Prioritizing Public Health Resources for COVID-19 Investigations: How Administrative Data Can Protect Vulnerable Populations

Mark Krass, Peter Henderson, Daniel E. Ho

Cite As: "Prioritizing Public Health Resources for COVID-19 Investigations: How Administrative Data Can Protect Vulnerable Populations," Health Affairs Blog, April 22, 2020. DOI: 10.1377/hblog20200420.729086

Introduction

In the face of the COVID-19 pandemic, scarcity is the order of the day.[1] Exponential growth may stress hospitals, testing laboratories, and medical supply chains to their limits.

One way of reducing this intense burden is by adopting public-health measures that can relieve pressure on the rest of the medical system.

Consider contact tracing, a measure central to South Korea's successful COVID-19 mitigation strategy.[2] Contact tracing is time and resource intensive. Interviewing every patient and getting in touch with their close contacts involves hours of work. In Wuhan, for instance, "more than 1800 teams of epidemiologists,[3] with a minimum of 5 people/team, are tracing tens of thousands of contacts a day." Understaffed public health departments are finding it tough to keep up, and some jurisdictions have abandoned contact tracing entirely.[4]

Allocation Beyond Contact Tracing

The same resource allocation problem plagues many other upstream public health interventions beyond contact tracing.

Public health officials require tools to help prioritize scarce investigative resources. One natural approach is to focus on cases most likely to infect the most vulnerable people. Imagine, for instance, the situation in King County, WA at the beginning of March. About a dozen people had received positive test results at that point in time. One of them lived at the Life Care Center (LCC) in Kirkland, Washington, a long-term care facility for frail and mostly older seniors. The top priority may be to deploy preventive mitigation measures to the LCC.

But there's a catch. When the cases number in the thousands, it's not so easy to sift out the LCC cases from the growing pile. Public health officials don't necessarily possess a list of buildings where high densities of vulnerable individuals live. And, in emergency situations like these, when public health staff are occupied with urgent operational tasks, there is no spare capacity to build such a list.

The Detroit Health Department reached out to our group at the Regulation, Evaluation, and Governance Lab (RegLab) at Stanford for help in triaging public health interventions. When we prototyped our approach around March 31, Detroit had some two dozen investigators combing through nearly 3,000 cases and prioritized investigations of cases at licensed nursing homes and elderly care facilities. Such investigations were meant to rapidly deploy first-order mitigation measures (e.g., when and how to isolate a resident or employee), not full-blown contact tracing. Yet Detroit, much to its credit, also suspected that the lists were incomplete. Where else should they be allocating the scarce resources? Was their existing approach optimal? Were they missing vulnerable populations?

Administrative Data to the Rescue

Where could one quickly find data sources to help prioritize public health investigations? Census data is one potential option, but microdata access to tabulate statistics at the building level is restricted. Another lies in CDC's Social Vulnerability Index,[5] which ranks census tracts based on statistics like poverty and high-density housing to inform emergency responses, but the index is a general measure not tied to known COVID-19 risk. The most advanced proposals for contact tracing have focused on using data from mobile devices[6] to automate contact tracing, which may be powerful, but also raise serious privacy and surveillance concerns.

Voter registration files present a less restrictive and more efficient alternative. Such files are publicly accessible, updated in close to real time, list individual voter residential addresses, and, most importantly for COVID-19, contain information about the critical risk factor of age. This data can rapidly generate a list of large buildings not only for Detroit, but nearly any metropolitan area.

Using Voter Registration Data: A Case Study

Once the voter file data is acquired, creating a preliminary list of large buildings with high densities of seniors is straightforward. First, to obtain a list of large buildings, we grouped the voter file by residential address and counted the number of registered voters at each address. Second, we computed two measures of the age distribution to help officials prioritize buildings by the vulnerability of residents: (a) the average age of registered voters residing at the address; (b) the number of registered voters above 65. Third, because the voter file contains information about each individual voter, we can develop cumulative risk scores at the building level, using CDC estimates[7] of hospitalization and fatality risk for confirmed cases of COVID-19 by age. [1] For each building, we used these to generate a weighted average risk score of severe outcomes for residents.

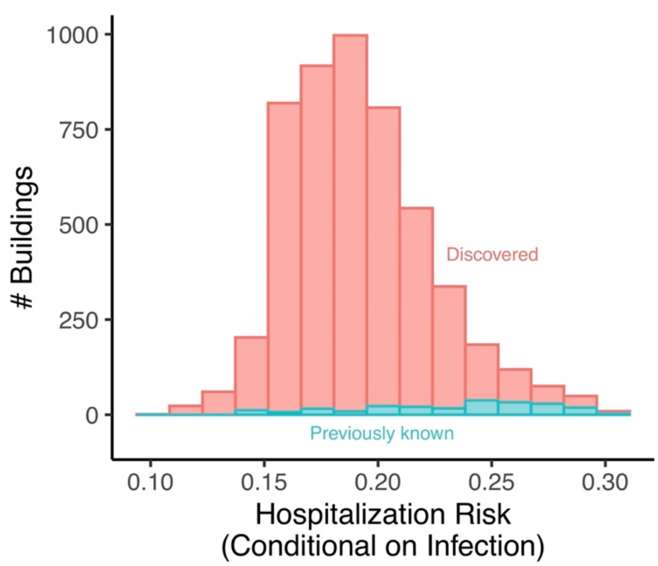
The approach can be scaled quickly and easily incorporated to help health departments prioritize case investigations more effectively. In Detroit, the additional list of facilities appears to generate at least three benefits. First, many facilities on Detroit's licensed list are small, housing only a half dozen residents. The voter file derived list uncovered elderly care communities with over a hundred residents above the age of 65. [2] Second, the licensed list may use the physical address of the management office, not the place where residents live. For some facilities, like dialysis centers, that makes sense: patients' permanent addresses should not be the focus. But for long-term residential communities, discrepancies between patient addresses and licensure data could prevent officials from homing in on early cases. Third, such administrative data allows for much more fine-grained prioritization calibrated to CDC estimates of risk.

To visualize the benefits of our approach, the exhibits below plot hospitalization risk, that is, the chance of being hospitalized conditional on infection, for "previously known" buildings based on licensing lists (turquoise color) and buildings residences "discovered" from voter registration rolls (salmon color).

Exhibit 1 indicates distribution of average residential risk across buildings. This panel shows that while Detroit's lists are indeed identifying some of the highest-risk buildings, its lists also have some buildings with lower risk and Detroit in fact has many more buildings at high risk. Roughly 90 of Detroit's licensed facilities have an average hospitalization risk of 25% or more. Voter registration rolls add 185 additional buildings above that hospitalization risk. The right panel plots the distribution of residents in six risk bins based on CDC estimates, instead of the average residential risk at the building. This panel illustrates how voter rolls help to uncover buildings that house a much larger proportion of vulnerable and elderly residents of Detroit.

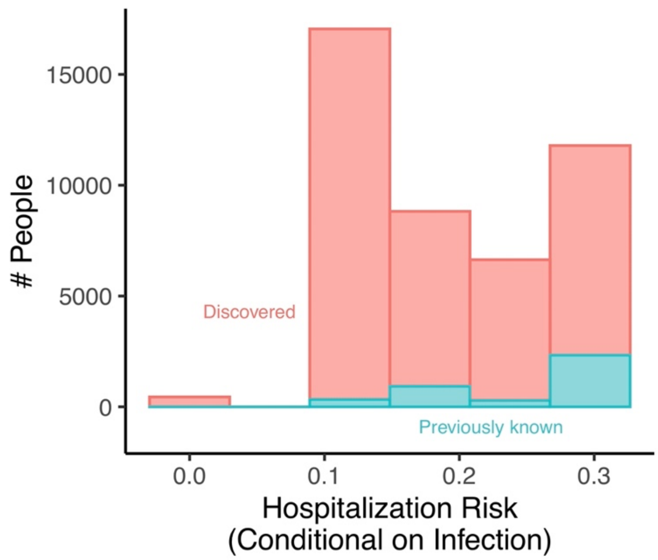
By using administrative data, Detroit might improve the effectiveness of its approach more than 5-fold, holding the investigative resources constant. [3]

Exhibit 1: Identifying high-risk buildings using voter files



Source: The underlying data sources were from the Detroit Health Department for previously known buildings and L2 voter files for Detroit. See technical notes below for additional detail.

Exhibit 2: Identifying high-risk individuals using voter files



Source: The underlying data sources were from the Detroit Health Department for previously known buildings and L2 voter files for Detroit. See technical notes below for additional detail.

Next Steps

We deployed this approach because it is fast, at a time when speed is required to compete with the spread of COVID-19, and could generate significant improvements in allocating scarce public health resources. Many areas of potential improvement and refinement remain.

First, one concern may be that relying on a list of registered voters would skew the prioritization demographically. Nationally, just 53.7% of eligible Hispanic adults reported[8] as registered to vote for the 2018 elections, while over 63% of potential White voters did. Non-citizens, who make up a growing share of the population in many urban areas, are not included in voter rolls. Yet this problem of demographic skew may be quantified and addressed. We can use Census tract data on demographics to adjust estimates of the vulnerable population, which may alleviate some of the skew. More importantly, the existing approach -- prioritizing licensed facilities -- is potentially more biased along two key dimensions: it focuses on smaller buildings and buildings that are *less* likely to be located in census blocks with immigrant communities. [4]

Second, our public infrastructure, including technology and law, remains ill-equipped to empower rapid deployment of these kinds of solutions. Lists of licensed facilities and associated meta-data are housed in decentralized and disparate fashion. Public health investigators can lack the technical capacity to ingest risk estimates. HIPAA and associated privacy concerns can slow down the sharing of critical data -- with cramped interpretations of its public health emergency exception -- when each day proves critical for mitigation. At the same time, COVID-19 has seen the academic and scientific worlds unite in unprecedented ways, and it will be critical to draw on these resources to help improve public health infrastructure.

In sum, our proof-of-concept shows the power of administrative data to augment the capacities of health departments. Ensuring fairness in allocation schemes is critically important, as is modernizing our public infrastructure. As more data becomes available, more sophisticated approaches based on machine learning and artificial intelligence can be deployed to protect the vulnerable.

Yet what the Detroit example shows most of all is that data-driven allocation of scarce public health resources will be critical in fighting this pandemic.

Technical Notes

(1) For age-adjusted risk estimates, we used the lower bound of CDC risk estimates within six age bins.

(2) To illustrate the magnitude of additional high-risk buildings discovered, of the 214 buildings found with >50 registered voters, only 38 were on the previously known list. Of 157 buildings with >25 estimated seniors, only 44 were previously known.

(3) Our estimate of the potential risk reduction is based on a simulated comparison of the current approach and the approach augmented with information from voter rolls. To simulate the current prioritization, we randomly sample 100 facilities from the list of known senior facilities. We assume that interventions reliably prevent infection for that building and that everyone in the building would be infected otherwise. If the hospitalization risk is accurate, the mean number of hospitalizations avoided by current interventions would be around 700. Under the same assumptions, ranking all known and discovered facilities by size and calculated hospitalization risk would result in a reduction of ~3900 hospitalizations.

(4) To illustrate the relative biases of the existing list and the augmented list, consider the following. According to data from the American Community Survey, the average building on Detroit's existing list belonged to a census tract in which 2.6% of people were foreign born. When we include the discovered buildings, the average building is in a tract with 3.7% foreign-born residents. This also holds true when we focus on known and discovered buildings with relatively high vulnerability to COVID-19, that is, in which the average risk of hospitalization is above 0.2. Vulnerable facilities on Detroit's existing list fell in census tracts where 2.6% of people were foreign-born, on average, while the average vulnerable building on the augmented list is in a tract with 3.0% foreign-born persons. Similarly, using data from CDC's Social Vulnerability Index, we found that the average new building falls in a tract where the percentage of persons with limited English proficiency was 1.22%, compared with 0.96% in the original list (0.97% vs. 0.92% for vulnerable buildings). Conversely, buildings on the new list fall in census tracts with fewer persons below the poverty line. The average high-risk building on the augmented list is in a census tract with 36.0% poverty, while the average high-risk building on the original list was in a tract with a 40.6% poverty rate (37.1% vs. 39.6% for all buildings). Such discrepancies can potentially be addressed by post-stratification.

References

1. Emanuel, Ezekiel J., Govind Persad, Ross Upshur, Beatriz Thome, Michael Parker, Aaron Glickman, Cathy Zhang, Connor Boyle, Maxwell Smith, and James P. Phillips. "Fair allocation of scarce medical resources in the time of Covid-19." (2020): 2049-2055.
2. Normile, Dennis. "Coronavirus cases have dropped sharply in South Korea. What's the secret to its success." *Science* 17, no. 03 (2020): 1-2.
3. Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19). https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf Date: Feb, 2020.
4. Gold, Jenny, Anna Maria Barry-Jester. "Testing Shortages Force Extreme Shift In Strategy By Local Health Officials." KHN. (2020). https://khn.org/news/testing-shortages-force-extreme-shift-in-strategy-by-local-health-officials/
5. Flanagan, Barry E., Edward W. Gregory, Elaine J. Hallisey, Janet L. Heitgerd, and Brian Lewis. "A social vulnerability index for disaster management." Journal of homeland security and emergency management 8, no. 1 (2011).
6. Servick, Kelly. "Cellphone tracking could help stem the spread of coronavirus. Is privacy the price?." *Science* (2020).
7. CDC COVID-19 Response Team, CDC COVID-19 Response Team, CDC COVID-19 Response Team, Stephanie Bialek, Ellen Boundy, Virginia Bowen, Nancy Chow et al. "Severe outcomes among patients with coronavirus disease 2019 (COVID-19)—United States, February 12–March 16, 2020." Morbidity and mortality weekly report 69, no. 12 (2020): 343-346.
8. U.S. Census Bureau. "Voting and Registration in the Election of November 2018." https://www.census.gov/data/tables/time-series/demo/voting-and-registration/p20-583.html