

Data appendix

State COVID-19 data

We obtain publicly released facility-level data on COVID-19 infections from 18 states. We began collecting this data in mid-April and continued to do so approximately every week until the week of July 10, 2020. Since some states do not report facilities with “closed outbreaks”—i.e. facilities with no current cases—we use the historical data to build a cumulative measure of whether a facility was ever infected as if they appeared on any list. These data should largely reflect all nursing homes that have ever reported a COVID-19 infection, though data is usually self-reported by facilities and may contain errors, and states also differ in the exact data that they report. Notably, Maryland only reported facilities that had cases after April 15, New York only reports deaths that occurred at the nursing home (rather than all deaths among nursing home residents), and four states only report nursing homes with 2+ cases. We matched the names on these lists to the administrative data on the universe of nursing homes. This allows us to calculate the number of deaths per bed at each nursing home, and to compare characteristics of nursing homes with low and high numbers of deaths. In terms of numbers of deaths, the 18 states in our sample represent over 80% of the total deaths from COVID at the time of data collection and contain all of the top 10 states. The states with many deaths for whom we do not have facility-level data include Texas, Ohio, Indiana, Arizona, Texas, and Virginia.

Eight of our sample states also released easily accessible data on confirmed or probable COVID-19 cases at a finer geography than county. These states and the lowest level of geography at which they supply data on cases were: Connecticut (town), Florida (town), Illinois (zip), Louisiana (tract), Massachusetts (town), Maryland (zip), Rhode Island (zip), and South Carolina (zip). For all states except Louisiana, we use the case rates as of the date the nursing home infection data was pulled. For Louisiana, we use data from May 31, because we have not been able to obtain data from the previous week.

Nursing home staff geography measures

To calculate our nursing home staff neighborhood characteristics, we first use the 2017 LODES Workplace Area Characteristics data to identify the nursing home's census block. This data records the number of workers in the education and health care sector who work on a given census block. We geocode the nursing home addresses, and compute a predicted number of nursing home workers based on the national ratio of workers to residents (1.6 million workers to 1.3 million residents), and the average number of residents for each facility from the Nursing Home Compare data. For 83% of facilities, health care employment on the coded census block is more than half the predicted employment, and we consider this a match. For the remaining 11% of facilities, we use the closest census block in the block group that meets this condition. This allows us to match an additional 11% of facilities, leaving about 511 facilities unmatched, which we exclude from our sample.

Using the census block chosen above, we then use the LODES Origin-Destination (OD) data to identify the home census blocks of workers who work on the same census block as the nursing home and belong to the “all other services” industry group. One concern is that there may be many more service sector employers that are on the same census block as the nursing home, and these employers have staff who live in completely different neighborhoods than the nursing home's staff. We find on the median block, 92% (IQR: [.62, 1]) of the service employment on these census blocks is in the education and health care sector. This gives us reassurance that we are not mostly picking up an entirely different type of employer on these blocks and that our geocoding is likely to be fairly accurate.

However, it is still possible that there are other education or health-care employers on the same block as the nursing home. One particular case of this is nursing homes located near hospitals. Indeed, we do find some blocks with unreasonably large numbers of health care workers for a nursing home (5000+). However, in general, we find that the total employment numbers are reasonable. Using the calculation for predicted nursing home employment above, the median block in our sample has an actual service sector employment to predicted employment ratio of 1.5 (IQR [1.1, 2.7]). Thus, we believe it is likely that the measured neighborhood characteristics will be largely representative of the types of neighborhoods where a facility's employees are likely to live.

To calculate the “share frontline” measure, we define a frontline worker as a worker in an essential industry (as defined in Tomer and Kane, 2020) in an occupation that cannot work from home (as defined in Dingel and Neiman, 2020).