Science Advances
AAAS

advances.sciencemag.org/cgi/content/full/sciadv.abc3054/DC1

# Supplementary Materials for

## Social connections with COVID-19-affected areas increase compliance with mobility restrictions

Ben Charoenwong, Alan Kwan* and Vesa Pursiainen

*Corresponding author. Email: apkwan@hku.hk

Published 23 October 2020, *Sci. Adv.* **6**, eabc3054 (2020)
DOI: 10.1126/sciadv.abc3054

**This PDF file includes:**

Sections S1 to S4
Tables S1 to S9
Fig. S1

# Supplementary materials

The supplementary materials include additional details on the variable definitions, summary statistics, and robustness checks.

# S1 Additional details on data sources

## S1.1 SafeGraph mobility data

Regarding coverage and quality, on an average day from February 1st, 2020 to March 31st, 2020, SafeGraph observes 18.75 million devices, approximately 5.6% of the U.S. population and about 10% of mobile devices. According to SafeGraph's analysis of user characteristics, SafeGraph posits that its sample is representative of the U.S. population based on its own study of income characteristics, age, and demographics of its users. In its full data available for academic use, SafeGraph calculates how long a user spends at home and/or work, how far the person has traveled from home, and how long the person stays at a given location away from home. A person is measured by a smartphone.

In our sample, the measure of home and work is derived from data prior to the pandemic using data in January 2020, based on "frequent day locations" and "frequent evening locations", where the former is used to define a "work place" and the latter is used to define "home". We believe this measure of social distancing is superior to explicitly measuring the distance travelled for the average individual within a region, since different regions are subject to different average travel distances and may introduce bias in our empirical analyses. In addition, for public health concerns, the actual distance travelled in itself is not of interest. The true interest for social distancing is in whether an individual interacts with others at any particular location. For that reason, we take the stance that looking at the percentage of people's flexible hours an individual spends completely at home is a useful metric for public health officials and policymakers. Common device locations are re-estimated every month as devices enter and exit the dataset.

To determine a home location, SafeGraph analyses 6 weeks of data during night-time hours (between 6 pm and 7am), and applies an undisclosed minimum data availability threshold in terms of the total data points as well as number of distinct days to assign a common night-time location for the device. All of these home locations are based on a census block group, the highest geographical resolution for which U.S. Census provides demographic information. Data for census block groups with fewer than 4 devices are excluded for privacy reasons. To determine a work location, SafeGraph uses 1 month of data to determine where the device is most frequently during traditional work hours and is not during the weekend or overnight.

Both the measure of home and work are subject to measurement bias. First, the number of mobile phones for which there are data changes over time and differs by region. Second, the classification of home and work is noisy and may flip home and work locations for night-time workers like nurses and doctors. Whether these measurement errors would introduce estimation bias and the direction of the bias in our empirical results would depend on whether the errors are systematically correlated with our measures of interest, either directly or through confounding unobserved variables that affect our measures of interest.

Regarding geographical bias in the first two points above, SafeGraph data documentation states that they have tested for geographical bias by comparing state-by-state numbers of home locations of the devices to the true proportions reported by the 2016 U.S. Census and find that they are similar, with the overall average percentage point difference $\leq 1\%$, a maximum bias of 3% and minimum bias of -3% in each state. In addition, a technical blog by data scientists at SafeGraph in October 2019 provides additional analyses with raw data samples and findings. Therefore, we do not expect large biases in our empirical results in any particular direction.

Regarding the measurement of home relative to work, the SafeGraph data will be biased in the home relative to work classification precision towards counties with more formal work as opposed to informal work. Entry into the sample depends on smartphone usage and mobile app usage across regions, so if smartphones are more used by professionals for work reasons compared to leisure, then states with lower cellphone penetration will tend to have users with more systematic home and work information. We do not expect a particular direction of bias from this scenario, but we expect heteroscedasticity in the error terms in the regression, which is accounted for by our county-level clustering of standard errors. Informal work would appear as if an individual is not complying with social distancing measures. If a county with fewer social connections has more informal work, then the low social connectedness would be correlated with lower effect of mobility restrictions.

To the extent that the informal work may be essential work, this would bias our estimates and interpretation of our data. However, even if counties with more social connections have more informal work (e.g. if cities with more people have more social connections and also a larger informal work sector), if mobility restrictions decrease informal work more in areas with more informal work, our empirical estimates would actually be biased downward, and the estimated difference in the effect of mobility restrictions ascribed to social connections may be an underestimate of the true effect.

Finally, additional documentation materials on other variables in their dataset that is available for non-profit public use can be found at `https://docs.safegraph.com/docs/places-manual`, which we last accessed on May 18, 2020.

## S1.2 Facebook social connection data

The Facebook Social Connectedness Index (SCI) used in this paper was provided by Facebook, which provided the data for academic use as part of its Data For Good initiative. The data are anonymized and aggregated, with the most granular unit in the data being a U.S. county. The SCI is a cross-sectional measure based on Facebook relationships as of April 2016. The data are available in two versions, county-to-foreign country relationships and within United States county-to-county relationships. We use the county-to-foreign country relationships in the main empirical analyses in our paper while we use the county-to-county relationships in the supplementary materials for additional analyses on the mechanism of information and misinformation flow in social networks. The level of the SCI itself is not informative or interpretable, but it is interpreted as a relative measure of the total number of Facebook friendship links between individuals located in two areas as of April 2016.

For the county-to-foreign country SCI measure, the county with the most friendship links for each country is normalized to 1,000,000. This normalization permits a comparison across different U.S. counties and the connection with a particular foreign country, but not from one U.S. county to multiple foreign countries. For example, a value of 20,000 from Los Angeles county to Italy and 10,000 from Los Angeles county to China does not mean that there are twice the number of connections between Los Angeles to Italy than Los Angeles to China. However, a value of 20,000 from Los Angeles county to Italy and a value of 40,000 from Manhattan to Italy implies that Manhattan has twice the number of friendship with Italy as Los Angeles. The data are rounded to the nearest integer, and the bottom 100 counties within the United States for each foreign country is winsorized, and the data does not include links to countries with a population of less than 1 million people according to official statistical agencies.

For the county-to-county SCI measure, the county with the most friendship links within the matrix is normalized to 1,000,000 (as opposed to normalizing counties for each connected county $j$, which would have been analogous to the normalization in the county-to-foreign country measure), assigned to the Los Angeles to Los Angeles connections, the county with the largest number of friendship links. The normalized data is rounded to the nearest 0.0025. Unlike with the county-to-foreign country SCI measure, the relative levels can be compared for any three-way pair of counties. For example, if the connection between Los Angeles county to Los Angeles county is 1,000,000 and the connection from Los Angeles county to Manhattan "county" is 500,000, and connection from Manhattan "county" to Cook County, Illinois is 250,000, then we can interpret that there are twice as many friendship connections within Los Angeles as those from Los Angeles with Manhattan, and four times the number of friendships within Los Angeles compared to those from Manhattan to Cook County.

Relative friend probability is an alternative measure based on the SCI that further normalizes the SCI to account for the local population of each region. For the county-to-foreign country measure, the relative friend probability is defined as

$$RFP = 100 \times \frac{SCI_{i,j}}{Population_j},$$

where $j$ is the foreign country considered, reported in millions. For the U.S. county-to-county measure, the relative friend probability is defined as

$$RFP = 10^{12} \times \frac{SCI_{i,j}}{Population_i \times Population_j},$$

where the populations of counties $i$ and $j$ are as reported by the U.S. census for county-to-county links. Reported by Facebook, the scaling by 100 and $10^{12}$ is to minimize the number of decimal places. Similar to the SCI, the level of relative friend probability itself is not informative or interpretable.

Additional information on the data are available at `https://dataforgood.fb.com/tools/social-connectedness-index/`, which we last accessed on May 18, 2020.

There are two reasons we use the SCI indices for China and Italy separately instead of pooling them together into one variable. First, unfortunately, Facebook has scaled the indices in a way that means that the index values are not comparable between different countries. The SCI values allow cross-sectional comparison of different US counties based on their connectedness to any single country, e.g. China. So, we can say which counties are more connected to China, and by how much more. However, we cannot say if a county is more connected to China than to Italy due to the scale of the variables. This also means that aggregating connectedness variables between different countries is also difficult. For example, an average of SCI China and SCI Italy would be a meaningless number, because these two values involve different scaling.

Second, it is not clear that we would want to combine the two. Ex ante, it is not obvious that connections to China and Italy have a similar effect on the effect of mobility restrictions. While this is what we find, we believe that showing the distinction is interesting in and of itself. As we show in the Supplementary materials Table S6, connections with different countries have different effects, so seeing whether there is a difference between the two first epicenters of the pandemic is interesting. This is particularly true as the countries in question are culturally very different. Similarly, this distinction allows us to perform a more nuanced analysis of the differences in the effect of social connections. For example, in Table 4 we show that there are differences in the relative effect of China and Italy that depend on the racial composition of the county itself. This analysis would not be possible with an aggregated measure.

## S1.3 Crowdsourced mobility restriction data

We accessed a crowdsourced dataset on April 1, 2020 when there was data on all states, which we independently verify against primary sources. The data we use in this study include shelter-in-place orders, local religious gathering restrictions, school closures, and business closures. While certain types of mobility restrictions are likely to be more effective than others in particular states, for simplicity our restriction index is simply the count of restrictions in each category. This value ranges from 0 to 5. While most variation is at the state level, some non-pharmaceutical interventions vary at the county-level. To ensure that our efforts did not miss key events, we verify against other lists collected and updated after our sample period. For example, there are similar efforts from Keystone Strategy and *socialdistancing.stanford.edu*. We retrieve the list from Keystone Strategy on May 31, 2020 - the latter was not released publicly at the time of our study. At the county-day level, our list of restrictions is 92% correlated with the same index calculated from Keystone strategy, suggesting we did not fail to incorporate substantial county-level mobility restrictions.

## S1.4 Additional variable definitions

In addition to the variable definitions discussed above, the remaining definitions include, “N cases”, the number of confirmed COVID-19 cases in a county which we use in calculating the variable $\ln(Cases_{i,t})$ from Equation 3. “N deaths” is the cumulative number of confirmed COVID-19-related deaths in the county. Median age is the median age of people in the county based on the latest American Community Survey data. Republican is an indicator variable taking the value of one if the vote share for Republicans was higher than for Democrats in the 2016 presidential election. Education is the share of the county population having a bachelor's degree or higher. White is the share of the county population classified as White in the U.S. Census. Asian is the share of the county population classified as Asian or Pacific Islander in the U.S. Census. Obesity rate is the share of the county population classified as obese. Diabetes rate is the share of the county population diagnosed with diabetes.

# S2 Summary statistics

Over 20 states initiated their first mobility restriction on March 16, 2020. States typically implement various non-pharmaceutical interventions simultaneously, and also appear to follow a ranking in terms of the restrictiveness. On average, states appear to first close schools and public places, then restrict gatherings and finally implement stay-at-home orders. The top 5 states have over three mobility restrictions, and throughout our sample from February 1 to March 30, 2020, the average state has 0.629 restrictions. By the end of the sample, the average number of mobility restrictions across states is 3.1.

Table S1 shows summary statistics for our sample at the county level. There is a large variation across counties and days in social distancing and the fraction of time spent at home. The county in the lowest 10th percentile spends on average 24.8% of flexible time at home while the county in the highest 10th percentile spends on average 44.6% of flexible time at home. The corresponding numbers for the fraction of those staying completely at home is 16.3% for the bottom 10th percentile and 35.1% for the top 10th percentile.

The number of restrictions is positively correlated with social distancing. Column 2 in Panel B of Table S2 shows the unconditional effects of mobility restrictions using the fixed effects specification we use in the main analysis. Unconditionally, social distancing increases when restrictions are imposed.

Finally, social connections to China and Italy are highly correlated at 0.948. The percentage of Asians within the county is 0.612 positively correlated with social connections to China and 0.590 positively correlated with social connections to Italy. The incidence of mobility restrictions also does not appear highly correlated with measures of social connections. County-level social connections to China and Italy are only 0.016 to 0.017 correlated with the number of restrictions.

**Table S1**
**Summary statistics**

Summary statistics for the county-day observations in the sample. The sample period is from February 1 to March 30, 2020.

| | Mean | Std | p10 | p50 | p90 |
|---|---|---|---|---|---|
| **Social distancing** | | | | | |
| Social distancing | 0.337 | 0.081 | 0.248 | 0.326 | 0.446 |
| Completely home | 0.248 | 0.074 | 0.163 | 0.236 | 0.351 |
| Full-time | 0.174 | 0.045 | 0.117 | 0.171 | 0.232 |
| Part-time | 0.100 | 0.037 | 0.054 | 0.096 | 0.150 |
| **Social Connectedness** | | | | | |
| SCI - China | 3519.549 | 10634.586 | 48.000 | 443.000 | 6692.000 |
| SCI - Italy | 4357.320 | 13331.747 | 64.000 | 539.000 | 8825.000 |
| Rel. prob. friend Chinese | 2.259 | 1.820 | 0.835 | 1.703 | 4.164 |
| Rel. prob. friend Italian | 2.707 | 1.983 | 1.081 | 2.131 | 4.995 |
| **Cases in county** | | | | | |
| N cases | 4.685 | 186.139 | 0.000 | 0.000 | 1.000 |
| N deaths | 0.068 | 3.190 | 0.000 | 0.000 | 0.000 |
| **Gov't restrictions** | | | | | |
| Restrictions (0-5) | 0.629 | 1.152 | 0.000 | 0.000 | 3.000 |
| Schools closed | 0.256 | 0.437 | 0.000 | 0.000 | 1.000 |
| Public places | 0.190 | 0.392 | 0.000 | 0.000 | 1.000 |
| Stay at home | 0.065 | 0.246 | 0.000 | 0.000 | 0.000 |
| Gatherings | 0.094 | 0.292 | 0.000 | 0.000 | 0.000 |
| Non-essential | 0.024 | 0.152 | 0.000 | 0.000 | 0.000 |
| **County variables** | | | | | |
| Population ('000) | 92.632 | 299.220 | 4.190 | 22.300 | 188.460 |
| Median age | 41.305 | 5.384 | 34.800 | 41.200 | 48.100 |
| Democrat | 0.154 | 0.361 | 0.000 | 0.000 | 1.000 |
| Republican | 0.846 | 0.361 | 0.000 | 1.000 | 1.000 |
| Education | 0.216 | 0.094 | 0.123 | 0.192 | 0.342 |
| **Race and ethnicity** | | | | | |
| White | 0.832 | 0.167 | 0.595 | 0.897 | 0.969 |
| Black | 0.090 | 0.145 | 0.003 | 0.023 | 0.301 |
| Asian | 0.014 | 0.028 | 0.001 | 0.006 | 0.031 |
| Native American | 0.019 | 0.074 | 0.001 | 0.003 | 0.027 |
| Other race | 0.045 | 0.046 | 0.013 | 0.032 | 0.092 |
| **Risk factors** | | | | | |
| Obesity rate | 0.328 | 0.057 | 0.253 | 0.330 | 0.397 |
| Diabetes rate | 0.104 | 0.038 | 0.061 | 0.099 | 0.155 |
| Population density ('000/sqm) | 0.233 | 1.522 | 0.004 | 0.038 | 0.359 |

**Table S1**
**Additional summary statistics (continued)**

Summary statistics for the county-day observations in the sample. The sample period is from February 1 to March 30, 2020.

| | Mean | Std | p10 | p50 | p90 |
|---|---|---|---|---|---|
| **County characteristics** | | | | | |
| Avg. household size | 2.124 | 0.387 | 1.627 | 2.155 | 2.563 |
| Personal income (USDbn) | 0.006 | 0.021 | 0.000 | 0.001 | 0.011 |
| PI per capita (USDk) | 44.145 | 12.745 | 32.620 | 41.981 | 57.069 |
| Stock participation | 0.162 | 0.068 | 0.082 | 0.158 | 0.244 |
| Self-employment | 0.167 | 0.039 | 0.122 | 0.163 | 0.217 |
| Unemployment | 0.036 | 0.028 | 0.013 | 0.030 | 0.062 |
| Avg. credit score | 698.484 | 28.313 | 658.999 | 701.329 | 732.419 |
| Social capital | -0.001 | 0.994 | -1.067 | -0.163 | 1.378 |
| Democrat vote | 31.606 | 15.249 | 15.012 | 28.382 | 53.215 |
| Republican vote | 63.683 | 15.546 | 41.827 | 66.778 | 80.907 |
| **Healthcare capacity** | | | | | |
| Hospitals per 100k | 5.981 | 9.846 | 0.000 | 2.742 | 15.049 |
| ICU beds per 100k | 14.681 | 28.043 | 0.000 | 0.000 | 38.622 |
| No health insurance | 0.101 | 0.051 | 0.047 | 0.092 | 0.168 |
| **Log SCI and case levels** | | | | | |
| ln(SCI - China) | 6.255 | 1.879 | 3.871 | 6.094 | 8.809 |
| ln(SCI - Italy) | 6.463 | 1.866 | 4.159 | 6.290 | 9.085 |
| ln(SCI - New York) | 1.785 | 2.299 | -0.904 | 1.439 | 5.061 |
| ln(N cases) | 0.206 | 0.730 | 0.000 | 0.000 | 0.693 |
| N | 187,965 | | | | |


| | Mean | Std | p10 | p50 | p90 |
|---|---|---|---|---|---|
| ln(SCI - China) | 6.255 | 1.879 | 3.871 | 6.094 | 8.809 |
| ln(SCI - Italy) | 6.463 | 1.866 | 4.159 | 6.290 | 9.085 |
| ln(SCI - New York) | 1.785 | 2.299 | -0.904 | 1.439 | 5.061 |
| ln(N cases) | 0.206 | 0.730 | 0.000 | 0.000 | 0.693 |
| N | 187,965 | | | | |

# S3 Economic magnitude and statistical importance of social connections

## S3.1 Economic magnitude of the effect of social connections

In the Results section, we mention that a one-standard-deviation increase in social connectedness to China and Italy increases the estimated effect of mobility restrictions by 45% and 47%, respectively, relative to a county with average social connectedness for each country. This section explains the calculation to obtain these numbers.

The average values of SCI China and SCI Italy are 3,520 and 4,357, and the standard deviations are 10,635 and 13,332 respectively. Thus, in natural logarithm form, these translate to 8.166 and 8.380. Correspondingly, when adding one standard deviation to the mean, the logarithm levels become 9.558 and 9.781, implying an increase of 1.392 and 1.401 in the log SCI respectively. Based on the estimated regression coefficients from columns 1 and 2 in Table 1 for the interactions with ln(SCI China) and ln(SCI Italy) of 0.0048 and 0.0052, this yields an incremental effect 0.0067 and 0.0073 respectively.

To get the average effect of restrictions from our estimates, we multiply the logarithmic average values of SCI China and SCI Italy of 8.166 and 8.380 by the interaction coefficients 0.0048 and 0.0052 and add the corresponding estimates of the restrictions coefficients of -0.0245 and -0.0280, which yields an estimated effect of restrictions of 0.0147 and 0.0156 for the counties with "average-level" connectedness to China and Italy.

Dividing the marginal increase of 0.0067 and 0.0073 from above by the corresponding average level effects translate into 45% and 47% of the average effect of restrictions for social connections with China and Italy respectively.

## S3.2 Statistical importance of restrictions and social connections

In this subsection, we study the incremental improvement in the $R^2$ measure for statistical importance. Starting from a benchmark model, we document that the incremental fraction of statistical variation in the social distancing (outcome variable) measure is captured by including the social connection measures. For clarity, we consider both an empirical specification excluding our fixed effects specification as well as a version with the fixed effects specification that we use in our analysis. The reason for including versions is that the fixed effects, due to their semi-parametric and high-dimensional nature, will capture a large fraction of variation.

In our baseline specifications with no fixed effects, we find that the differences in the estimated effects of mobility restrictions statistically attributable to social connections are 10% as important as the main effect of mobility restrictions themselves. Panel A of Table S2 shows the regression output when including only the log local number of cases, then adding mobility restriction measures, followed by social connections, and finally mobility restrictions interacted with social connection measures. Moving from column 1 to 2, including the mobility restrictions improves the $R^2$ measure from 0.145 to 0.228, or an increase of 0.083, equivalent to a relative increase of approximately 57%. Further including social connection measures, which only contain cross-sectional variation from Facebook users as of April 2016 improves the fit slightly to 0.232. Finally, including the mobility restrictions interacted with mobility restrictions increases $R^2$ from 0.232 to 0.241, for an increase of 0.009, equivalent to a relative increase of approximately 3.9%. Comparing the increase in $R^2$ going from columns 3 to 4 with the improvement from column 1 to 2 shows that the relative improvement in statistical fit from including the social connection interacted with mobility restriction gives an $R^2$ improvement of around 10%. This number suggests accounting for the differential effects of mobility restrictions due to social connections is about 10% of the relative statistical importance of the main effect from mobility restrictions themselves.

In the specification including the full state-by-day and county fixed effects used in our main analyses, we find a larger improvement in relative statistical fit due to differences in estimated effects of mobility restrictions statistically attributable to social connections. Panel B of Table S2 shows the regression output when including only the set of fixed effects, log local number of cases, then adding mobility restriction measures, followed by mobility restrictions interacted with social connection measures. Moving from column 1 to 2, including the log local cases improves the $R^2$ by 0.02. However, further including mobility restrictions does not generate a noticeable increase in $R^2$ as most of the variation in mobility restrictions is at the state-level, which gets absorbed by the state-by-day fixed effects. Despite that, moving from column 3 to 4 by including the mobility restrictions interacted with mobility restrictions generates a noticeable increase in $R^2$ from 0.749 to 0.756, for an increase of 0.007. This increase is 35% relative to the improvement when including log local cases, over 5 times larger than the 6% relative increase in the specification with no fixed effects (which is calculated by dividing the increase in $R^2$ from column 3 to 4 by the $R^2$ in column 1. In other words, this compares the improvement due to the interaction of mobility restrictions and social connections with the baseline when only including log local cases). Therefore, when using this more stringent empirical specification, we find that the relative statistical importance of the difference in estimated effects of mobility restrictions statistically attributable to social connections is higher than a baseline model.

These results not only suggest that social connections with China and Italy statistically significantly increase compliance with restrictions, but also that they are statistically important. Finally, we note that even if the goodness of fit measure $R^2$ did not improve noticeably, it would not mean social connections are not economically important on the margin. Although, on average, there may be many personal or county-specific drivers of social distancing, the information flow from social connections may still have a large impact on behavior. That is, given all the personal and county-specific drivers, the information flow may be incrementally important in affecting whether a person increases or decreases social distancing. However, in this case, our results suggest social connections with China or Italy are both statistically and economically significant for the estimated effects of mobility restrictions.

**Table S2**
**R-squared analysis**

The dependent variable is *Social distancing*. *Restrictions* is the number of the five restriction types currently adopted in the county. *SCI China* and *SCI Italy* are the Social Connectedness Index values between the county and China and Italy, respectively. *N cases* is the current number of confirmed cases in the county. Heteroscedasticity-consistent standard errors, double-clustered by county and day, are shown in parentheses.

**Panel A: SCI China and Italy, excluding fixed effects**

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| ln(N cases) | 0.0418*** | 0.0244*** | 0.0261*** | 0.0157*** |
| | (0.0028) | (0.0012) | (0.0014) | (0.0021) |
| Restrictions (0-5) | | 0.0239*** | 0.0234*** | -0.0047 |
| | | (0.0040) | (0.0040) | (0.0071) |
| ln(SCI - Italy) | | | 0.0068*** | 0.0053*** |
| | | | (0.0014) | (0.0014) |
| Restr. x ln(SCI Italy) | | | | 0.0041*** |
| | | | | (0.0007) |
| ln(SCI - China) | | | -0.0085*** | -0.0084*** |
| | | | (0.0013) | (0.0014) |
| Restr. x ln(SCI China) | | | | 0.0007 |
| | | | | (0.0006) |
| N | 184,831 | 184,831 | 184,831 | 184,831 |
| $R^2$ | 0.145 | 0.228 | 0.232 | 0.241 |

**Significance levels: * 0.1, ** 0.05, *** 0.01.**

**Panel B: SCI China and Italy, with fixed effects**

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| ln(N cases) | | 0.0245*** | 0.0244*** | 0.0113*** |
| | | (0.0013) | (0.0013) | (0.0009) |
| Restrictions (0-5) | | | 0.0058 | -0.0280*** |
| | | | (0.0050) | (0.0042) |
| Restr. x ln(SCI Italy) | | | | 0.0042*** |
| | | | | (0.0005) |
| Restr. x ln(SCI China) | | | | 0.0011** |
| | | | | (0.0005) |
| State-Day FE | Yes | Yes | Yes | Yes |
| County FE | Yes | Yes | Yes | Yes |
| N | 185,356 | 184,766 | 184,766 | 184,766 |
| $R^2$ | 0.729 | 0.749 | 0.749 | 0.756 |

**Significance levels: * 0.1, ** 0.05, *** 0.01.**

# S4 Robustness checks and additional empirical results

This section in the supplementary materials presents five additional robustness tests and empirical evidence consistent with relevant information flow through social connections. First, we perform a number of additional analyses to confirm that our results are not driven by differences in risk. We find qualitatively similar results studying the sub-sample excluding counties in New York and counties with large Asian fraction of the population. We also show that obesity as a proxy for COVID-19 risk yields results similar to that using diabetes rates, discussed in the main text.

Second, we consider alternative measures of social connections which account for the population within a county in addition to the number of Facebook users. We find similar results as our main empirical specification, which suggests our results are not simply due to differential participation in the Facebook social network.

Third, we study whether social connections with China and Italy amplify the behavioral responses in a county in response to additional local cases. We find larger responses to local cases in social distancing. However, these effects are dominated by the impact of social connections and mobility restriction measures, consistent with existing economic theory for rational inattention.

Fourth, we study the effect of social connections with foreign countries other than China or Italy and social connections within the United States. Adding other, ex-ante less informative, countries does not affect the impact of social connections with China or Italy, suggesting that our empirical results are not simply due to higher social connections to foreign countries overall.

Finally, we also consider including intra-United States social connections by studying connections to New York and an SCI-weighted number of local cases as a measure of local transmission of COVID-19 information in Section S4.5. We find that social connections with counties within the United States that are more affected by COVID-19 also increase estimated effects of mobility restrictions, consistent with our interpretation of the relevant informational role of social connections.

## S4.1 Additional proxies for risk

In Table S3, we perform analyses with different sample restrictions as well as an interaction analysis including the obesity rate of the county. First, in columns 1 and 2, we exclude all counties based in the state of New York and the adjacent states (Connecticut, Massachusetts, New Jersey, Pennsylvania, Rhode Island, and Vermont). This is because New York had experienced the worst outbreak in the U.S. in our sample period from February 1, 2020 through March 31, 2020, and may thus represent the highest levels of perceived risk. Second, we exclude the 20% of U.S. counties with the highest share of Asians (columns 3 and 4). With these subsamples, our main results remain qualitatively unchanged and statistically significant. In columns 5 and 6, we study the interaction between obesity levels and the effect of social connections, given obese people are likely to be more vulnerable to COVID-19. Similar to the results on diabetes, we find that more obese populations comply with restrictions better and are less affected by additional information via social connections.

**Table S3**
**Robustness: Additional proxies for risk**

The dependent variable is *Social distancing*. *Restrictions* is the number of the five restriction types currently adopted in the county. *SCI China* and *SCI Italy* are the Social Connectedness Index values between the county and China and Italy, respectively. *Obesity* is the share of county population classified as obese. *N cases* is the current number of confirmed cases in the county. Heteroscedasticity-consistent standard errors, double-clustered by county and day, are shown in parentheses.

| | Ex. NY and adj. states | | Ex. top 20% Asian | | Obesity | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Restr. x ln(SCI China) | 0.0048*** | | 0.0052*** | | 0.0088*** | |
| | (0.0004) | | (0.0005) | | (0.0011) | |
| Restr. x ln(SCI Italy) | | 0.0052*** | | 0.0056*** | | 0.0093*** |
| | | (0.0004) | | (0.0005) | | (0.0011) |
| Restr. x ln(SCI China) x Obesity | | | | | -0.0132*** | |
| | | | | | (0.0030) | |
| Restr. x ln(SCI Italy) x Obesity | | | | | | -0.0133*** |
| | | | | | | (0.0030) |
| Restr. x Obesity | | | | | 0.0497** | 0.0522** |
| | | | | | (0.0199) | (0.0204) |
| Restrictions (0-5) | -0.0249*** | -0.0280*** | -0.0258*** | -0.0295*** | -0.0370*** | -0.0414*** |
| | (0.0040) | (0.0043) | (0.0041) | (0.0046) | (0.0076) | (0.0078) |
| ln(N cases) | 0.0126*** | 0.0119*** | 0.0100*** | 0.0091*** | 0.0107*** | 0.0099*** |
| | (0.0009) | (0.0009) | (0.0009) | (0.0009) | (0.0009) | (0.0009) |
| State-Day FE | Yes | Yes | Yes | Yes | Yes | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 168,069 | 168,069 | 147,795 | 147,795 | 184,766 | 184,766 |
| $R^2$ | 0.744 | 0.745 | 0.734 | 0.735 | 0.756 | 0.757 |

**Significance levels: * 0.1, ** 0.05, *** 0.01.**

## S4.2 Relative friend probabilities

Relatedly, one might ask if the Social Connectedness Index captures level of connectedness rather than friendships composition. To disentangle the composition of friendships from a general level of connectedness in a county, Table S4 repeats our main analysis with relative friend probabilities rather than the SCI. Rather than scaling by the product of the number of Facebook users in a geography pair, relative friend probabilities (RFP) scaling by populations, defined as $RFP_{i,j} = 10^{12} \times \frac{Social\ Connectedness\ Index_{i,j}}{Population_i \times Population_j}$. We find similar results. A one-standard-deviation increase in the RFP with China increases the effect of mobility restrictions on social distancing by 33% and 37%, respectively. As in the main specification, the RFP with Italy exhibits a larger effect. When placing connections to China and Italy in the same specification, the measure for Italy renders the effect of RFP with China statistically insignificant.

**Table S4**
**Gov't restrictions vs. social connectedness**

The dependent variable is *Social distancing*. *Restrictions* is the number of the five restriction types currently adopted in the county. *RFP China* and *RFP Italy* are the relative friend probabilities of the county for China and Italy, respectively. *N cases* is the current number of confirmed cases in the county. Heteroscedasticity-consistent standard errors, double-clustered by county and day, are shown in parentheses.

| | (1) | (2) | (3) |
|---|---|---|---|
| Restr. x RPF China | 0.0012*** | | 0.0002 |
| | (0.0002) | | (0.0002) |
| Restr. x RPF Italy | | 0.0017*** | 0.0016*** |
| | | (0.0003) | (0.0003) |
| Restrictions (0-5) | 0.0040 | 0.0036 | 0.0034 |
| | (0.0046) | (0.0048) | (0.0047) |
| ln(N cases) | 0.0224*** | 0.0213*** | 0.0212*** |
| | (0.0012) | (0.0012) | (0.0012) |
| State-Day FE | Yes | Yes | Yes |
| County FE | Yes | Yes | Yes |
| N | 184,766 | 184,766 | 184,766 |
| $R^2$ | 0.750 | 0.750 | 0.750 |

**Significance levels: * 0.1, ** 0.05, *** 0.01.**

## S4.3 Social connections and local threat level

Beyond the transmission of information from abroad, another first order consideration may be to study whether the social network carries information even absent of any policy change. If social connections indeed facilitate information flow, we should expect to find some effect on the number of local cases as well, as people with information from their social connections in China or Italy will take local cases more seriously and thus socially distance more.

Table S5 studies the interaction of social connections with local cases and also includes our main variable of interest, the social mobility restrictions interacted with social connection measures. Conceptually, we expect a weakly positive relation between incremental (as opposed to large, discrete) information flows and changes in behavior. Statistically, by disentangling any differential responses to local information and mobility restrictions, we are also able to identify whether the original estimated coefficient represents incremental response to the mobility restrictions as well as local cases.

We find that social connections appear to amplify the effect of local cases as well. Counties with more social connections to either Italy or China socially distanced more on average, with a one-standard-deviation increase in the log social connections increasing the social distancing effect of local cases by around 8%. However, these results appear dominated by the impact of mobility restrictions, as columns 2 and 4 show that when including mobility restrictions, the interaction effect of log local cases with social connections are either no longer statistically significant, or slightly change signs.

These empirical findings are in line with the existing literature on rational inattention. For example, one theory for behavior is as a function of information as a threshold condition, where some action is taken only when the information received is large enough to cross some threshold [29]. If the number of cases flow continually with no large jumps, then the information flow may not be sufficient to cause action. Therefore, we would expect smaller results if we interact the SCI with local cases.

**Table S5**
**Social distancing and local cases vs. social connections**

The dependent variable is *Social distancing*. *SCI China* and *SCI Italy* are the Social Connectedness Index values between the county and China and Italy, respectively. *N cases* is the current number of confirmed cases in the county. Heteroscedasticity-consistent standard errors, double-clustered by county and day, are shown in parentheses.

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| ln(Cases) x ln(SCI China) | 0.0008** | -0.0007* | | |
| | (0.0003) | (0.0003) | | |
| Restr. x ln(SCI China) | | 0.0049*** | | |
| | | (0.0004) | | |
| ln(Cases) x ln(SCI Italy) | | | 0.0008** | -0.0005 |
| | | | (0.0004) | (0.0003) |
| Restr. x ln(SCI Italy) | | | | 0.0053*** |
| | | | | (0.0005) |
| Restrictions (0-5) | | -0.0262*** | | -0.0292*** |
| | | (0.0042) | | (0.0045) |
| ln(N cases) | 0.0171*** | 0.0182*** | 0.0167*** | 0.0160*** |
| | (0.0035) | (0.0032) | (0.0038) | (0.0032) |
| State-Day FE | Yes | Yes | Yes | Yes |
| County FE | Yes | Yes | Yes | Yes |
| N | 184,766 | 184,766 | 184,766 | 184,766 |
| $R^2$ | 0.749 | 0.756 | 0.749 | 0.756 |

**Significance levels: * 0.1, ** 0.05, *** 0.01.**

## S4.4 Social connections with other foreign countries

We study the role of social connections to China and Italy, as they were the first countries to experience major outbreaks of COVID-19 and hence were substantially ahead of the United States in terms of the spread of the epidemic. Therefore, social connections with these countries are likely to be useful for Americans for obtaining information about the seriousness of the threat. Using a similar logic, connections with countries that are *behind* America in terms of the spread of the virus are likely to be less useful for obtaining information, and serve as a test for the information channel of social connections. To test this prediction, we perform a similar analysis using the social connectedness with a number of large countries that U.S. citizens are likely to have significant social ties with, but which have experienced later and/or less severe COVID-19 outbreaks than the U.S. We classify countries as less severe than the U.S. if the date when they reached a daily death rate of 0.1 deaths per million was after the U.S., and if their daily death rate as of March 31, 2020, was lower than that of the U.S.

Table S6 above shows a series of tests incrementally including more countries close to the United States as well as close to China. The results are consistent with our interpretation in the paper. In comparison with China and Italy, social connections with countries with less experience of COVID-19 than the U.S. are generally not associated with higher estimated effects of mobility restrictions on social distancing. When including all countries in the same regression, the estimated coefficients for China and Italy remain significantly positive and larger in magnitude than those for other countries. Consistently across columns 1 through 7, we find that the interactions of mobility with social connections with China and Italy both retain their positive estimates and are in fact quantitatively unaffected when including social connections with these other countries. Our conclusion based on this analysis is that our main empirical results are not driven by openness or social connections acting as a proxy for altruism.

**Table S6**
**Robustness check: Social connectedness to informative vs. less informative countries**

The dependent variable is *Social distancing*. *SCI China* and *SCI Italy* are the Social Connectedness Index values between the county and China and Italy, respectively. *N cases* is the current number of confirmed cases in the county. Heteroscedasticity-consistent standard errors, double-clustered by county and day, are shown in parentheses.

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| Restr. x ln(SCI China) | 0.0011** | 0.0009 | 0.0008 | 0.0010* | 0.0010* | 0.0011* | 0.0011* |
| | (0.0005) | (0.0005) | (0.0005) | (0.0006) | (0.0006) | (0.0006) | (0.0006) |
| Restr. x ln(SCI Italy) | 0.0042*** | 0.0037*** | 0.0036*** | 0.0037*** | 0.0036*** | 0.0038*** | 0.0039*** |
| | (0.0005) | (0.0008) | (0.0008) | (0.0008) | (0.0008) | (0.0008) | (0.0009) |
| Restr. x ln(SCI Canada) | | 0.0008 | 0.0006 | 0.0009 | 0.0008 | 0.0009 | 0.0011 |
| | | (0.0009) | (0.0009) | (0.0009) | (0.0010) | (0.0010) | (0.0010) |
| Restr. x ln(SCI Mexico) | | | 0.0004 | 0.0005 | 0.0005 | 0.0005 | 0.0006* |
| | | | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0004) |
| Restr. x ln(SCI India) | | | | -0.0007 | -0.0007 | -0.0006 | -0.0005 |
| | | | | (0.0005) | (0.0005) | (0.0005) | (0.0006) |
| Restr. x ln(SCI Philippines) | | | | | 0.0002 | 0.0003 | 0.0004 |
| | | | | | (0.0006) | (0.0006) | (0.0006) |
| Restr. x ln(SCI Vietnam) | | | | | | -0.0005 | -0.0005 |
| | | | | | | (0.0004) | (0.0004) |
| Restr. x ln(SCI Brazil) | | | | | | | -0.0007 |
| | | | | | | | (0.0008) |
| Restrictions (0-5) | -0.0280*** | -0.0288*** | -0.0284*** | -0.0293*** | -0.0291*** | -0.0300*** | -0.0297*** |
| | (0.0042) | (0.0043) | (0.0043) | (0.0043) | (0.0044) | (0.0044) | (0.0044) |
| ln(N cases) | 0.0113*** | 0.0112*** | 0.0112*** | 0.0113*** | 0.0113*** | 0.0113*** | 0.0113*** |
| | (0.0009) | (0.0009) | (0.0009) | (0.0009) | (0.0009) | (0.0009) | (0.0009) |
| State-Day FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 184,766 | 184,766 | 184,766 | 184,766 | 184,766 | 184,766 | 184,766 |
| $R^2$ | 0.756 | 0.756 | 0.756 | 0.756 | 0.756 | 0.756 | 0.756 |

**Significance levels: * 0.1, ** 0.05, *** 0.01.**

## S4.5 Social connections within the United States

Table S7 studies whether social connections to New York affect the responses to mobility restrictions in the same way as social connections to China and Italy. To distinguish the information effect from social connections predicting travelling, we exclude all counties located in the state of New York, as well as the adjacent states, including Connecticut, Massachusetts, New Jersey, Pennsylvania, Rhode Island, and Vermont.

Table S8 studies the effect of COVID-19 cases in all socially connected counties in the U.S. To avoid the results being driven by physical proximity and a potentially higher infection risk, we repeat the analysis using different minimum distances between the county itself and the socially connected counties. Column 1 studies the interaction of mobility restrictions with the SCI-weighted number of cases in all socially connected counties, column 2 only excludes counties less than 100 miles away from the county, and column 3 those less than 200 miles away. In all cases, we find statistically significant results consistent with our main results: higher SCI-weighted cases, which we interpret as a proxy for the information flow from a county's social connections, increase the impact of mobility restrictions on social distancing.

**Table S7**
**Gov't restrictions and local cases vs. social connectedness**

The dependent variable is *Social distancing*. The analysis using *SCI New York* excludes all counties located in the state of New York, as well as the adjacent states (Connecticut, Massachusetts, New Jersey, Pennsylvania, Rhode Island, and Vermont). Heteroscedasticity-consistent standard errors, double-clustered by county and day, are shown in parentheses.

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Restr. x ln(SCI New York) | 0.0046*** | 0.0032*** | | |
| | (0.0004) | (0.0006) | | |
| Restr. x ln(SCI China) | | 0.0002 | | |
| | | (0.0006) | | |
| Restr. x ln(SCI Italy) | | 0.0016** | | |
| | | (0.0007) | | |
| Restrictions (0-5) | -0.0012 | -0.0106** | | |
| | (0.0036) | (0.0053) | | |
| ln(Cases) x ln(SCI New York) | | | 0.0007** | -0.0017 |
| | | | (0.0003) | (0.0012) |
| ln(Cases) x ln(SCI China) | | | | -0.0001 |
| | | | | (0.0015) |
| ln(Cases) x ln(SCI Italy) | | | | 0.0032** |
| | | | | (0.0016) |
| ln(N cases) | 0.0106*** | 0.0106*** | 0.0218*** | 0.0054 |
| | (0.0009) | (0.0009) | (0.0022) | (0.0072) |
| State-Day FE | Yes | Yes | Yes | Yes |
| County FE | Yes | Yes | Yes | Yes |
| N | 168,659 | 168,659 | 168,659 | 168,659 |
| $R^2$ | 0.745 | 0.745 | 0.738 | 0.738 |

**Significance levels: * 0.1, ** 0.05, *** 0.01.**

**Table S8**
**Gov't restrictions vs. SCI-weighted U.S. cases**

The dependent variable is *Social distancing*. Heteroscedasticity-consistent standard errors, double-clustered by county and day, are shown in parentheses.

| | (1) | (2) | (3) |
|---|---|---|---|
| Restr. x ln(SCI-w. US cases) | 0.0044*** | | |
| | (0.0011) | | |
| ln(SCI-w. cases - min dist. 0) | 0.0357*** | | |
| | (0.0035) | | |
| Restr. x ln(SCI-w. US cases) (100) | | 0.0075*** | |
| | | (0.0015) | |
| ln(SCI-w. cases - min dist. 100) | | 0.0307*** | |
| | | (0.0044) | |
| Restr. x ln(SCI-w. US cases) (200) | | | 0.0070*** |
| | | | (0.0018) |
| ln(SCI-w. cases - min dist. 200) | | | 0.0287*** |
| | | | (0.0054) |
| Restrictions (0-5) | -0.0102 | -0.0320*** | -0.0308*** |
| | (0.0061) | (0.0080) | (0.0095) |
| ln(N cases) | 0.0153*** | 0.0200*** | 0.0207*** |
| | (0.0011) | (0.0011) | (0.0010) |
| State-Day FE | Yes | Yes | Yes |
| County FE | Yes | Yes | Yes |
| N | 184,766 | 184,766 | 184,766 |
| $R^2$ | 0.755 | 0.753 | 0.752 |

**Significance levels: * 0.1, ** 0.05, *** 0.01.**

## S4.6 Accounting for types of work

For a number of reasons, mobility restrictions may affect households differently depending on their economic characteristics. Social connections may be correlated with these economic characteristics. First, the pandemic may have instigated large workforce adjustments. Given that our social distancing measure, $\frac{Completely\ Home_{i,t}}{Total\ Device\ Count_{i,t} - Working_{i,t}}$, is partially deflated by work, one might wonder if our measure of social distancing is not driven by changes in those staying at home but rather by adjustments in the workforce. It is unknown whther the exact timing of workforce adjustments coincide with mobility restrictions. To be clear, it could be important in our analysis only if the magnitude of these workforce adjustments could be related to international social connectedness. For example, those who are socially connected via social media abroad may achieve such ties through working in import/export or in a global service firm. As telecommunications occur on a regular basis anyway, remote work might be more possible in these firms.

Second, prior work suggests that different sub-groups may have economic constraints that prevent optimal health decisions [26] . For example, it may be that high *SCI* areas may be areas with more ability to socially distance merely because such households are better equipped to withstand economic shocks and not attend work.

We assess the extent to which workforce-related adjustments could be part of our analysis in two ways. First, we decompose our statistical measure into the fraction staying completely home versus the change in the percentage of devices classified as displaying work-like behavior. Second, we construct a battery of economic indicators to conduct split-sample analysis and construct alternative measures of social distancing that account for differences in the fraction of jobs in each county that may have been differently affected by workforce adjustments.

We first assess the magnitude of Table S9 presents our results. After restrictions come into force, socially connected counties exhibit a larger quantitative response in terms of staying at home than in changes in the percentage of devices working. Columns 1 and 2 present the estimates on *CompletelyHome*. The coefficient is positive. Columns 3 and 4 suggest that the relationship between the percentage of devices classifed as working and the coefficient of interest $\beta$ is negative. This provides us some evidence that workforce changes do not account for our results. It also suggests that given that we are subtracting out working people from the population, socially connected counties may have a larger *denominator*, suggesting that the working fraction has gone down in these counties. This might mean a larger base of people against which we rate social distancing, biasing against our finding.

**Table S9**
**Accounting for different types of work**

| | CompletelyHome | | Working | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| ln(SCI China) × Restr. | $0.004^{***}$ | | $-0.001^{***}$ | |
| | (0.0003) | | (0.0002) | |
| ln(SCI Italy) × Restr. | | $0.004^{***}$ | | $-0.001^{***}$ |
| | | (0.0003) | | (0.0002) |
| Restrictions | $-0.021^{***}$ | $-0.024^{***}$ | 0.001 | 0.002 |
| | (0.003) | (0.004) | (0.003) | (0.003) |
| ln(Cases) | $0.012^{***}$ | $0.012^{***}$ | $-0.007^{***}$ | $-0.006^{***}$ |
| | (0.001) | (0.001) | (0.001) | (0.001) |
| State-Day FE | Yes | Yes | Yes | Yes |
| County FE | Yes | Yes | Yes | Yes |
| N | 185,356 | 185,356 | 185,356 | 185,356 |
| $R^2$ | 0.845 | 0.846 | 0.833 | 0.833 |

### S4.6.1 Measuring workforce changes

The Covid-19 pandemic led to numerous changes in the workforce that could have impacted local counties differently in a manner correlated with social connectedness. We construct alternative measures of social distancing based on estimations of three different groups: the fraction of workers who became unemployed, the fraction of remote workers, and the fraction of workers in "critical" jobs.

First, unemployment may have increased. Mobility restrictions often prevent or discourage attendance at restaurants and other public venues. While heterogeneous across states or even cities within a state, such restrictions, coupled with economic uncertainty and pessimism, may have led to tremendous rises in local unemployment. We can account for March 2020 unemployment numbers in our analysis using data from Bureau of Labor Statistics (BLS). We note that the record unemployment numbers observed by the BLS was released in April 2020, occurring after our sample period. We estimate the fraction of the workforce that was previously employed based on the American Community Survey.

Second, immediately and over time some businesses became classified as critical, thereby exempting some industries from restrictions. We define the critical work population. We obtain a list from the Department of Homeland Security, which defines critical occupations by an Occupational Information Network (O*NET) code. Public tabulation of occupational breakdowns by county does not exist. Therefore, we take data from Linkup, which provides 200 million job postings since 2007. Linkup provides, for each job posting, the zip code where the work is meant to be done as well as the occupation code. Using Linkup occupational codes, we construct an estimate of the percentage of jobs in a county defined as critical using job postings from the year 2016 onward. Of course, we note that job postings reflect demand for workers, not headcount. Thus, differential turnover in a particular occupation could lead to skews in our results. Still, we argue that showing robustness to an albeit noisy measure of critical workers is better than not attempting to account for such workers at all.

Third, certain types of work shifted from the office to remote locations, presumably one's own home. In either of these cases, there may be shifts in stay-at-home behavior imposed by labor market conditions. It is not clear how one should account for these workers as employers have discretion in how they comply with social distancing. For example, famously, Elon Musk has tried to defy California state mandates to not operate his Tesla factory. Moreover, looking at crowdsourced data collection efforts such as Stayinghome.club, of 911 companies which reported working from home, 492 required it, and 381 encouraged it. Also, to the extent that remote work is mandatory and people cannot attend the office, it's not clear if workers are precluded from traveling to friends' homes or sitting in coffee shops. However, we collect four different data sources to account for this possibility.

We estimate remote work using two approaches. First, following [41], we estimate at the county-level the fraction of jobs that can be done remotely. [41] estimates the fraction of jobs by industry which are feasibly done remotely. Feasibility is "based on responses to two Occupational Information Network (O*NET) surveys covering "work context" and "generalized work activities." For example, if answers to those surveys reveal that an occupation requires daily "work outdoors" or that "operating vehicles, mechanized devices, or equipment" is very important to that occupation's performance, [they] determine that the occupation cannot be performed from home." Their data are available on GitHub.

To calculate industry employment shares, we use the Quarterly Workforce Indicators produced by the United States Census. Using their verbatim, the Quarterly Workforce Indicators (QWI) "provide local labor market statistics by industry, worker demographics, employer age and size. Unlike statistics tabulated from firm or person-level data, the QWI source data are unique job-level data that link workers to their employers." The data provide statistics on the universe of employment. We extract the Q2 2020 and calculate county-level industry average employment shares since 2014.

Second, we collect data from Aberdeen's Computer Intelligence Technology Database. Aberdeen collects data via telephone research interviews. In 2019, Aberdeen covered roughly 3.2 million establishments in the United States, covering about 85% of all establishments with employment size over 10 and the vast majority of US corporations. They provide (1) the number of employees at the establishment, and (2) the number of workers they classify as remote, and (3) the number of laptops, which we imagine to be a proxy for a job that can be done remotely. Of course, this measure is imperfect as an employee may have two laptops and such conditions are not sufficient or necessary for being able to work remotely. Thus, we propose two measures of being able to work remotely: the fraction of employees classified as mobile, and the fraction of employees with laptops. We aggregate this data to the county level.

Fourth, there was an increase in delivery work. Amazon, for example, claimed to want 100,000 additional workers to meet demand for delivered products. Delivery might also increase if service providers such as restaurant workers turn to delivering products. Given such workers have no common day-time location, our social distancing measure would not be appropriate. Luckily, the effects of delivery drivers on our statistics is not likely to be large. The Bureau of Labor Statistics estimates that 1,449,100 jobs were in the light-truck or delivery industries, while the size of the labor force was over 163 million people in 2018. Therefore, even if there was a large increase in delivery workers, the increase may not be sufficient to affect the aggregate statistics and introduce a large bias in our empirical results.

To mitigate the concerns that an increase in delivery work could confound our results, we gathered data on where Grubhub was operational. We choose food services because food services is the one type of service very likely to increase. Other types of delivery driving, such as Lyft or Uber, actually likely contracted over this period – with both Uber and Lyft announcing layoffs given a vast reduction in people taking public transit. Meanwhile, statistics suggest food delivery has increased due to restaurant closures. We chose Grubhub because Grubhub, by far, has the largest coverage in America, serving as a publicly listed firm operational for over 16 years, and 5.29 billion of market capitalization as of June 5, 2020. We grabbed all cities in which Grubhub was operational as of the beginning of June. One might wonder about a small look-ahead bias. Though, we do not find any evidence of Grubhub entering markets after the pandemic (we tried to look at a Wayback machine to get the data as of earlier this year or last, but this did not yield a sensible result). Given the list of cities in which Grubhub reports operations in, there are in fact only 728 counties in which Grubhub actually operates. Although Grubhub claims to operate in over 5,000 cities across the US, we discovered because many cities in the United States are in fact quite small, by some definitions there are 19,506 cities in the United States as of the 2019 according to the US Census. For example, Ithaca and Hanover are small cities although they are thought of to others as college towns. Therefore, one county can incorporate many cities. As Grubhub operates in 1/4th of cities, coverage of nearly 1/4th of counties (728 counties of the total of just over 3,200) is sensible.

#### S4.6.2 Results

in this section, we can see how our estimates will change based on toggling different assumptions for whether and how we account for economic characteristics of households, remote work, unemployment or critical workers. We summarize all of this analysis in Figure S1. In this figure, each data point represents a different empirical specification, with red points representing those using *ln(SCI China)* while blue points represent *ln(SCI Italy)*. The y-axis is the coefficient estimate and we present three standard error bars, which is very conservative. Shown at the bottom are indicator variables that describe our different empirical specifications.

First, we present empirical specifications around constrained and unconstrained groups. Relative to the full sample presented in column 1, we find no evidence that economic characteristics nullify our main result with positive and significant coefficients using either social connectedness to China or Italy for low-credit, high-credit, low-income, or high-income households. There is some evidence in column 2 that low-credit households are less likely to socially distance after the first restriction relative to high-income households, giving some credence to the theory about constrained household choices [26]. However, measuring constraints by income differences between constrained and unconstrained households is minimal. Finally, segmenting by areas where Grubhub delivers versus where it does not, we find a lower coefficient in areas where Grubhub delivers. However, the coefficients are similar, within 25% of each other. This suggests that socially connected households may be less constrained households, but the economic characteristics are not sufficiently large so as to account for our main finding.

Second, we present specifications perturbing our original measure of social distancing. This measure assumes those who are going to work are making a rational choice. Assuming some people work remotely or were laid off, it's arguable that what might appear to be an increase in social distancing could actually instead be people being sent home from work due to a workforce adjustment. On the other hand, people working remotely or laid off can choose not to remain home. For example, they can work elsewhere such as at a Starbucks or a visit a friend's home. Therefore, we consider both of these possibilities in different versions our alternative social distancing measures, incrementally incorporating different assumptions into our measure one at a time.

For each social distancing measure that factors in remote work, we present the statistic adjusting remote work by different measures of the fraction of jobs that can be done remotely. As mentioned previously, we have three measures: (1) the fraction of industry jobs that can be done remotely, per [41], (2) the fraction of workers in a county with laptops according to Aberdeen, (3) the fraction of workers in county classified as mobile by Aberdeen, and (4) the fraction of employees with laptop multiplied by 3 (with a max of 100%). This is meant to simulate a relatively extreme case where workers can adjust relatively rapidly.

We present results from 15 different perturbations of our social distancing measure, derived from seven different formulas. Across our different social distancing measures, the median correlation was 79%, suggesting the measures capture very similar information. Our point estimates do change but are statistically significant in all cases, and more often than not the point estimate increases, suggesting that some potential concerns about instantaneous operationalization of workforce change are unlikely to drive our results.

- Definition 1: Our original measure, $\frac{CompletelyHome}{1-\%working}$.
- Definition 2: $\frac{CompletelyHome}{1-\%workforce}$ - Nobody is fired, nobody is allowed to work remotely (yet).
- Definition 3: $\frac{CompletelyHome}{1-\%workforce+\Delta unemployed^{march2020}}$ - Here, we factor in those laid off or

furloughed by the end of March. Nobody is allowed to work remotely (yet).

- Definition 4: $\frac{CompletelyHome}{1-\%workforce*\%critical}$ - We can assume everyone who is not a critical worker optionally chooses to go to the office or be completely at home. Some non-critical workers can still not be at home.

- Definition 5: $\frac{CompletelyHome}{1-\%workforce-\%remote-\%critical}$ - Remote work is optional. Assumes remote workers are not critical. Non-critical, non-remote workers are laid off.

- Definition 6: $\frac{CompletelyHome-\%workforce*\%remote}{1-\%workforce}$ - No workers laid off, much less remote workers. However, we assume those who are remote working are forced to remain home, and not voluntarily engaging in social distancing. Therefore, they should be removed from the numerator.

- Definition 7: $\frac{CompletelyHome-\%workforce*\%remote}{1-\%workforce+\Delta unemployed^{march2020}}$ - The union of all changes. Should account for all non-critical workers who were not laid off. Unemployment rate changes should be exhaustive of all workers removed from the workforce.

- Definition 8: Same as Definition 7, except we factor in changes to the workforce only after the first restriction to test the senstivity of our finding to accounting for the timing of when these workforce changes were made.

In all specifications, we find statistically significant results. In fact, our point estimates actually increase when we account for remote work and workers being sent home. This is consistent with our previous finding that the fraction of people classified as working in counties socially connected to China and Italy is lower after restrictions.

**Figure S1:** Coefficient estimates using alternative economic assumptions