disease incidence on contact with a simple linear regression model and examine the remaining unexplained variability in contact.

Survey data

We analyzed survey responses of the number of non-household contacts for all 50 US states and the District of Columbia using data from the US COVID-19 Trends and Impact Survey (CTIS) [31]. The CTIS was created by the Delphi Research Group at Carnegie Mellon University and distributed through a partnership with Facebook. Beginning in April 2020, a random state-stratified sample of active Facebook users were invited daily to take the survey about COVID-19 and report how many people they had direct contact with outside their household, where contact was defined as "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." Contacts were disaggregated by settings outside the home in the survey question: work, shopping for groceries or other essentials, social gatherings, or other. We analyze the sum of contacts across all these settings starting in June 2020. We removed responses with more than 72 contacts in the last 24 hours (the 95th percentile) and performed sensitivity analyses of this truncation point (Figures S27 to S30). Finally, we generated response weights to match county age and sex distributions and post-stratified the data to adjust for unrepresentative sampling (details in supplement). To test the sensitivity of our estimates to including race/ethnicity in the raking weights, we compared mean contact estimates at the state-month level for September 2020 - April 2021 (Figure S26).

Spatiotemporal contact estimation

To address noise and low sample sizes in the data, we estimated county-week mean non-household contacts with hierarchical generalized additive models (GAMs). We use penalized 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. Thus, information is shared within states but not between. The model is as follows:

$$\text{mean contact}_i = \beta_0 + f(t) + f_i(t) + \epsilon_i$$

where

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

where i represents each county and t each week [39]. The coefficients $\beta_{j,k}$ are estimated for each of the K basis functions $b_{j,k}$.

County urbanicity was determined using the NCHS 2013 urban-rural classifications, where 1 indicates large central metropolitan areas, and 6 represents rural non-core areas [40].

We fit GAMs to estimate contact by urbanicity, age, gender, race/ethnicity, and setting at the county or state level for pandemic and baseline conditions (more below). Additional details can be found in the supplement.

Baseline contact estimation

To estimate contact in the absence of a pandemic, we used a linear regression model of weekly contact (as modeled using the GAMs above) predicted by the 4-week rolling average of county and national incidence [41] and the average state-level Oxford Stringency Index [42] (Figure S22). We hypothesized that increases in disease incidence would lead to a linear decrease in contact rates, which is supported by the survey data (Figures S20, S21). The model is as follows:

$$\text{mean contact}_{i,t} = \beta_{0_i} + \beta_{1_i} * \text{incidence}_{\text{county } i_t} + \beta_{2_i} * \text{incidence}_{\text{national}_t} + \beta_{3_i} * \text{policy index}_{\text{state}_t} + \epsilon_{i,t}$$

where i represents each county, t represents each week, and incidence represents the 4-week rolling average (mean of the previous four weeks) at the appropriate scale.

Contact after controlling for disease was defined as

$$\text{baseline contact}_{i,t} = \phi_i \ (\beta_{0_i} + \epsilon_{i,t})$$

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 thru December of year y. The mobility data is obtained from the SafeGraph Social Distancing dataset [43]. We added the residual to account for temporal changes in contact not captured by the incidence and policy data. We scale by pre-pandemic mobility data to account for the drastic decrease in contact upon SARS-CoV-2 introduction in the United States; mobility and contact patterns are highly correlated (Figure S19).

This analysis was limited to the period from October 2020 to April 2021 to encompass a full "wave" of COVID-19 in the US.

Contributors

JCT did the analyses, interpreted the findings, and drafted and edited the manuscript. ZS contributed to the analyses and edited the manuscript. VC interpreted the findings and edited the manuscript. SB conceived and supervised the study, interpreted the findings, and edited the manuscript. JCT, ZS, and SB accessed and verified the data. All authors were responsible 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 analyze these data will be made 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 CMU Delphi.

Acknowledgments

The authors thank the Carnegie Mellon University Delphi team for sharing the US COVID-19 Trends and Impact Survey results openly and freely.

References

1. 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 Mar;5(3):e74.

2. Bansal S, Grenfell BT, Meyers LA. When Individual Behaviour Matters: Homogeneous and Network Models in Epidemiology. Journal of The Royal Society Interface. 2007;4:879–891.

3. Gostic KM, McGough L, Baskerville EB, Abbott S, Joshi K, Tedijanto C, et al. Practical Considerations for Measuring the Effective Reproductive Number, Rt. PLOS Computational Biology. 2020 Dec;16(12):e1008409.

4. Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. Superspreading and the Effect of Individual Variation on Disease Emergence. Nature. 2005 Nov;438(7066):355–359.

5. Rohani P, Zhong X, King AA. Contact Network Structure Explains the Changing Epidemiology of Pertussis. Science. 2010 Nov;330(6006):982–985.

6. 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.

7. Kretzschmar ME, Rozhnova G, Bootsma MCJ, van Boven M, van de Wijgert JHHM, Bonten MJM. Impact of Delays on Effectiveness of Contact Tracing Strategies for COVID-19: A Modelling Study. The Lancet Public Health. 2020 Aug;5(8):e452–e459.

8. 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):964–971.

9. Overton CE, Stage HB, Ahmad S, Curran-Sebastian J, Dark P, Das R, et al. Using Statistics and Mathematical Modelling to Understand Infectious Disease Outbreaks: COVID-19 as an Example. Infectious Disease Modelling. 2020 Jan;5:409–441.

10. Gimma A, Munday JD, Wong KLM, Coletti P, van Zandvoort K, Prem K, 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 Medicine. 2022 Mar;19(3):e1003907.

11. Coletti P, Wambua J, Gimma A, Willem L, Vercruysse S, Vanhoutte B, et al. CoMix: Comparing Mixing Patterns in the Belgian Population during and after Lockdown. Scientific Reports. 2020 Dec;10(1):21885.

12. 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 Computational Biology. 2023 Sep;19(9):e1011453.

13. Viboud C, Bjørnstad ON, Smith DL, Simonsen L, Miller MA, Grenfell BT. Synchrony, Waves, and Spatial Hierarchies in the Spread of Influenza. Science. 2006 Apr;312(5772):447–451.

14. Favier C, Schmit D, Müller-Graf CDM, Cazelles B, Degallier N, Mondet B, et al. Influence of Spatial Heterogeneity on an Emerging Infectious Disease: The Case of Dengue Epidemics. Proceedings of the Royal Society B: Biological Sciences. 2005 Jun;272(1568):1171–1177.

15. 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 Aug;.

16. 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 Jan;37(5):732–741.

17. 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 Dec;334(6061):1424–1427.

18. Kummer AG, Zhang J, Jiang C, Litvinova M, Ventura PC, Garcia MA, et al. Evaluating Seasonal Variations in Human Contact Patterns and Their Impact on the Transmission of Respiratory Infectious Diseases. Infectious Diseases (except HIV/AIDS); 2022.

19. DeStefano F, Haber M, Currivan D, Farris T, Burrus B, Stone-Wiggins B, et al. Factors Associated with Social Contacts in Four Communities during the 2007–2008 Influenza Season. Epidemiology & Infection. 2011 Aug;139(8):1181–1190.

20. 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. American Journal of Epidemiology. 2008 Nov;168(9):1082–1090.

21. Breen CF, Mahmud AS, Feehan DM. Novel Estimates Reveal Subnational Heterogeneities in Disease-Relevant Contact Patterns in the United States. PLOS Computational Biology. 2022 Dec;18(12):e1010742.

22. 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 and Surveillance. 2023 Mar;9(1):e42128.

23. Tiu A, Susswein Z, Merritt A, Bansal S. Characterizing the Spatiotemporal Heterogeneity of the COVID-19 Vaccination Landscape. American Journal of Epidemiology. 2022 Sep;191(10):1792–1802.

24. 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 Infectious Diseases. 2021 Sep;21(1):1009.

25. 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.

26. Feehan DM, Mahmud AS. Quantifying Population Contact Patterns in the United States during the COVID-19 Pandemic. Nature Communications. 2021 Feb;12(1):893.

27. Nelson KN, Siegler AJ, Sullivan PS, Bradley H, Hall E, Luisi N, et al. Nationally Representative Social Contact Patterns among U.S. Adults, August 2020-April 2021. Epidemics. 2022 Sep;40:100605.

28. 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.

29. Garnier R, Benetka JR, Kraemer J, Bansal S. Socioeconomic Disparities in Social Distancing During the COVID-19 Pandemic in the United States: Observational Study. Journal of Medical Internet Research. 2021 Jan;23(1):e24591.

30. 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 Nov;16(11):e0259665.

31. Salomon JA, Reinhart A, Bilinski A, Chua EJ, La Motte-Kerr W, Rönn MM, et al. The US COVID-19 Trends and Impact Survey: Continuous Real-Time Measurement of COVID-19 Symptoms, Risks, Protective Behaviors, Testing, and Vaccination. Proceedings of the National Academy of Sciences. 2021 Dec;118(51):e2111454118.

32. CMMID COVID-19 working group, Jarvis CI, Van Zandvoort K, Gimma A, Prem K, Klepac P, et al. Quantifying the Impact of Physical Distance Measures on the Transmission of COVID-19 in the UK. BMC Medicine. 2020 Dec;18(1):124.

33. 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 Apr;12:e80466.

34. Pullano G, Alvarez-Zuzek LG, Colizza V, Bansal S. Characterizing US Spatial Connectivity: Implications for Geographical Disease Dynamics and Metapopulation Modeling. medRxiv; 2024.

35. Read JM, Lessler J, Riley S, Wang S, Tan LJ, Kwok KO, et al. Social Mixing Patterns in Rural and Urban Areas of Southern China. Proceedings of the Royal Society B: Biological Sciences. 2014 Jun;281(1785):20140268.

36. 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.

37. Kleynhans J, Tempia S, McMorrow ML, von Gottberg A, Martinson NA, Kahn K, 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 Jun;21(1):1055.

38. Mousa A, Winskill P, Watson OJ, Ratmann O, Monod M, Ajelli M, et al. Social Contact Patterns and Implications for Infectious Disease Transmission – a Systematic Review and Meta-Analysis of Contact Surveys. eLife. 2021 Nov;10:e70294.

39. Pedersen EJ, Miller DL, Simpson GL, Ross N. Hierarchical Generalized Additive Models in Ecology: An Introduction with Mgcv. PeerJ. 2019 May;7:e6876.

40. Ingram DD, Franco SJ. 2013 NCHS Urban-Rural Classification Scheme for Counties. Vital and Health Statistics Series 2, Data Evaluation and Methods Research. 2014 Apr;(166):1–73.

41. The New York Times. Coronavirus (Covid-19) Data in the United States; 2021. https://github.com/nytimes/covid-19-data.

42. Hale T, Angrist N, Goldszmidt R, Kira B, Petherick A, Phillips T, et al. A Global Panel Database of Pandemic Policies (Oxford COVID-19 Government Response Tracker). Nature Human Behaviour. 2021 Apr;5(4):529–538.

43. Safegraph. Social Distancing Metrics;. https://docs.safegraph.com/docs/social-distancing-metrics.

44. Meta. User Guide for the COVID-19 Trends and Impact Survey Weights; 2022. Version 1.

45. Pasek J. Anesrake: ANES Raking Implementation; 2018. https://CRAN.R-project.org/package=anesrake.

46. Wood SN. Generalized Additive Models: An Introduction with R. Second edition ed. Chapman & Hall/CRC Texts in Statistical Science. Boca Raton: CRC Press/Taylor & Francis Group; 2017.

47. Feehan DM, Cobb C. Using an Online Sample to Estimate the Size of an Offline Population. Demography. 2019 Dec;56(6):2377–2392.

48. Kiti MC, Aguolu OG, Zelaya A, Chen HY, Ahmed N, Batross J, et al. Changing Social Contact Patterns among US Workers during the COVID-19 Pandemic: April 2020 to December 2021. Epidemics. 2023 Dec;45:100727.

49. Google. COVID-19 Community Mobility Report; 2021. https://www.google.com/covid19/mobility?hl=en.

Supplementary Materials

Additional CTIS details

The COVID-19 Trends and Impact Survey was created by researchers in the Delphi Group at Carnegie Mellon University and distributed via Facebook to active users 18 or older starting in April 2020. On a daily basis, a random state-stratified sample of Facebook users are invited to take the survey at the top of their news feed. These users will not be re-invited to take the survey for at least thirty days. The survey asks a broad range of questions related to COVID-19 symptoms and behaviors, with variations across each of the thirteen different waves spanning from April 2020 to June 2022. In this study, we make use of data from Wave 3 (May 21, 2020 - September 7, 2020), Wave 4 (September 8, 2020 - November 23, 2020), Wave 5 (November 24, 2020 - December 18, 2020), Wave 6 (December 19, 2020 - January 11, 2021), Wave 7 (January 12, 2021 - February 7, 2021), Wave 8 (February 8, 2021 - March 1, 2021), and Wave 10 (March 2, 2021 - May 19, 2021, Wave 9 was skipped for numbering purposes).

Weights were provided by Facebook for each response to adjust for non-response and coverage bias at the daily-state level [44]. Briefly, each weight describes the number of people represented by a respondent based on their age, gender, location, and date of response. These weights were calculated via a two-step process using inverse propensity score weighting based on respondents' demographics recorded in their Facebook user profiles to adjust the survey sample to reflect active Facebook users, followed by revisions of these weights using post-stratification so that the survey sample reflects the general population [31]. Because these weights were not at the same scale of our data analysis, we did not use them and instead performed raking to calculate our own weights as described below.

To address the lack of representation in this social media-administered survey, we calculated county-week weights for each observation using the `anesrake` package [45] and the US Census American Community Survey's 2021 data on county age and sex distributions. Age and sex distributions were based on each county's population over 18 to match the survey sample. We use three categories for age (18-24, 25-54, and 55+) and two categories for sex (ACS)/gender (CTIS) (male/man and female/woman). Subsetting age further results in issues of nonconvergence. Including race or ethnicity was not possible in our raking approach as race/ethnicity data were only available for half the survey period and because convergence was not feasible with it. We did not rake on household size because our age and gender raked sample sufficiently captured county household size distributions (Figures S36 and S37). We estimated mean contact for each county-week using a weighted mean of the responses with the calculated weights. We excluded observations with missing age or sex from the raking process and assigned equal weights to observations from county-weeks that did not converge.

As part of our data processing, we dropped responses missing fips codes, outside the 50 states and the District of Columbia, and from county-weeks with fewer than three responses.

Spatiotemporal contact estimation details

Generalized additive models were fit with a Gaussian family and identity link using the `bam` function with the `fREML` estimator from the `mgcv` version 1.8-39 R package [46]. To enforce smoother fits for all 50 states, the basis dimensions were set at maximum 30 for the weekly smooth and maximum 15 for the county-level factor smooth with marginal penalty 2 for both smooths, and a gamma value of 2 for each model. For DC, the weekly smooth had a maximum basis dimension of 10 with a penalized cubic regression spline. Diagnostic checks via `gam.check()` were completed to ensure that most p-values were not significant, k-indexes were close to 1, residuals were roughly normally distributed, and there was no heteroscedascity.

Baseline contact estimation details

The VIF values for national and county incidence were high, but our estimates of are contact robust to the removal of county incidence from the model which reduces all VIFs below 10 (Figure S32).

Contact by demographic or social classification model details

Age & Gender

Using spatiotemporal GAMs, we estimated mean contact in four age groups for each county-week (18-54, 55-64, 65-74, 75+). We had to collapse age groups to have sufficient sample size, but note that contact rates among the 18-54 age group were quite similar prior to aggregation (Figure S31). This analysis was limited to counties with 5 or more responses in each age group for each of the 30 weeks from October 2020 through April 2021 (224 counties). For each age group separately, we estimated pandemic contact using the spatiotemporal GAMs and baseline contact using linear regression as described above. The same steps were followed for the two gender classes in the survey; there were 889 counties with sufficient sample size to estimate contact by gender. Given these sample size constraints, these results may not be representative of differences in contact patterns between age (and possibly gender) groups in the most rural US counties. This setup does allow us to compare age- and gender-specific regression coefficients describing the association between contact and disease incidence for each age/gender class. The VIF values for national and county incidence were high, as in the main model, but our estimates of contact are quantitatively the same if county incidence is removed to reduce all VIFs below 10 and our coefficient results are qualitatively consistent (Figures S33 to S35).

Race/ethnicity

Using spatiotemporal GAMs, we estimated mean contact in five race/ethnicity categories (Asian, Black, Hispanic, other, and white) for each state-week. This analysis was limited to states with 10 or more responses in each racial/ethnic category for each of the 30 weeks from October 2020 through April 2021 (24 states). Due to limited sample size, this analysis had to be performed at the state instead of the county level. For each racial/ethnic category separately, we estimated pandemic contact using the spatiotemporal GAMs and baseline contact using linear regression as described above with a couple exceptions. In the original GAMs we allowed for information sharing between counties within the same state; for these race/ethnicity-specific GAMs we instead allow for information sharing between states in the same census region of the US (West, Midwest, Northeast, South). For the linear regression, we predict contact using only national incidence to avoid collinearity issues that arise when using state-level incidence as well.

Setting

Using spatiotemporal GAMs, we estimated mean contact in four settings (work, social gatherings, shopping for essentials, and other) for each county-week. This analysis was limited to counties with 10 or more responses in each setting for each of the 30 weeks from October 2020 through April 2021. For each setting separately, we estimated pandemic contact using the spatiotemporal GAMs and baseline contact using linear regression as described above.

Comparison to existing contact studies

We validated our findings with other work both before and during the pandemic, in Figures S1 to S18 below.

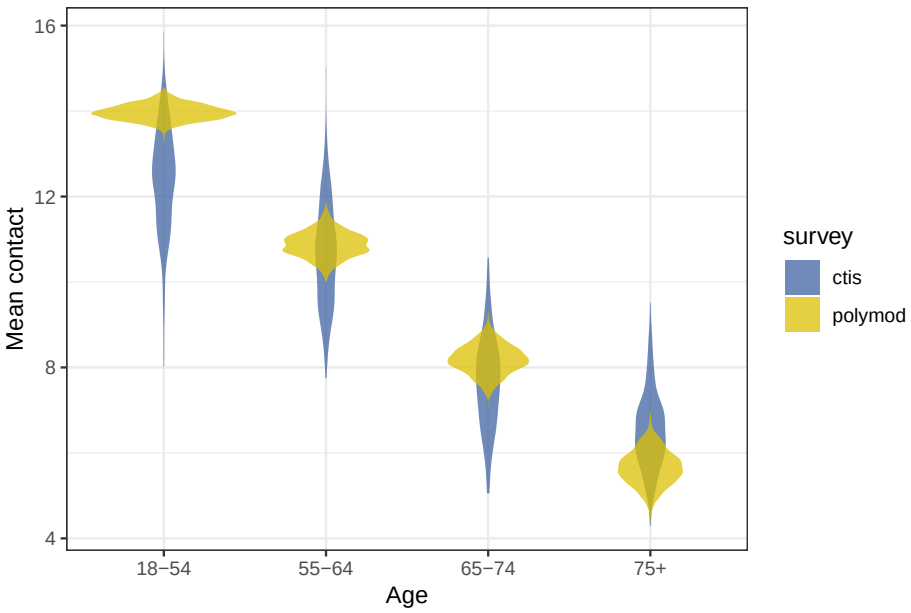

Figure S1. Comparison between our baseline contact estimates by age with those from the POLYMOD study [1]. Yellow violins are composed of bootstrapped means from 1,000 samples from the POLYMOD study (all responses equally weighted, truncated at 72 contacts, over all countries). Blue violins show the distribution of mean county-age group baseline estimates inferred from the CTIS.

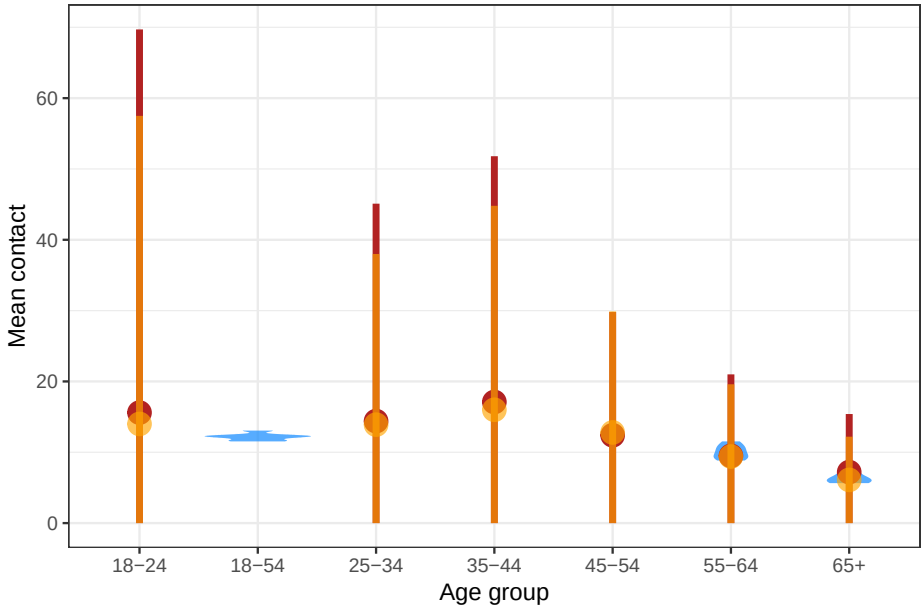

Figure S2. Comparison between our baseline contact estimates by age with those from [19]. Destefano and colleagues' [19] analyzed contact patterns across four counties in North Carolina, USA during the 2007-08 flu season, with 3,845 survey respondents. They measured both the mean number of speaking interactions (orange), defined as face to face conversations lasting at least one minute, and mean number of close proximity contacts (red), defined as within 6 feet for at least fifteen minutes. Point ranges represent one standard deviation above and below the mean (truncated at 0). Blue violins show the distribution of mean county-age group baseline estimates for seven North Carolina counties (those with sufficient group sample size) inferred from the CTIS.

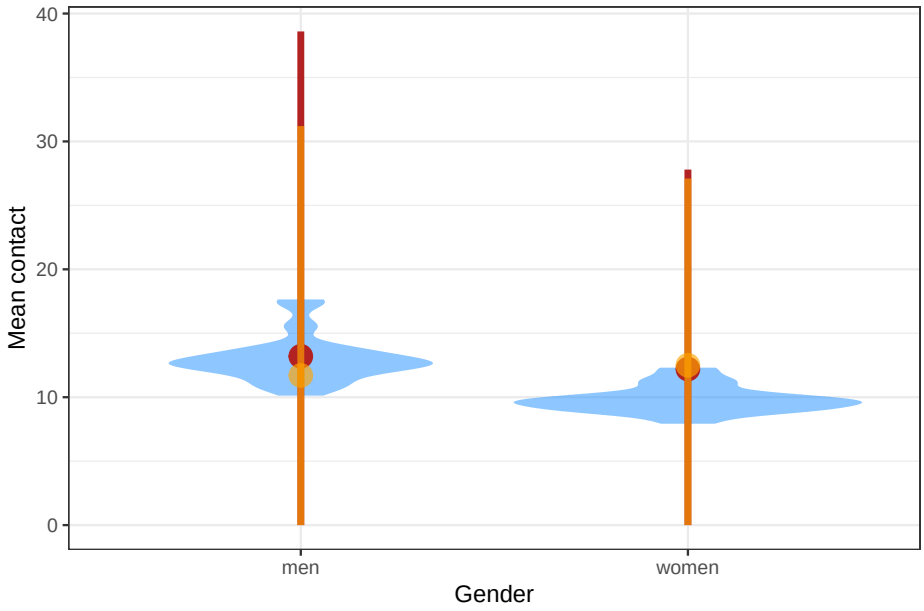

Figure S3. Comparison between our baseline contact estimates by gender with those from [19]. Destefano and colleagues' [19] analyzed contact patterns across four counties in North Carolina, USA during the 2007-08 flu season, with 3,845 survey respondents. They measured both the mean number of speaking interactions (orange), defined as face to face conversations lasting at least one minute, and mean number of close proximity contacts (red), defined as within 6 feet for at least fifteen minutes. Point ranges represent one standard deviation above and below the mean (truncated at 0). Blue violins show distribution of mean county-gender baseline estimates for 36 North Carolina counties (those with sufficient group sample size) inferred from the CTIS.

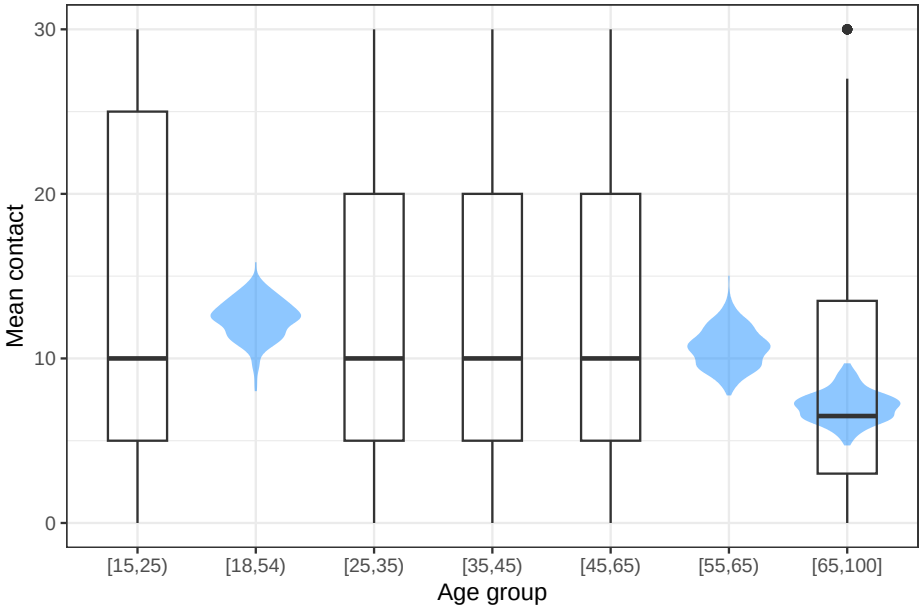

Figure S4. Comparison between our baseline contact estimates by age with those from [47]. Feehan and Cobb [47] conducted a national study of 3,076 Facebook users 15 years of age and older in the US before the pandemic (2015). They defined a contact as a face to face conversation containing at least three words. Boxplots were constructed using the raw data downloaded from `https://doi.org/10.7910/DVN/M74AJ4` `fb_ego.tab` file. Blue violins show distribution of mean county-age group baseline estimates inferred from the CTIS.

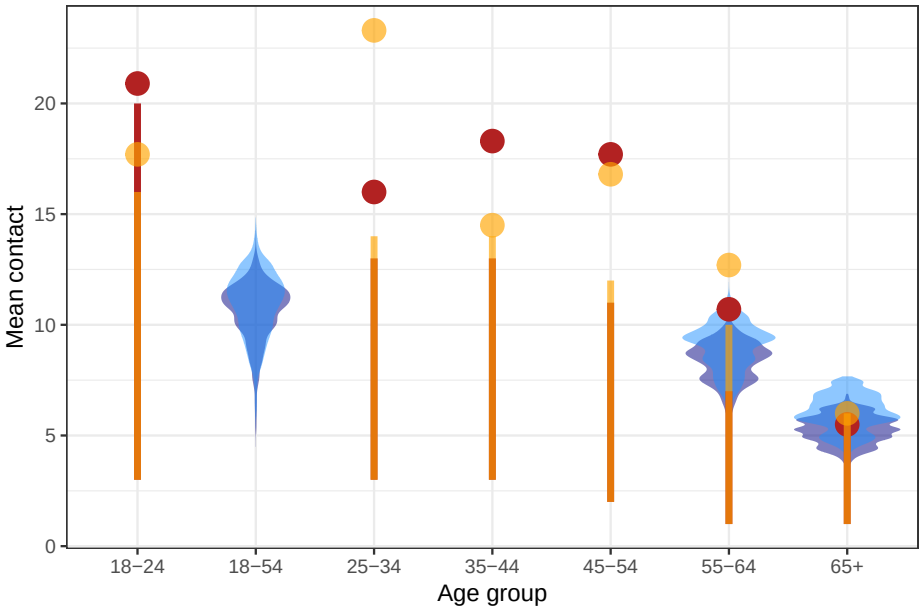

Figure S5. Comparison between our pandemic contact estimates by age with those from [27]. Nelson and colleagues' [27] conducted a nationally representative study of 3,112 individuals in the US during two periods, August - December 2020 (red) and March - April 2021 (orange). Contacts were defined as interactions within 6 feet involving the exchange of at least 3 words or involving physical touch. Point ranges show the mean, 25th, and 75th percentiles; note that [27] truncated contacts at 50 per day. Blue violins show distribution of mean county-age group estimates from the CTIS: dark blue for September - December 2020 and light blue for March - April 2021 to match the time periods in Nelson et al.

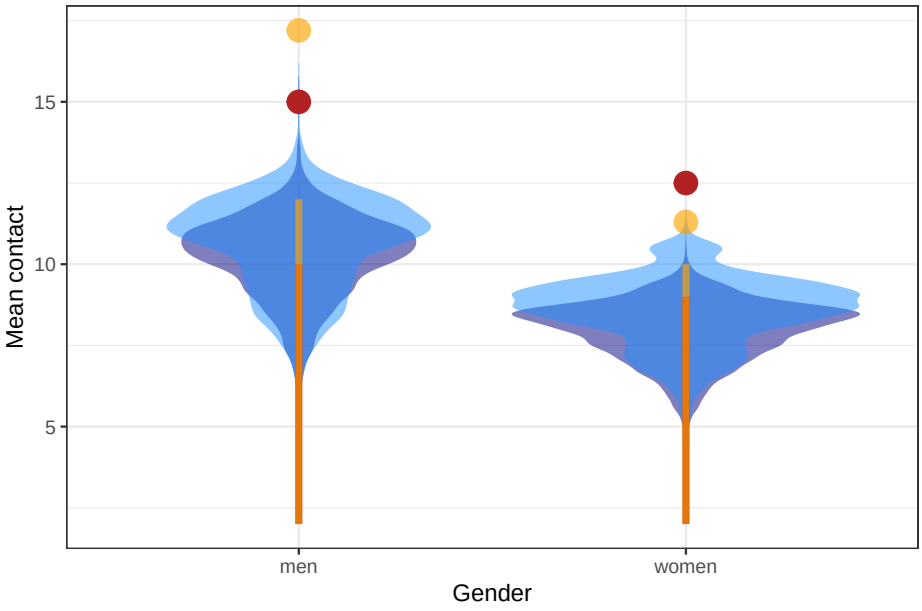

Figure S6. Comparison between our pandemic contact estimates by gender with those from [27]. Nelson and colleagues' [27] conducted a nationally representative study of 3,112 individuals in the US during two periods, August - December 2020 (red) and March - April 2021 (orange). Contacts were defined as interactions within 6 feet involving the exchange of at least 3 words or involving physical touch. Point ranges show the mean, 25th, and 75th percentiles; note that [27] truncated contacts at 50 per day. Blue violins show distribution of mean county-gender group estimates from the CTIS: dark blue for September - December 2020 and light blue for March - April 2021 to match the time periods in Nelson et al.

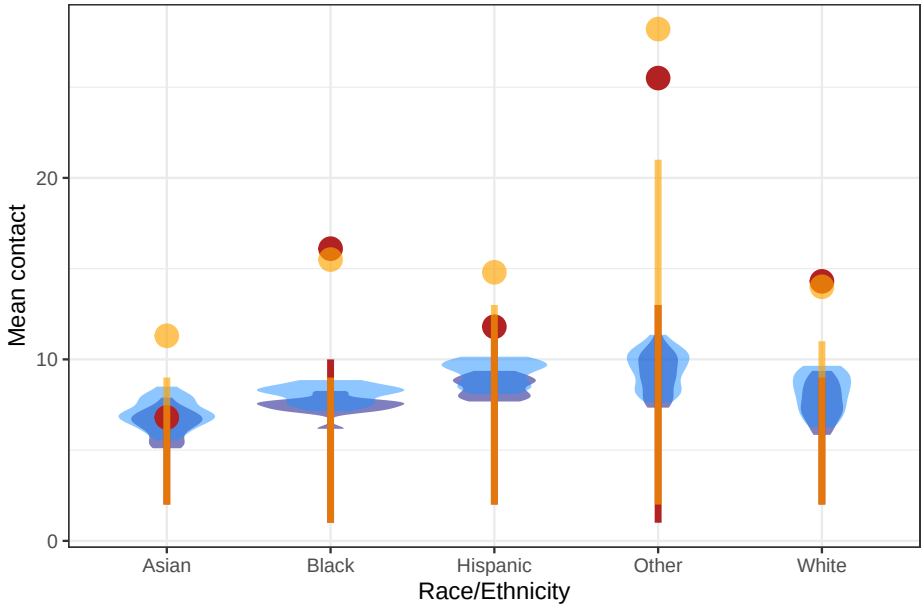

Figure S7. Comparison between our pandemic contact estimates by race/ethnicity with those from [27]. Nelson and colleagues' [27] conducted a nationally representative study of 3,112 individuals in the US during two periods, August - December 2020 (red) and March - April 2021 (orange). Contacts were defined as interactions within 6 feet involving the exchange of at least 3 words or involving physical touch. Point ranges show the mean, 25th, and 75th percentiles; note that [27] truncated contacts at 50 per day. Blue violins show distribution of mean state-race/ethnicity group estimates from the CTIS: dark blue for September - December 2020 and light blue for March - April 2021 to match the time periods in Nelson et al.

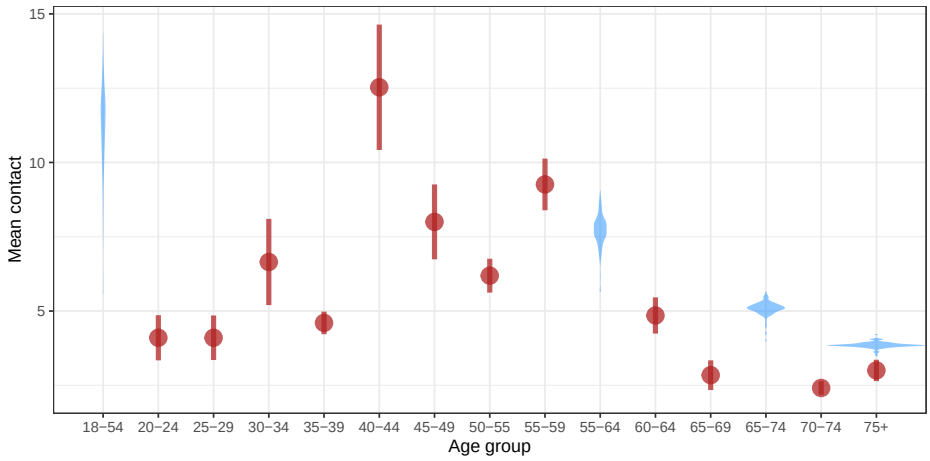

Figure S8. Comparison between our pandemic contact estimates by age with those from [25]. Dorelien and colleagues' [25] measured contact patterns in 2,083 children and adults in Minnesota, USA in April and May 2020. Contacts were defined as two-way conversations with 3+ words in the physical presence of another person or physical skin-to-skin contact. Red point ranges denote mean contact and one standard deviation above and below the mean from uncensored contact data. Blue violins show distribution of mean county-age group estimates for all 87 Minnesota counties from the CTIS for the first two weeks of May 2020 (these data are available in the CTIS but not as reliable).

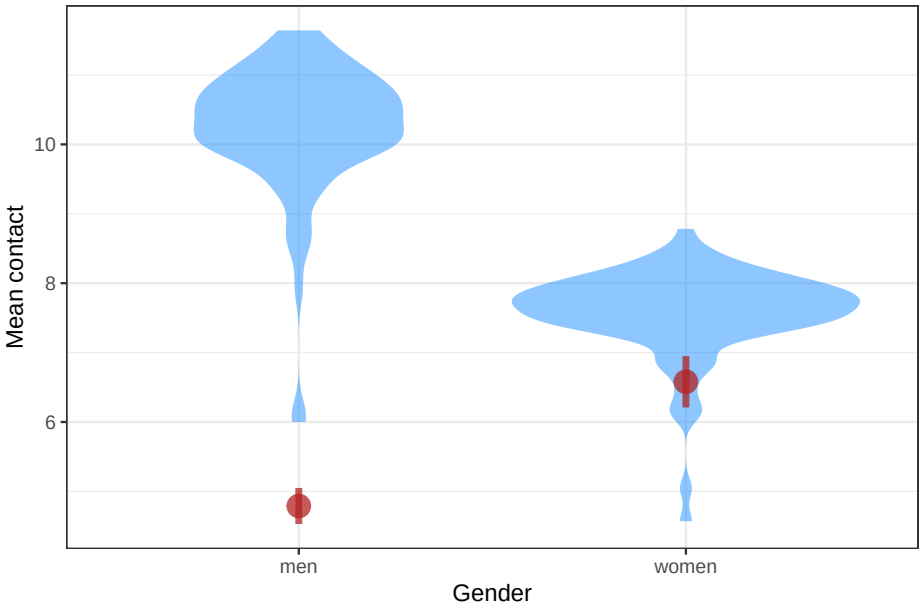

Figure S9. Comparison between our pandemic contact estimates by gender with those from [25]. Dorelien and colleagues' [25] measured contact patterns in 2,083 children and adults in Minnesota, USA in April and May 2020. Contacts were defined as two-way conversations with 3+ words in the physical presence of another person or physical skin-to-skin contact. Red point ranges denote mean contact and one standard deviation above and below the mean from uncensored contact data. Blue violins show distribution of mean county-gender group estimates for all 87 Minnesota counties from the CTIS for the first two weeks of May 2020 (these data are available in the CTIS but not as reliable).

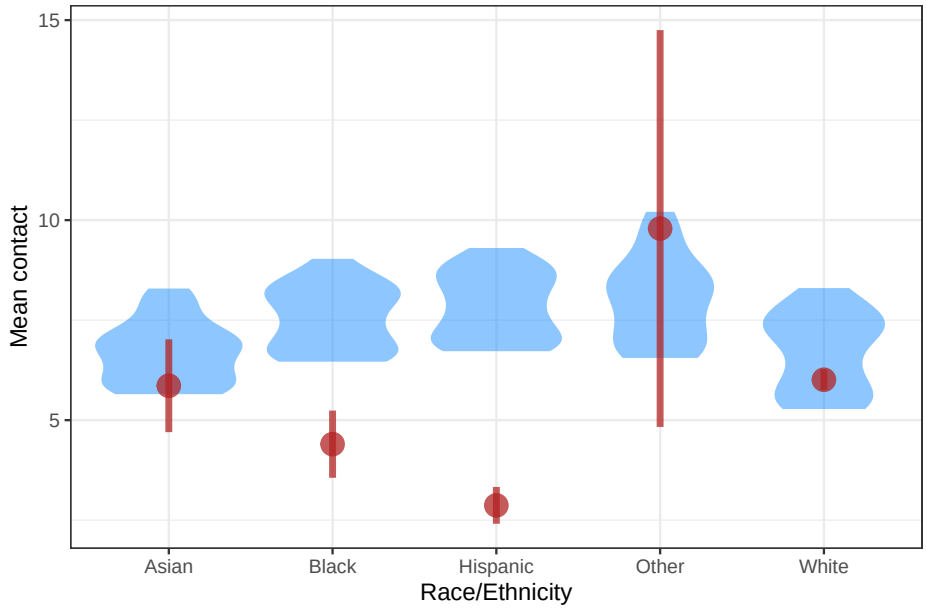

Figure S10. Comparison between our pandemic contact estimates by race/ethnicity with those from [25]. Dorelien and colleagues' [25] measured contact patterns in 2,083 children and adults in Minnesota, USA in April and May 2020. Contacts were defined as two-way conversations with 3+ words in the physical presence of another person or physical skin-to-skin contact. Red point ranges denote mean contact and one standard deviation above and below the mean from uncensored contact data. Blue violins show distribution of mean Minnesota race/ethnicity estimates from the CTIS for each week from September 2020 to April 2021 (a non-overlapping time period).

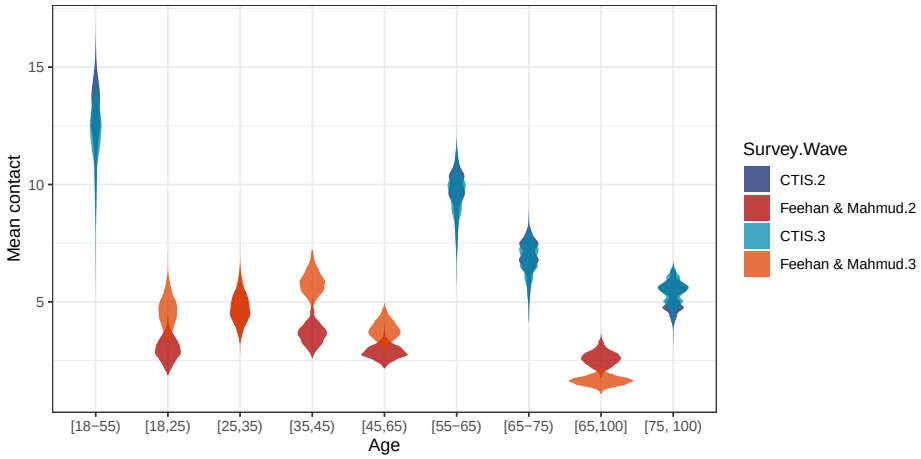

Figure S11. Comparison between our pandemic contact estimates by age with those from [26]. Feehan and Mahmud [26] conducted a nationally representative survey of contact patterns in the US including one week in June (Wave 2) and one week in September 2020 (Wave 3). Across the two waves, they had 2,641 respondents who reported fewer than 72 contacts. Red and orange violins show the bootstrapped mean contact for June and September 2020, respectively, using provided weights from [26]. (Note that we have imposed a truncation point on their data to match our analysis so the weights may no longer be accurate.) Blue and teal violins show the distribution of mean county-age group estimates from the CTIS for the same weeks in June and September, respectively.

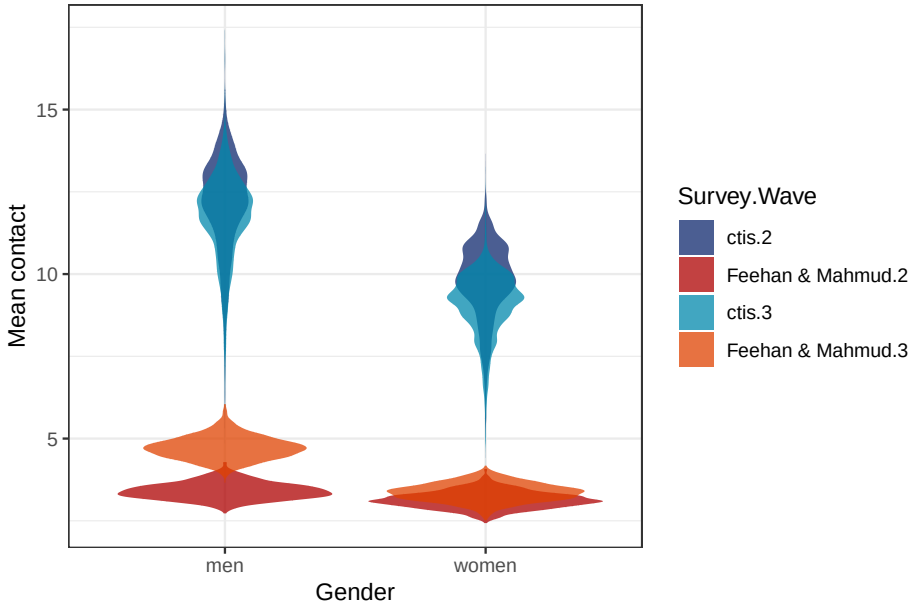

Figure S12. Comparison between our pandemic contact estimates by gender with those from [26]. Feehan and Mahmud [26] conducted a nationally representative survey of contact patterns in the US including one week in June (Wave 2) and one week in September 2020 (Wave 3). Across the two waves, they had 2,641 respondents who reported fewer than 72 contacts. Red and orange violins show the bootstrapped mean contact for June and September 2020, respectively, using provided weights from [26]. (Note that we have imposed a truncation point on their data to match our analysis so the weights may no longer be accurate.) Blue and teal violins show the distribution of mean county-gender estimates from the CTIS for same weeks in June and September, respectively.

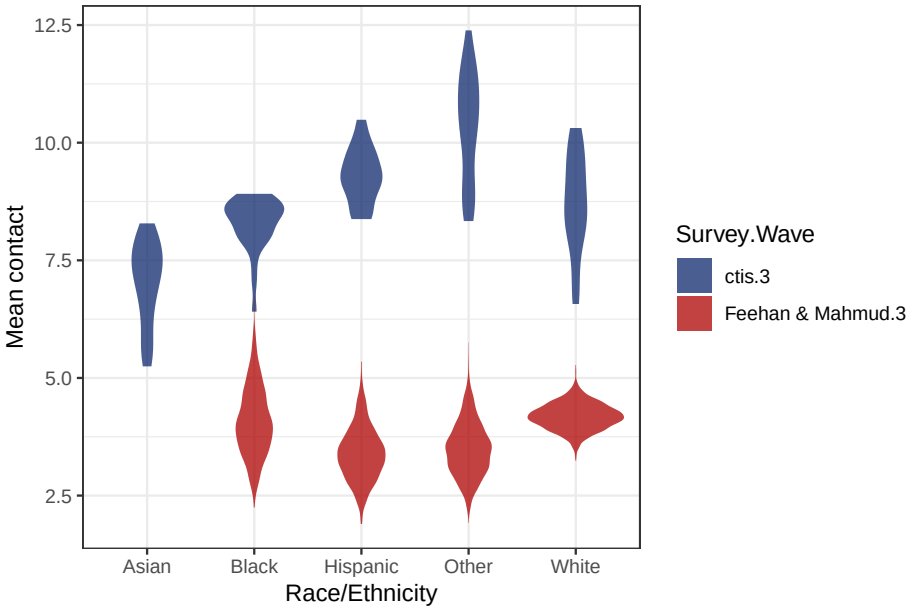

Figure S13. Comparison between our pandemic contact estimates by race/ethnicity with those from [26]. Feehan and Mahmud [26] conducted a nationally representative survey of contact patterns in the US including one week in September 2020 (Wave 3). In this wave, they had 1,514 respondents who reported fewer than 72 contacts. Red violins show the bootstrapped mean contact for September 2020, using provided weights from [26]. (Note that we have imposed a truncation point on their data to match our analysis so the weights may no longer be accurate.) Blue violins show distribution of mean state-race/ethnicity estimates from the CTIS for the same week in September 2020.

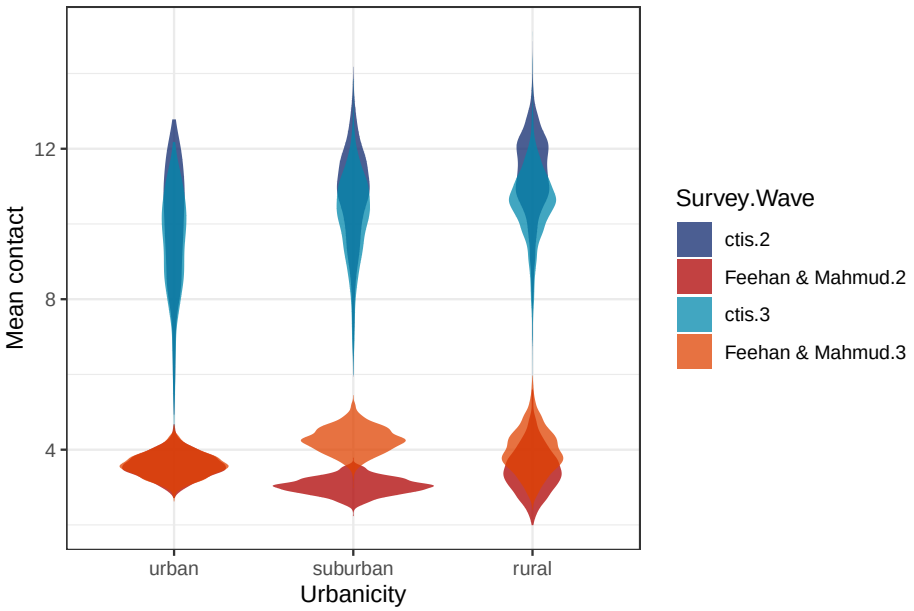

Figure S14. Comparison between our pandemic contact estimates by urbanicity with those from [26]. Feehan and Mahmud [26] conducted a nationally representative survey of contact patterns in the US including one week in June (Wave 2) and one week in September 2020 (Wave 3). Across the two waves, they had 2,641 respondents who reported fewer than 72 contacts. Respondents were classified as living in urban, suburban, or rural areas. Red and orange violins show the bootstrapped mean contact for June and September 2020, respectively, using provided weights from [26]. (Note that we have imposed a truncation point on their data to match our analysis so the weights may no longer be accurate.) Blue and teal violins show the distribution of mean county estimates from the CTIS for same weeks in June and September, respectively. CTIS data are at the county level; therefore, in this figure counties in NCHS classes 1 and 2 have been designated as urban, 3 and 4 as suburban, and 5 and 6 as rural for visualization purposes.

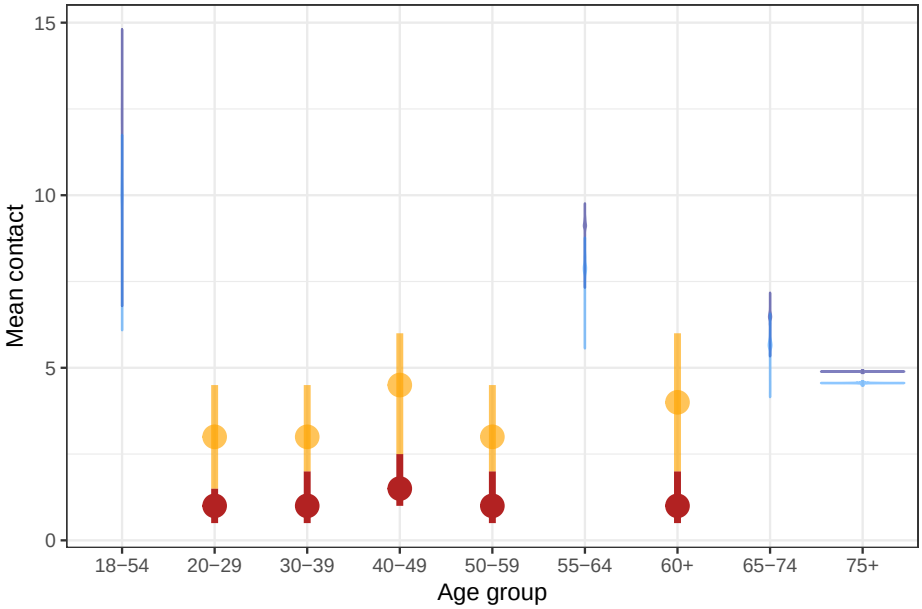

Figure S15. Comparison between our pandemic contact estimates by age with those from [48]. Kiti and colleagues' surveyed contact patterns across 647 employees of five private companies based in Georgia, USA over two days in April - June 2020 and November 2020 - January 2021. Contacts were categorized as interactions within 6 feet for less than 20 seconds, conversational contacts, or physical contacts, and included contacts in the home. Red and orange point ranges show the means and interquartile ranges of contact divided by two (to represent one day's worth of contacts) for April - June 2020 and November 2020 - January 2021 periods, respectively. Blue and navy violins show the distribution of mean county-age estimates for 158 counties in Georgia from the CTIS for the same time periods, respectively.

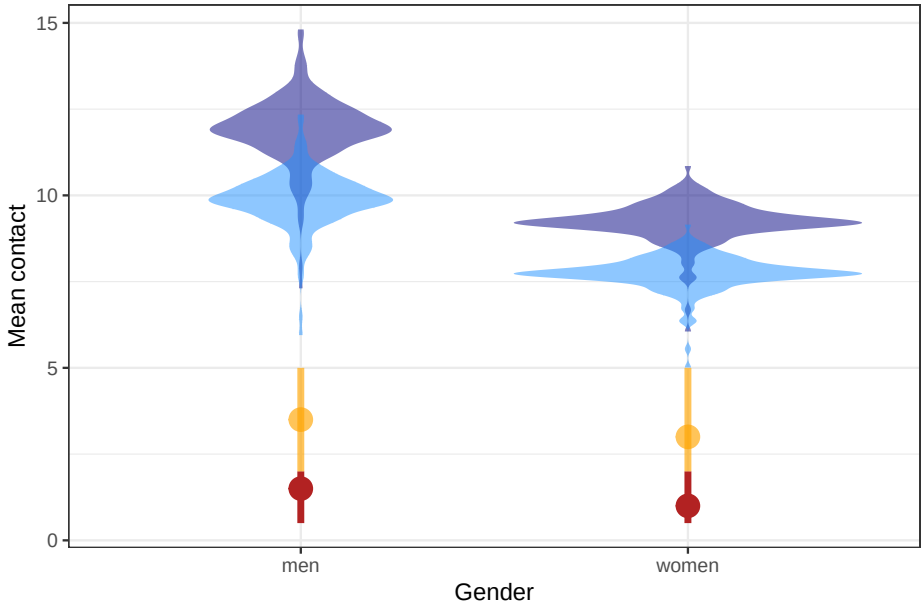

Figure S16. Comparison between our pandemic contact estimates by gender with those from [48]. Kiti and colleagues' surveyed contact patterns across 647 employees of five private companies based in Georgia, USA over two days in April - June 2020 and November 2020 - January 2021. Contacts were categorized as interactions within 6 feet for less than 20 seconds, conversational contacts, or physical contacts, and included contacts in the home. Red and orange point ranges show the means and interquartile ranges of contact divided by two (to represent one day's worth of contacts) for April - June 2020 and November 2020 - January 2021 periods, respectively. Blue and navy violins show the distribution of mean county-gender estimates for 158 counties in Georgia from the CTIS for the same time periods, respectively.

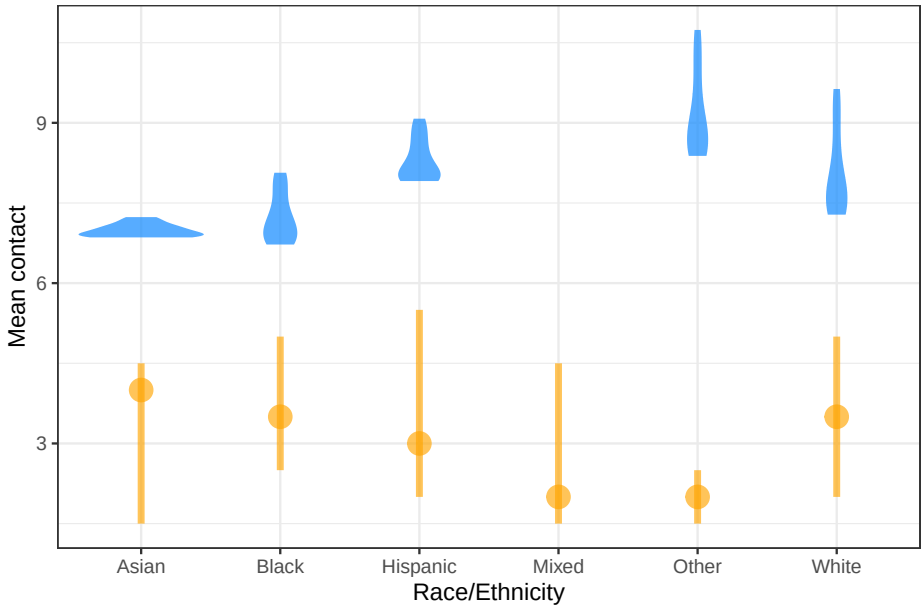

Figure S17. Comparison between our pandemic contact estimates by race/ethnicity with those from [48]. Kiti and colleagues' surveyed contact patterns across 647 employees of five private companies based in Georgia, USA over two days in November 2020 - January 2021. Contacts were categorized as interactions within 6 feet for less than 20 seconds, conversational contacts, or physical contacts, and included contacts in the home. Orange point ranges show the means and interquartile ranges of contact divided by two (to represent one day's worth of contacts) for November 2020 - January 2021 periods. Blue violins show the distribution of mean state-race/ethnicity estimates for Georgia from the CTIS for each week in the same time period.

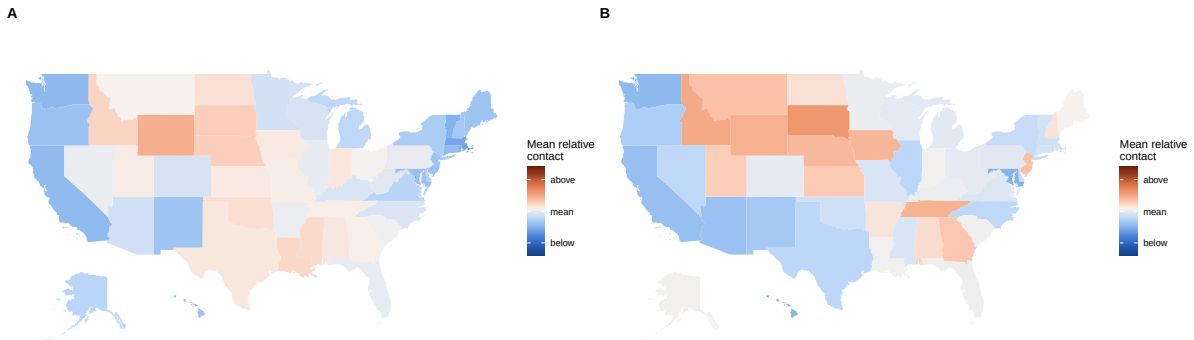

Figure S18. Comparison between (A) our findings and (B) Breen et al. [21] at the state level.

Supplementary Figures

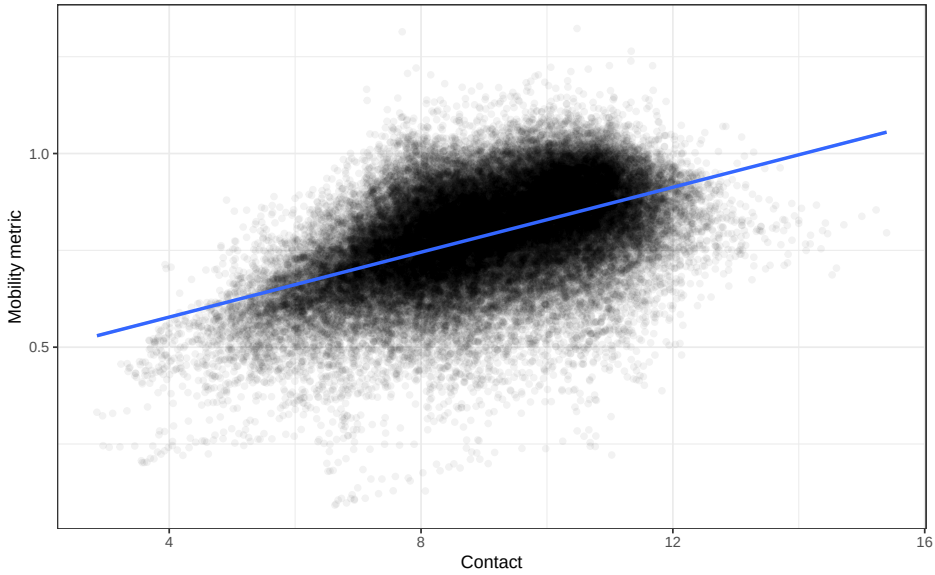

Figure S19. Mobility and contact are highly correlated. Data from September - December 2020. Mobility metric is the ratio of weekly 2019 to 2020 Safegraph visitor counts for each county.

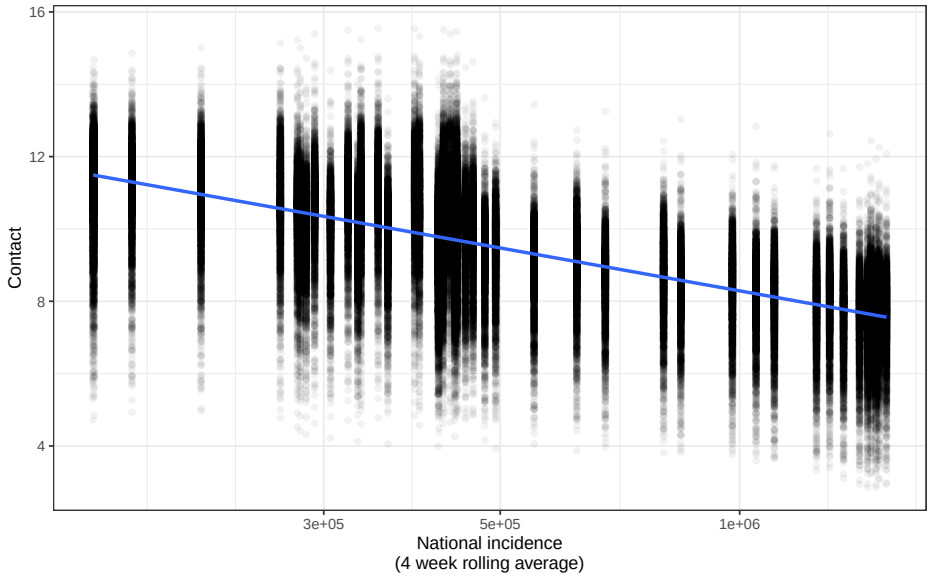

Figure S20. National incidence and contact are correlated. Data from June 2020 - April 2021. National incidence is the centered 4 week rolling average.

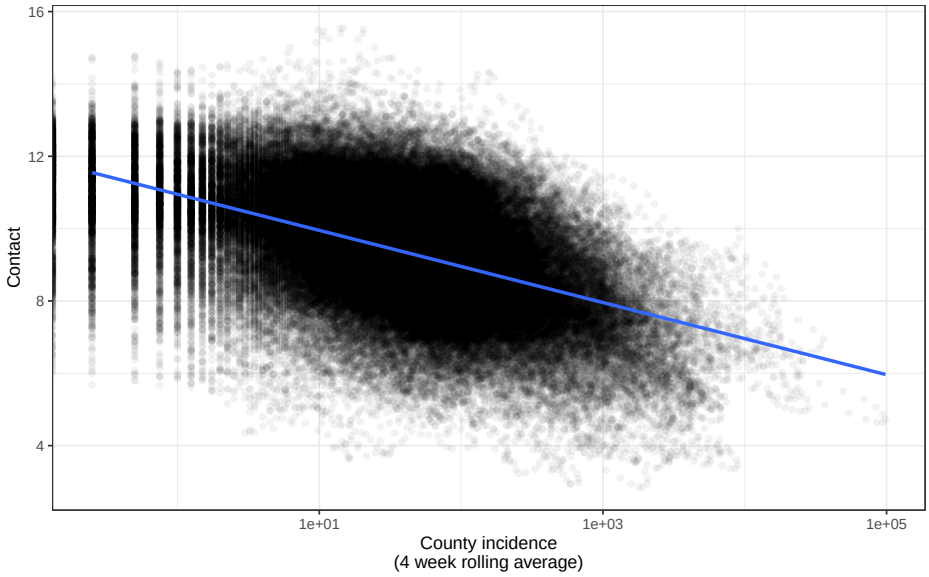

Figure S21. County incidence and contact are correlated. Data from June 2020 - April 2021. County incidence is the centered 4 week rolling average.

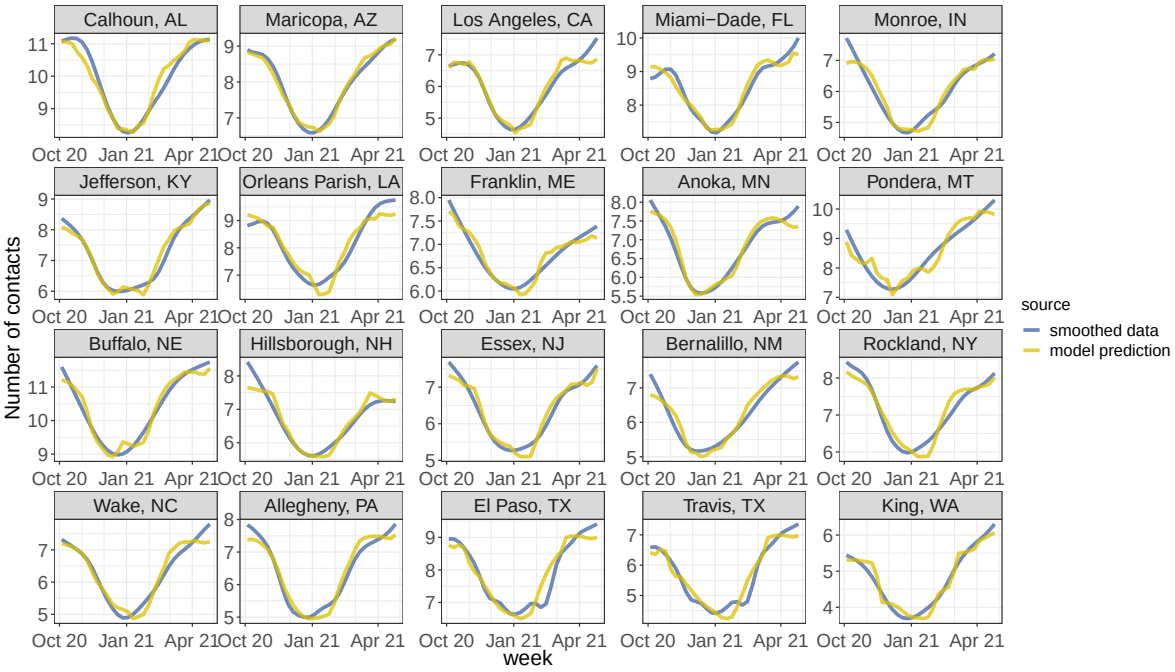

Figure S22. Smoothed contact data versus linear combination of incidence and policy terms in regression shows good fit between fluctuations in disease and contact.

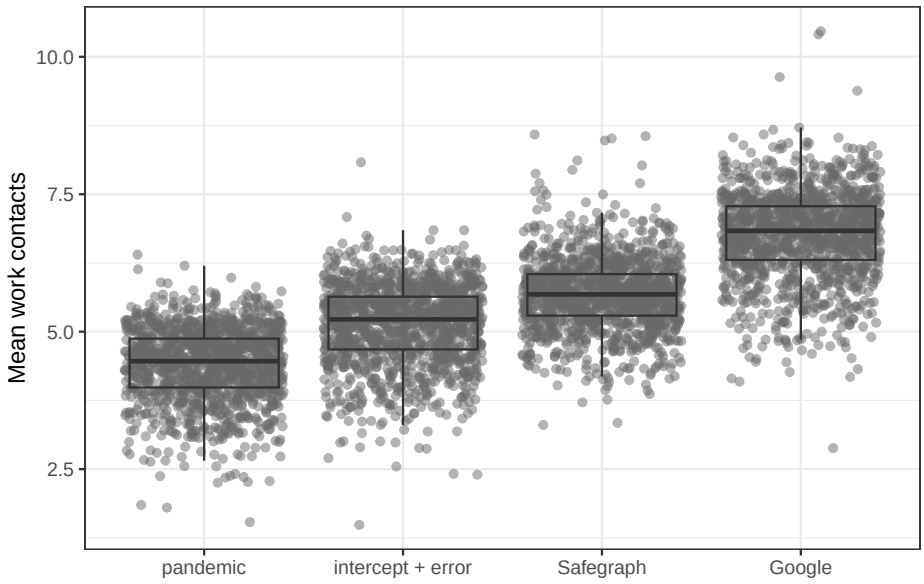

Figure S23. Using Google work-specific mobility data [49] to infer baseline work contacts yields higher estimates than Safegraph mobility data aggregated across all settings. Each point represents a county.

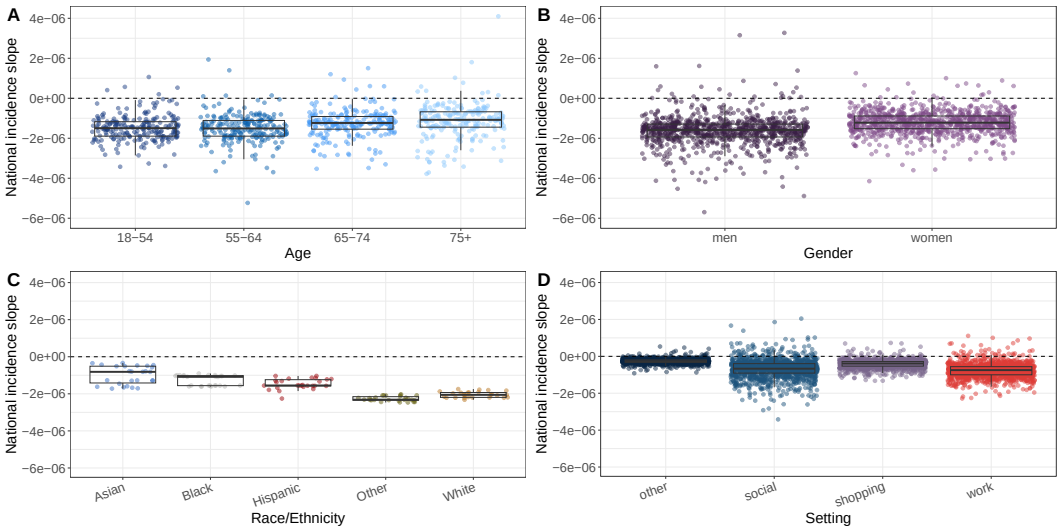

Figure S24. National disease incidence regression coefficient by age, gender, race/ethnicity, and setting. More negative coefficients indicate a stronger negative relationship between national incidence and non-household contacts, calculated during the baseline regression.

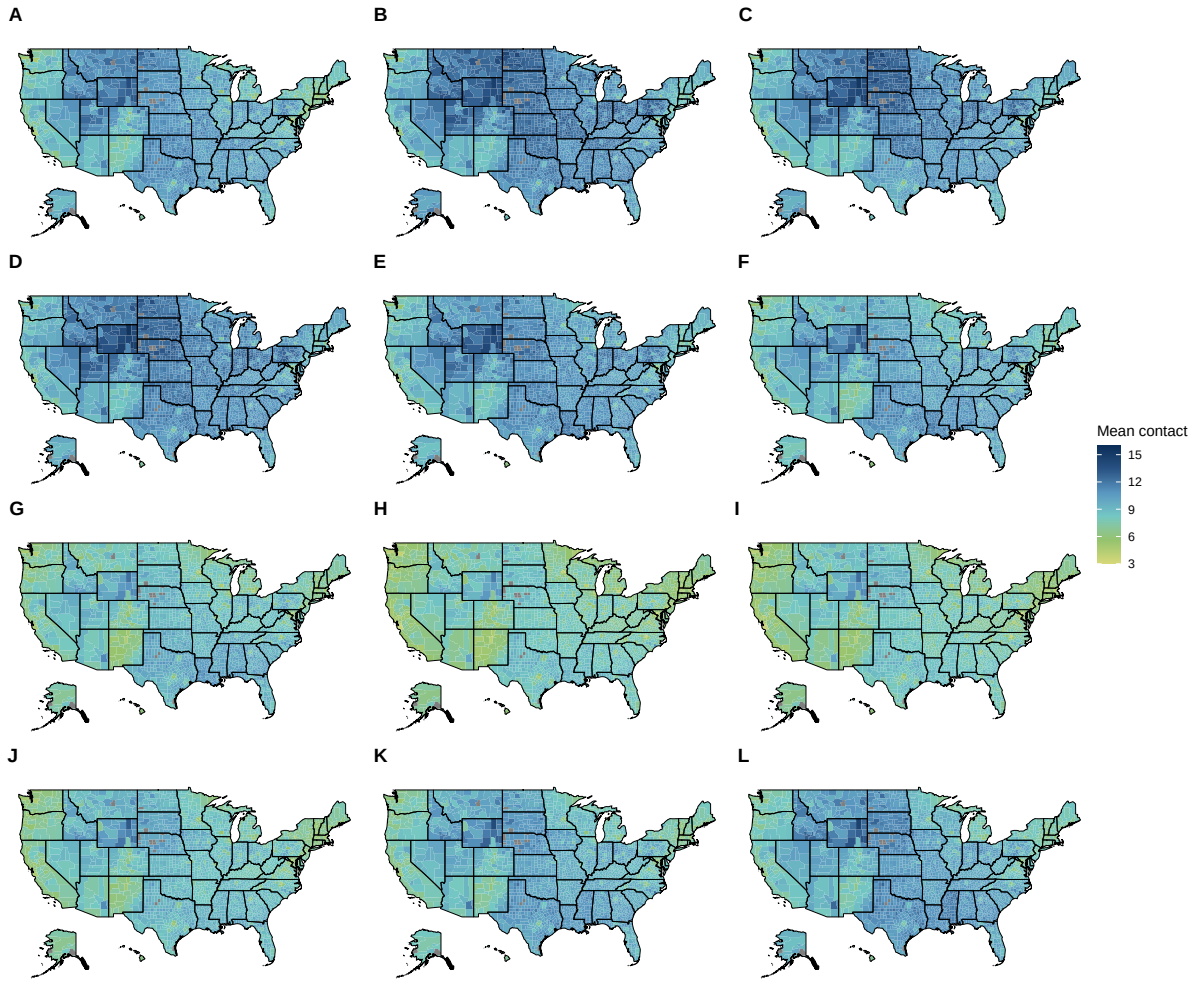

Figure S25. Fitted contact values by month. (A) May 2020 through (L) April 2021. While the mean changes, the spatial heterogeneity is relatively consistent.

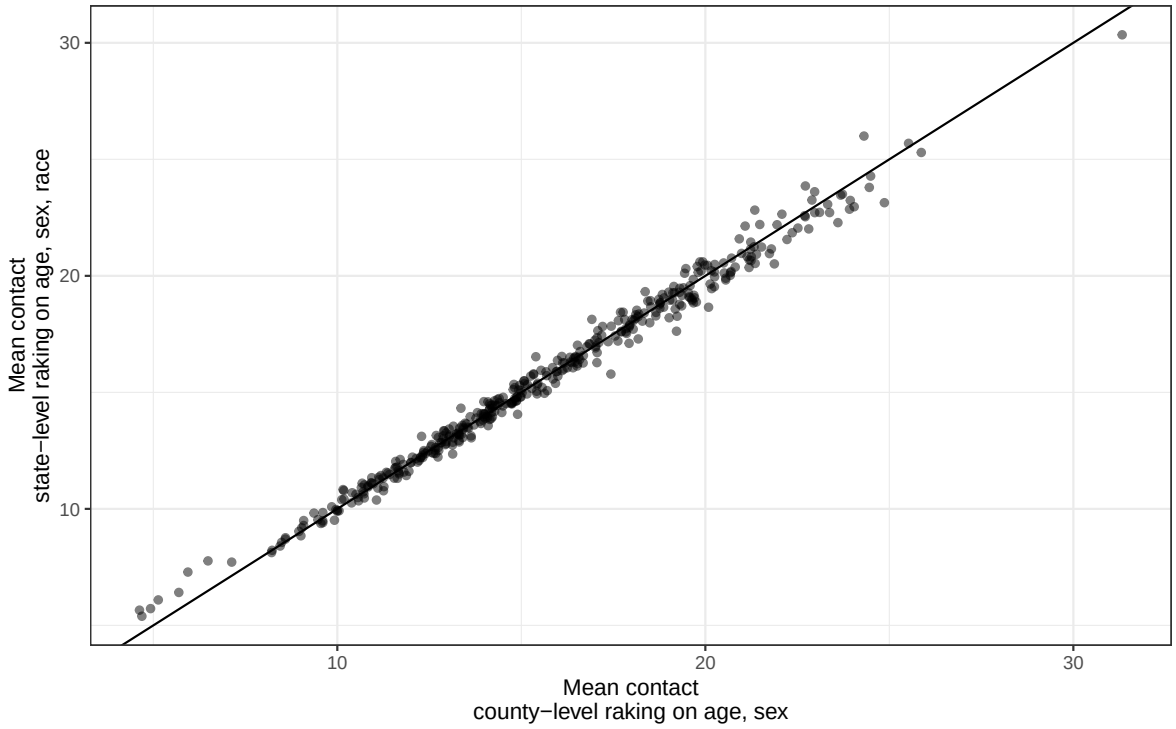

Figure S26. Pandemic contact estimates are minimally affected by raking on race in addition to age and sex. X-axis value is generated by raking at the county-level on age and sex and aggregating estimates to the state-month level. Y-axis value is generated by raking at the state-level on age, sex, and race and aggregating estimates to the monthly level.

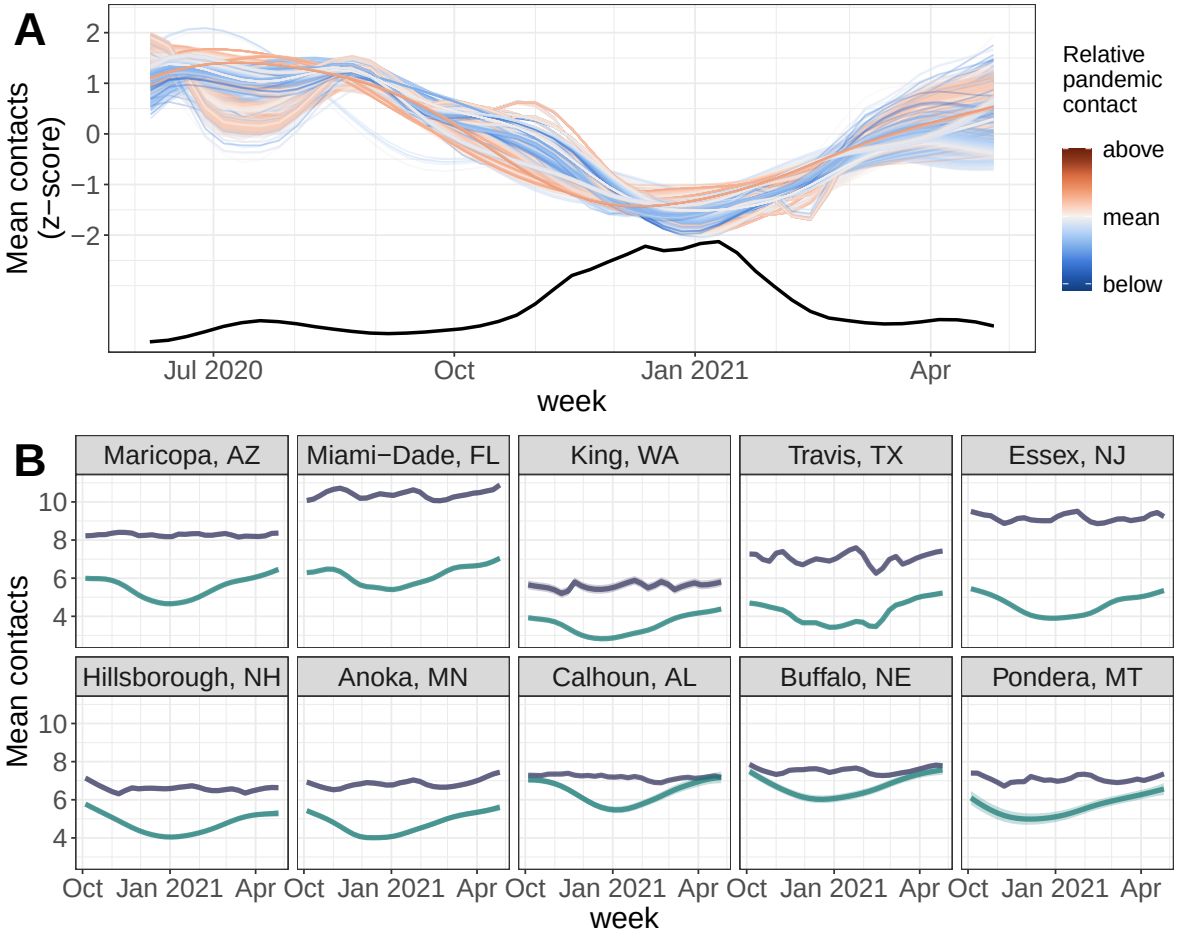

Figure S27. Counties have similar contact dynamics over time and relatively stable contact after controlling for disease using 90th percentile truncation (36 contacts). (A) Most counties had higher contact during the summer of 2020 and all had lower contact during the winter of 2020-21. Counties that experienced a dip in contact in summer 2020 were typically in states that exhibited higher incidence during that time. Each line represents a county colored by mean contact relative to the national mean; zscored contact relative to each county's mean is shown to allow comparison between time series despite the large range of mean contact values across counties. Black line shows the centered 3-week rolling average of national incidence for context. (B) Contact in the absence of disease (teal) is effectively constant over time compared to observed contact during the pandemic (blue). We controlled for disease using a linear regression predicting contact from county and national disease incidence and state policy data. This analysis is restricted to October 2020 to April 2021 to encompass a full wave of COVID-19. Shaded areas represent one standard error above and below the fitted contact value or estimated non-pandemic value.

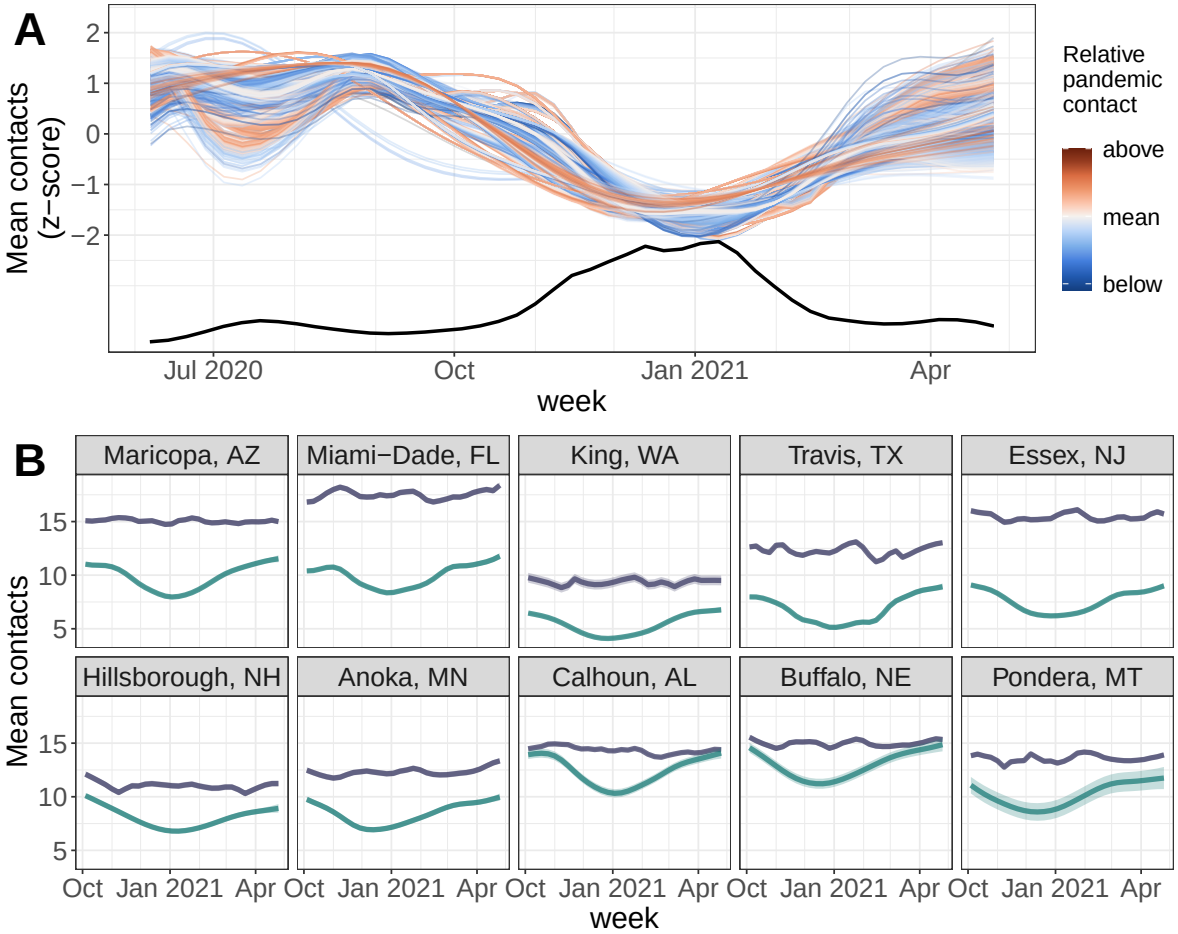

Figure S28. Counties have similar contact dynamics over time and relatively stable contact after controlling for disease using 97th percentile truncation (108 contacts). (A) Most counties had higher contact during the summer of 2020 and all had lower contact during the winter of 2020-21. Counties that experienced a dip in contact in summer 2020 were typically in states that exhibited higher incidence during that time. Each line represents a county colored by mean contact relative to the national mean; zscored contact relative to each county's mean is shown to allow comparison between time series despite the large range of mean contact values across counties. Black line shows the centered 3-week rolling average of national incidence for context. (B) Contact in the absence of disease (teal) is effectively constant over time compared to observed contact during the pandemic (blue). We controlled for disease using a linear regression predicting contact from county and national disease incidence and state policy data. This analysis is restricted to October 2020 to April 2021 to encompass a full wave of COVID-19. Shaded areas represent one standard error above and below the fitted contact value or estimated non-pandemic value.

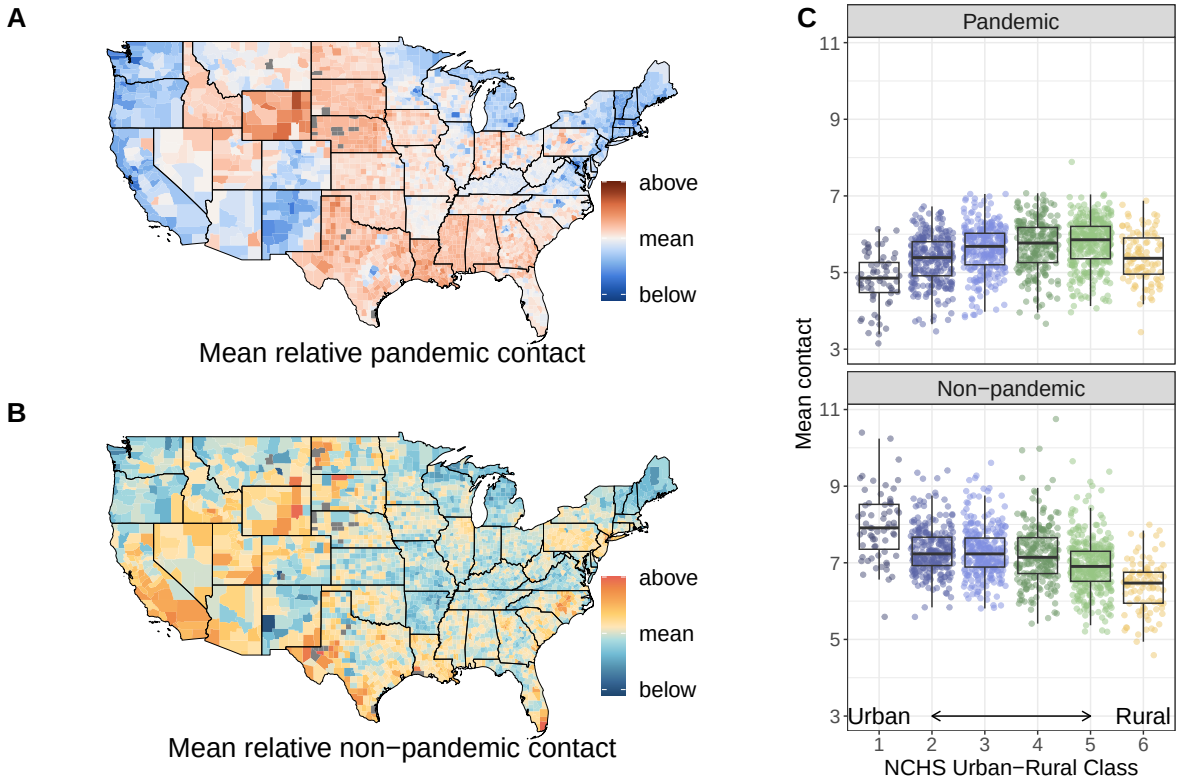

Figure S29. Contact is spatially heterogeneous regardless of disease incidence, but the urban-rural gradient reverses after controlling for disease incidence using 90th percentile truncation point (36 contacts). (A) Map of mean number of non-household contacts per county relative to the national mean during the pandemic (October 2020 - April 2021). There is high spatial heterogeneity in contact, even within states. Gray counties did not have sufficient sample size to estimate contact. (B) Map of inferred mean number of non-household contacts per county relative to the national mean in a baseline scenario. Spatial heterogeneity in contact remains high, though which counties are above and below the national mean has shifted from the pattern observed during the pandemic. (C) Mean number of contacts for each county decreases with increasing urbanicity during the pandemic but increases with urbanicity during inferred non-pandemic times. Only counties with 10 or more responses per week each week (Oct 2020 - April 2021) are included. Reversal of pandemic urban-rural gradient is stronger with this truncation.

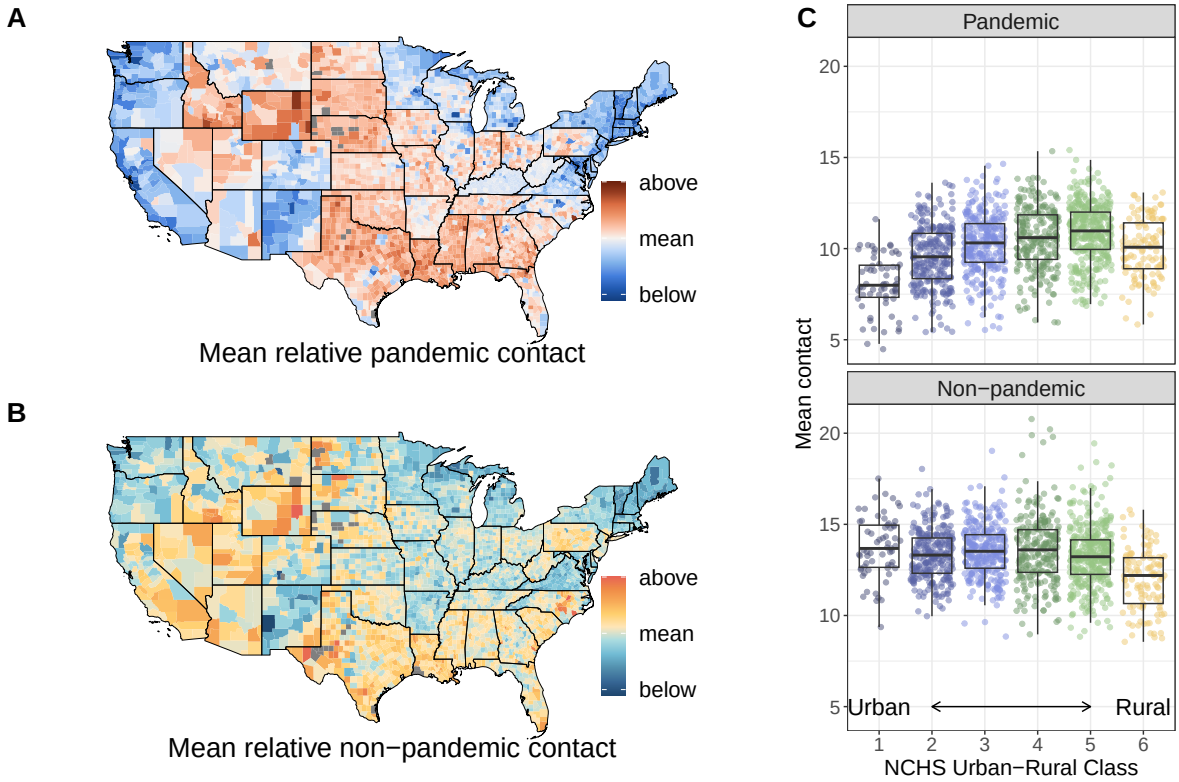

Figure S30. Contact is spatially heterogeneous regardless of disease incidence, but the urban-rural gradient reverses after controlling for disease incidence using 97th percentile truncation point (108 contacts). (A) Map of mean number of non-household contacts per county relative to the national mean during the pandemic (October 2020 - April 2021). There is high spatial heterogeneity in contact, even within states. Gray counties did not have sufficient sample size to estimate contact. (B) Map of inferred mean number of non-household contacts per county relative to the national mean in a baseline scenario. Spatial heterogeneity in contact remains high, though which counties are above and below the national mean has shifted from the pattern observed during the pandemic. (C) Mean number of contacts for each county decreases with increasing urbanicity during the pandemic but increases with urbanicity during inferred non-pandemic times. Only counties with 10 or more responses per week each week (Oct 2020 - April 2021) are included.

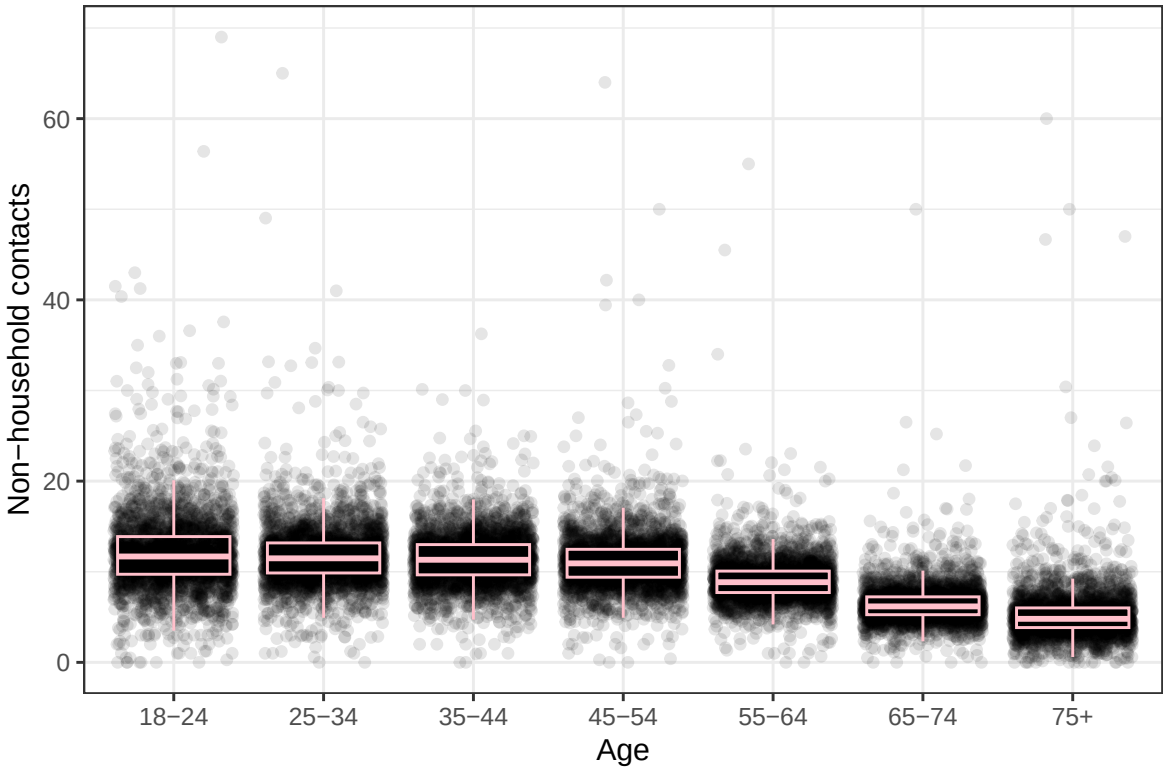

Figure S31. Age groups 18-54 have roughly the same number of contacts. Each point represents a county, raw mean contact for the period May 2020 through April 2021.

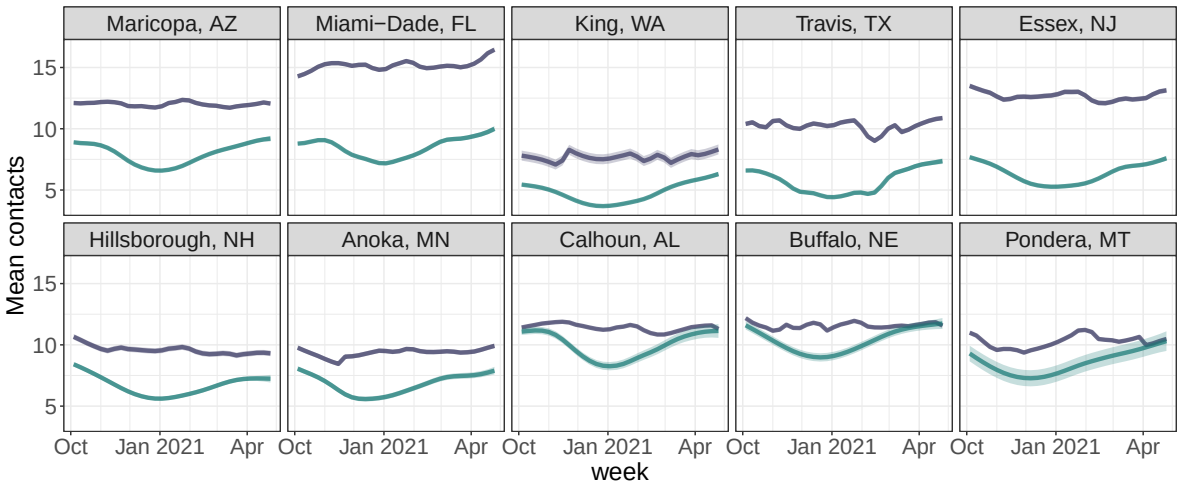

Figure S32. Baseline contact estimates do not change meaningfully when county incidence is excluded from linear regression model, compare to Figure 1B.

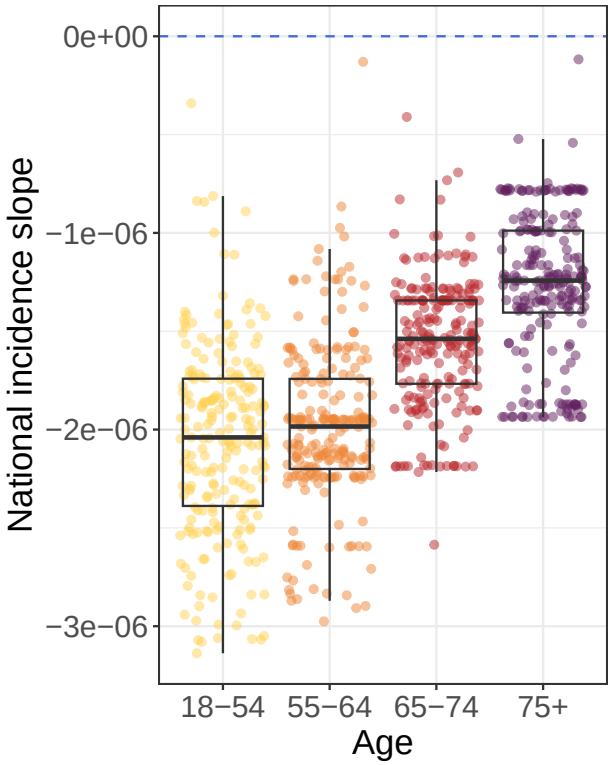

Figure S33. Regression coefficients by age group do not change qualitatively when county incidence is excluded from linear regression model, compare to Figure 3A.

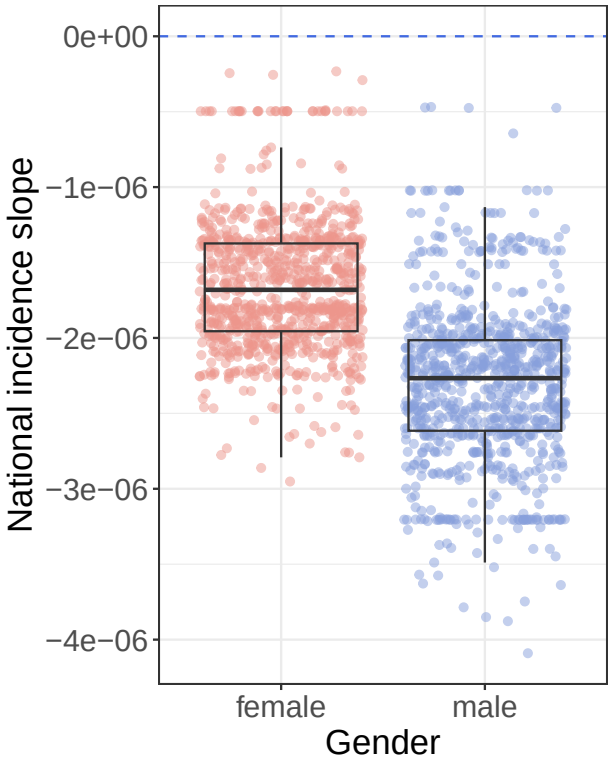

Figure S34. Regression coefficients by gender do not change qualitatively when county incidence is excluded from linear regression model, compare to Figure 3B.

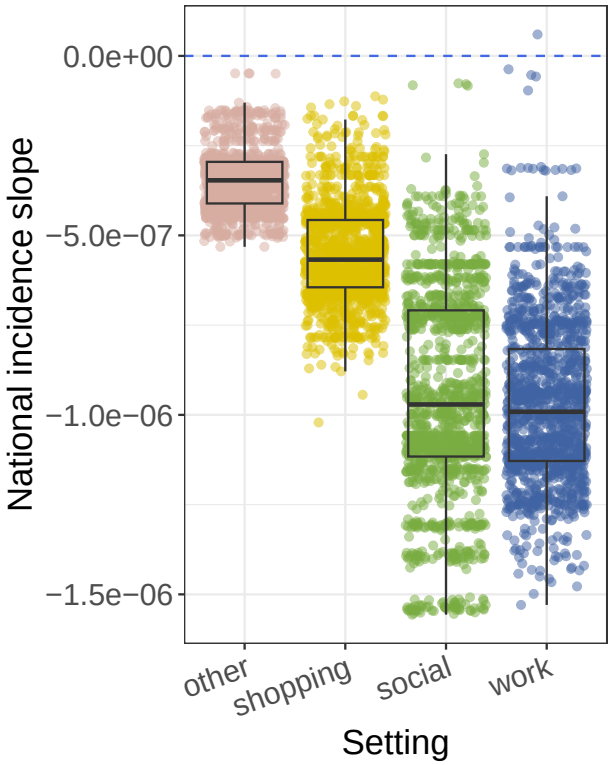

Figure S35. Regression coefficients by contact setting do not change qualitatively when county incidence is excluded from linear regression model, compare to Figure 3D.

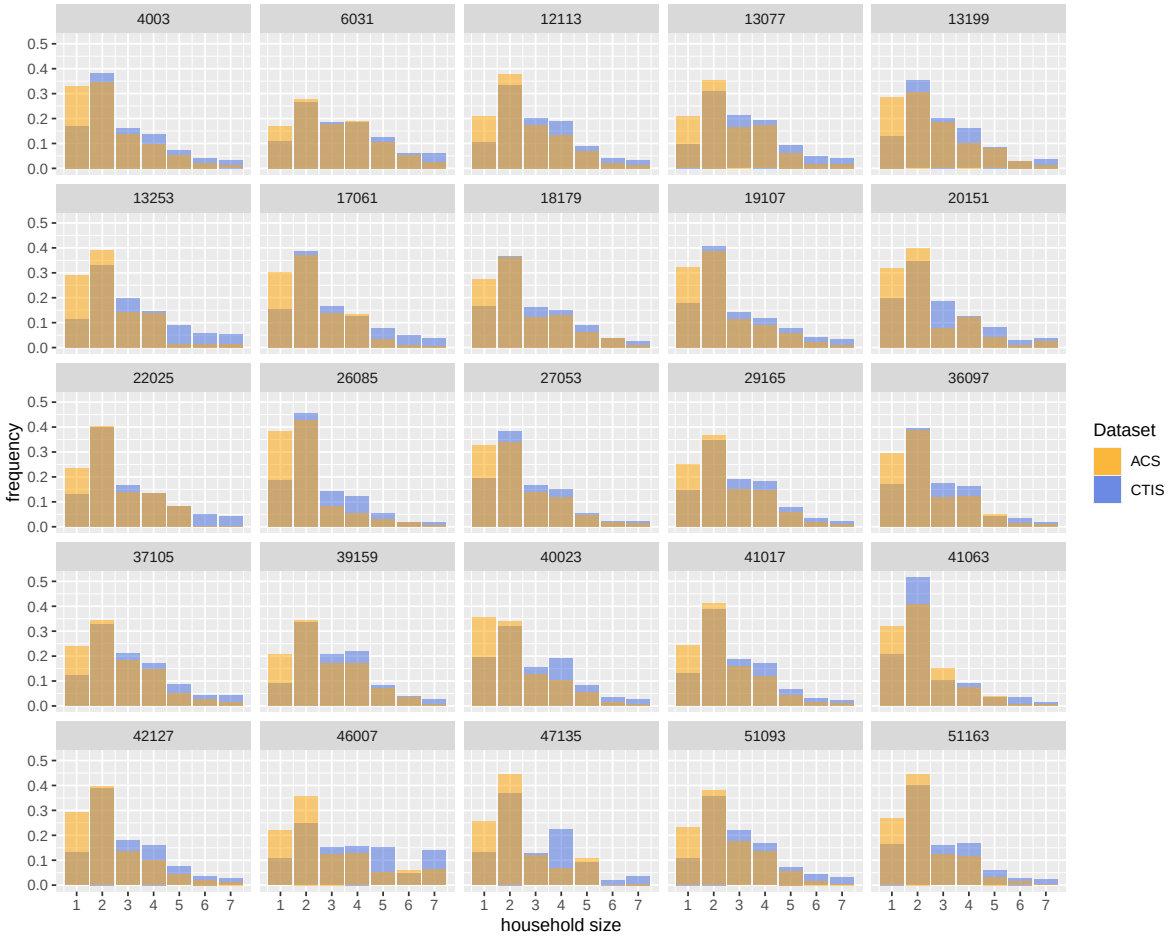

Figure S36. Comparison between household size in raked contact data and census data for a subsample of counties shows that survey decently captures range of household sizes and we do not need to rake on household size.

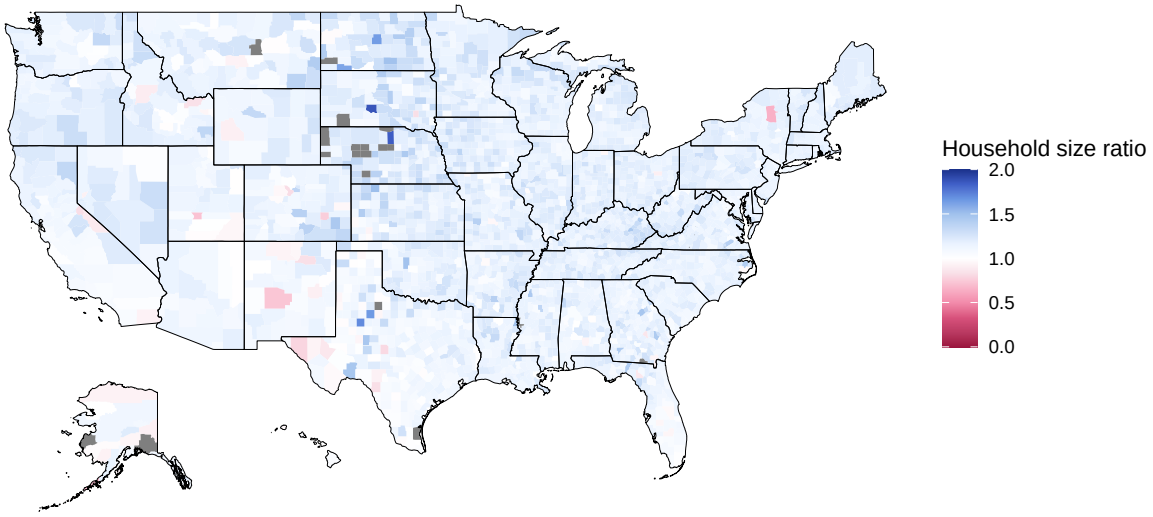

Figure S37. Ratios of mean survey household size and census household size are close to 1.