

INFORMING ON PUBLIC HEALTH EFFECTS AND DINING RISKS AS THE COVID-19 THREAT EVOLVES: A DATA-DRIVEN APPROACH

Saahil Sundaresan

Mountain View High School

Mountain View, California

{s67sea}@gmail.com

ABSTRACT

During the COVID-19 pandemic, consumers have faced significant additional health risks and challenges when engaging in out-of-home activities such as socialising and dining out. The present availability of a vast array of pandemic-related data, not only provided by the government and health authorities, but also generated through social media, presents significant opportunities in data science to conduct analysis and build tools to help citizens understand COVID-19 risks during this challenging period. In this paper, we provide an overview of some of our AI-related work in identifying COVID-19 risks by area, and risks of dining out. With the vaccination campaign underway and as more and more people get inoculated each day, our focus now shifts to studying how COVID-related risks evolve in response to the vaccinations, and the effects of new virus variants and vaccine hesitancy on those risks.

BACKGROUND AND MOTIVATION

The COVID-19 pandemic has created difficult and unexpected circumstances for billions of people across the globe. The countless decisions people have to make on a daily basis, in regards to school, work, eating out, holidays, and more, have only been complicated by the novel dangers of spread of the disease or news of virus mutation. Every in-person interaction with the outside world now carries an inherent risk to the health and safety of one's self and family, which must be carefully and objectively considered in order to minimise one's chance of contracting the virus. At the time of this writing, according to official data (CDC (2020)) the United States has more than 28 million documented COVID-19 cases, and the death toll stands at over 500,000. The virus has wreaked havoc in the economy, particularly affecting small businesses; over 100,000 businesses have closed, impacting employment significantly.

About a century ago, it took a private detective 4 months to track down Typhoid Mary, an asymptomatic pastry cook who was behind the typhoid outbreak in New York, but not before she had infected over 50 citizens. With the availability of massive data in the modern era, as well as the tools to process them, the speed and scale at which analysis can be conducted today is markedly different. A recent paper in *Nature* (Chang et al. (2020)), uses mobility movement data of 98 million people at various points of interest including as restaurants, places of worship and the workplace, to explain risk of COVID infection and spread. It notes that the economically challenged have less control over their mobility since they have to work to provide for their families, placing them at a higher risk,

In recent years, massive demography, health and other data has become easily accessible and downloadable, as provided by government agencies. In addition, social media sites such as Yelp, Twitter, Facebook, and Google have made vertical dimensional data available in a structured form through APIs. The data.gov website has over 500 datasets related to COVID-19 alone, providing significant opportunities for research on topics related to COVID. Advances in big data techniques in machine learning and data science, and the availability of cloud computing resources for research, has made processing such data and building inference algorithms based on the data relatively straightforward. ICLR workshop themes on public health and WWW data, as studied by Shaban-Nejad et al. (2014), has extensively discussed how intelligence is created by sophisticated informatics platforms that

collect and integrate data from multiple sources, and apply analytics to generate insights that will improve decision-making at individual and societal levels.

PUBLIC DATA AND COVID-19 RISKS

Our initial interest in using COVID-19 and other data to build tools for public awareness of infection risks began in spring 2020, when infections were just starting to pick up in the US. At a time when little to no information was accessible to the public, and before the widespread prevalence of online visualisation tools to track the virus, we foresaw the importance of developing a measure of localised risk, so that users could make travel and other decisions in an informed way. We built a service called RankMe (Sundaresan (2020))¹. RankMe uses county-level data collected by the US Census, NY Times, CDC, NIH and other organisations to score each county in the United States in the categories of transmission risk, county preparedness, and socio-economic protections offered to residents. This is done via an algorithm that takes into account over 15 factors, including population density, poverty rates, case data, age demographics, hospital beds and ventilators per capita, as well as relative change in light pollution levels compared to pre-pandemic levels. Data for each of these factors, compared with data from all counties in the US, is then used to generate categorised risk scores for each of the three major areas listed, which are in turn combined to produce final score for each county nationwide. This allows users to understand the risk levels in their area and plan activities accordingly, and also to compare risk levels across areas.

COVID-19, SOCIALIZING AND DINING OUT

The restaurant and takeout industry has in recent decades become ingrained in the American culture and experience. A recent study (Fourth 2020) found that 56% of Americans eat out at a restaurant, order takeout or have food delivered at least 2-3 times per week. Even during the pandemic, on any given day, 21% of Americans go to a restaurant for a meal (Crabtree (2020)). The food service and preparation industry is also the second-largest job sector in the United States, and serves as a vital backbone of the economy. In addition, with the lack of other entertainment options, visiting bars and restaurants with friends and family has been a way to maintain social connection and mental health.

Even before the pandemic, there have been numerous potential health risks associated with dining out, ranging from contamination of food resulting in food poisoning to potential contraction of various illnesses from fellow diners. Research has demonstrated the ability of restaurant health inspections and food safety ratings to significantly reduce cases of food poisoning and food-borne illnesses (Liu et al. (2020)). In New York City and Los Angeles, restaurants were required to visibly display given letter grades based on inspection results. Studies (Wong et al. (2015), Simon et al. (2005) and Jin & Leslie (2009)) have found marked improvement in restaurant hygiene and a drop in food-related hospitalisations following implementation of these directives.

According to Ellis (2020), these risks have only multiplied during the pandemic; as bars and restaurants a common place of public congregation in today's world, they pose a significant potential risk both as places where COVID-19 may be contracted, and also as hotspots that can contribute to community transmission. Additionally, restaurants are places of maskless interaction, since people have to remove their face coverings to eat, making it hard to adhere to the guidelines proposed by the CDC. Health authorities have issued various guidelines for social distancing, mask wearing, and increased use of hand-washing and hand sanitiser use to prevent spread of the virus, but the message from public health officials has oftentimes been contradictory or incomplete, and the measures prescribed are often not consistently implemented in establishments.

While the CDC recommends against unnecessary public interaction, 3 in 10 Americans recently surveyed (Crabtree (2020)) demonstrated interest in socialising and dining, if permitted, in full-capacity restaurants, and over half of those surveyed showed interest if the restaurants are 25% full. In a controlled study (Fisher et al. (2020)) across 10 states in the US, researchers found that adults who tested positive for COVID-19 were approximately twice as likely to have reported dining at a restaurant in the 14 days before falling ill than those who tested negative. Additionally, while masks

¹RankMe was a Global Finalist in the NASA COVID-19 Space App Challenge held in early summer of 2020

Name	Location	Yelp rating	RRISK score	Reason for shutdown
Blue Bar & Grill	Yonkers, NY	4.7	51 (Orange)	Disorderly patron, health violations
Fozzy's Grill	Rockford, IL	4.6	32(Deep Red)	Did not comply with dining rules
Dolphin Grill	Salisbury, MA	4.0	52(orange)	Rear entrance allowance
Republic Garden	Bethesda, MD	3.5	39(Red)	Not enforcing soc.distancing protocols
Plan B Bar	Rockville, MD	4.5	48 (Red)	Letting patrons play pool
Fire Station 1 Restaurant	Silver Spring, MD	4.6	44 (Red)	Overcrowding

Table 1: In this table we show some restaurants and their calculated risk scores. The restaurants were closed down for violations. Our system accurately predicted the risks of these establishments and warned potential customers against visiting, even though they had higher Yelp scores. In the color coding red is highest risk and orange is high risk. There were other restaurants and bars in the neighborhood that were permitted to remain open, to which our algorithm gave markedly higher ratings.

are often presented as a solution to COVID-19 risk, the data showed that wearing masks did not influence the risk in a significant way compared to being in close proximity through dining.

The risks of COVID-19 are even higher among minority communities. According to recent data from Marshall (2020) at the Mayo clinic, American Indian and Alaska Native people had an age-adjusted COVID-19 hospitalisation rate of about 5.3 times that of non-Hispanic white people. COVID-19 hospitalisation rates among Black and Hispanic or Latino individuals were both about 4.7 times the rate of non-Hispanic white people. Additionally, given that about 30% of food preparation workers in the United States are from minority communities, these risks are multiplied even further.

Given these concerns, and the fact that Americans remain heavily dependant on restaurants and takeout even throughout the pandemic, there is a significant need for a service that can provide an unbiased and accurate model of COVID-19 contraction risk for restaurants and other food establishments.

USING PUBLIC DATA TO IDENTIFY DINING RISKS

In our paper (Sundaresan et al. (2021)), we describe a machine learning-based system that takes publicly available data, some provided by the government, some curated from media, and others available via social media and consumer forums like Yelp, to automatically calculate risk of COVID-19 for individual food establishments in the US and communicate that information to the public. Our data, spanning 7 dimensions, scales to millions of rows of information, and is dynamic and updated daily. For each business establishment, risk calculation is based on factors. Having multiple dimensions of features allows us to identify risks at different hierarchies, for example "Juice bars" versus "Juice bars at a certain location". Factors considered in our algorithm include local population densities and case counts, as well as restaurant reviews, written user sentiments, health inspection data and food type. The details of the algorithm and the system are in Sundaresan et al. (2021). The system is able to measure at least substantial risk for approximately 90% of restaurants that were later closed down, and flag them as high risk in 70% of cases. Some examples of such restaurants are shown in Table .

The system was also useful in identifying two interesting phenomena of restaurant risks: establishments of relatively low risk in comparatively high-risk areas (Figure 1), and variations in risk even in geographically and demographically similar establishments (Figure 2).

VACCINES AND DINING

As we have moved to the next phase of the COVID-19 pandemic, vaccines are rapidly becoming available. The United States began its mass vaccination campaign in mid-December, and approximately 65 million doses have been administered since then. A lack of complete data transparency makes it difficult to scrutinise the demographics of those already vaccinated, but with a current administration rate of around 1.5 million doses per day, it is reasonable to presume vaccination of the majority of the adult population of the United States by the end of 2021. Given this, and as rates of

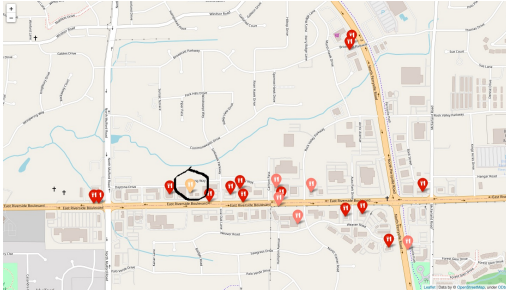


Figure 1: A map of E Riverside Blvd, IL 61111. Almost all restaurants except one are high to very high risk according to RRISK. Eating out in this area is risky except for one restaurant that offers lowered risk.

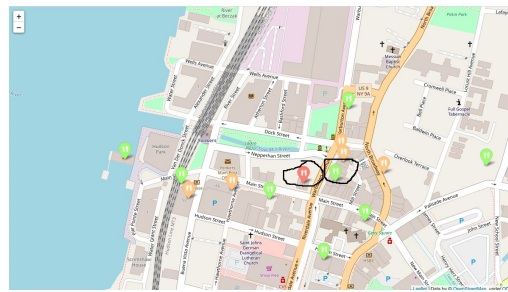


Figure 2: Area around Warburton Avenue, Yonkers, NY 10701. Two restaurants, La Bella Havana and Guapo,, both serve the same type of food (Cuban cuisine) and are located in close proximity to one another, yet have starkly different risk assessments.

reported infections and hospitalisations continue to fall, the end of the pandemic seems to be fast-arriving for many. After a year of pandemic-era lockdowns, this hope, as well as COVID fatigue, has left people anxious for a return to normal. The public is beginning to push for the reopening of schools, gyms, concert halls and places of worship, and many are becoming more relaxed in their personal caution regarding the virus.

Despite all the positive developments, it is important to note that substantial concerns still remain, given the relatively small proportion of the population currently vaccinated, and the recent emergence of new viral variants, which have been found to be more resistant to the effects of the vaccine, leaving its true practical effectiveness unknown. The question thus arises: what procedures should food establishments follow to continue to ensure a safe dining experience? what risks should potential patrons be aware of?

As some states start releasing vaccination data with detailed county-by-county statistics, we have begun looking to analyse the effects of vaccinations on calculated restaurant risks. While the results are still preliminary, we are beginning to see a positive shift in estimated COVID-19 risk of restaurants in those areas with a higher proportion of residents vaccinated, including in parts of California such as Mono County, and other areas of the country including New Mexico and South Dakota. This needs to be considered with caution, however; it is still too early to tell how much, and how soon, of an effect the vaccine will have on dining risk on a macro scale.

We do, however, make some observations regarding protocols that should be observed by food establishments. Firstly, as more of the population gets vaccinated, and restrictions start to lessen, there still remains a significant potential for those who have not been vaccinated to transmit the virus and spread it, and for those who are vaccinated to unknowingly carry it as well. To mitigate this concern, we recommend that food establishments require proof of vaccination from patrons who wish to dine in. Additionally, with the vaccination campaign in as early a stage as it is in right now, public health experts generally advise against frequent in-person food consumption, movie theater visits and other higher-risk activities, and the public should continue to exercise due vigilance and caution regarding the virus. Restaurant owners should also continue to follow all COVID-19-related safety and hygiene protocols as stringently as possible. As new data becomes available we will continue to study the effects of the vaccine, and of relaxation of social gathering and dining restrictions, and build appropriate tools or alerts in our system to let citizens take informed action.

CONCLUSIONS AND FUTURE WORK

In this paper we discussed the use of publicly available datasets to assess COVID-19 risks and risks of dining out, and stressed the importance of such work to the public during the pandemic. We believe that with such analytical tools and systems, we can help people make safe decisions for themselves and their families. Our work is still evolving as new data like vaccine related data becomes available, and so we have designed a flexible framework that allows us not only to incorporate user responses back into the system and account for additional factors including vaccinations and

spread of virus variants, but also to expand its scope beyond dining to other uses including gyms or movie theaters. We also hope to in the near future provide insights and feedback to the businesses in order to help them improve the safety of their establishments. Finally, we plan to expand our framework beyond the United States; as coronavirus is a global issue, including data from other countries will help significantly in humanity's fight against COVID-19.

ACKNOWLEDGEMENTS

The authors would like to thank NASA, as well as the University of California, San Diego and the Border Solutions Alliance for the opportunity to work on this project.

REFERENCES

- CDC. Corona virius(covid-19) update, November 2020.
<https://www.cdc.gov/coronavirus/2019-ncov/index.html>.
- S. Chang, E. Pierson, W Pang, G Jaline, B Redbird, D Grusky, and J Leskovec. Mobility network models of covid-19 explain inequities and inform reopening. *Nature*, November 2020.
<https://doi.org/10.1038/s41586-020-2923-3>.
- Steve Crabtree. Which americans are most likely to be eating out again?, 2020. <https://news.gallup.com/poll/316622/americans-likely-eating-again.aspx>.
- R Ellis. Dining out linked to increased covid risks, September 2020.
<https://www.webmd.com/lung/news/20200911/dining-out-linked-to-increased-covid-19-risk>.
- K Fisher, M Tenforde, L Feldstein, C Lindsell, N Shapiro, C Files, K Gibbs, H Erickson, M Prekker, J Steingrub, M Exline, D Henning, J Wilson, S Brown, I Peltan, T Rice, D Hager, A Ginde, K Talbot, J Casey, C Grijalva, B Flannery, and M Patel. Community and close contact exposures associated with covid-19 among symptomatic adults over 18 years in 11 outpatient health care facilities — united states, july 2020. *Morbidity and Mortality Weekly Report*, 69(36):1258–1264, September 2020.
- Fourth. Truth about dining out. <https://tinyurl.com/y5v3mjhn>.
- Ginger Jin and Phillip Leslie. Incentives for restaurant hygiene. *American Economic Journal (MicroEconomics)*, 1:237–267, 1 2009.
- J Liu, C Rehm, R Micha, and Mozaffarian. D. Quality of meals consumed by us adults at full-service and fast-food restaurants, 2003–2016: Persistent low quality and widening disparities. *The Journal of Nutrition*, 120(4):873–883, April 2020.
- W Marshall. Coronavirus infection by race: What's behind the health disparities, August 2020.
<https://www.mayoclinic.org/diseases-conditions/coronavirus/expert-answers/coronavirus-infection-by-race/faq-20488802>.
- Arash Shaban-Nejad, David Buckeridge, and John Brownstein. *World Wide Web and Public Health Intelligence*, volume Technical report WS-14-14. 2014. ISBN 978-1-57735-675-2.
- Paul Simon, Phillip Leslie, Grace Run, Ginger Jin, Roshan Reporter, Arturo Aguirre, and Jonathan Fielding. Impact of restaurant hygiene grade cards on foodborne-disease hospitalizations in los angeles county. *Journal of environmental health*, 67:32–6, 56; quiz 59, 03 2005.
- Saahil Sundaresan. Rankme, 2020. <https://covid19.spaceappschallenge.org/challenges/covid-challenges/human-factors/teams/rankme/project>.
- Saahil Sundaresan, Shafin Khan, Faraz Rahman, and Chris Huang. Risk: Analysing covid-19 risks in food establishments. In *Studies in Computational Intelligence*. 2021. Accepted manuscript: <https://tinyurl.com/t8rtndxj>.
- Melissa Wong, Wendy McKelvey, Kazuhiko Ito, Bryan Jacobson, and Daniel Kass. Impact of a letter-grade program on restaurant sanitary conditions and diner behavior in new york city. *American Journal of Public Health*, 105:81–87, 3 2015. doi: 10.2105/AJPH.2014.302404.
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4330857/>.