Tracking COVID-19 in Real Time: Leveraging Publicly-Available Data Sources to Inform Hospital Operations and Infection Prevention Practices

Kevin P O'Callaghan MB BCh[1,2,3] , Kyle Winser BSc[4], Vaidehi Mehta MS MPH[4], Eimear Kitt MB BCh[1,2,3], Coralee Del Valle Mojica MD[1,2,3], Caryn Kerman MD[1,5], Susan Coffin MD MPH[1,2,3], Julia S Sammons MD MSCE[1,2,3]

[1]Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, [2]Division of Infectious Diseases, Children's Hospital of Philadelphia, [3]Department of Infection Prevention & Control, Children's Hospital of Philadelphia, [4]Center for Healthcare Quality and Analytics, Children's Hospital of Philadelphia, [5]Division of General Pediatrics, Subsection of Hospital Medicine, Children's Hospital of Philadelphia

Background

- The SARS-CoV-2 pandemic led to an urgent need for data regarding national, regional, and local prevalence to inform public health practices.
- Hospitals and healthcare systems additionally found themselves equally in need of up-to-date data in order to develop best practices around patient screening, isolation and use of personal protective equipment.
- We describe a data analytics tool developed at our institution which uses publicly-available data sources to track county-level disease prevalence of COVID-19, in order to delineate risk for individual patients based on their geographic origin.

Objectives

- To develop a data analytics tool using publicly-available data sources to track county-level disease prevalence of COVID-19 in near real-time
- To deploy this tool as part of a hospital-wide screening program identifying patients from high-risk regions

Methods

- We investigated a variety of public data sources tracking COVID-19 case counts, assessing for (1) frequency of updates, (2) a granular level of geographic detail, optimally to a zip-code or county level, and (3) completeness of the data.
- We selected the Johns Hopkins University Center for Systems Science and Engineering COVID-19 dashboard data set, which is publicly available in a comma-separated value (CSV) file format.
- The data set incorporates counts of new diagnoses per day, per county, using Federal Information Processing System (FIPS) codes to delineate each county. The dataset is updated daily, with adjustments made for backdated data correction as needed.
- We developed an internal data analytics dashboard which allowed for direct comparison of calculated period prevalence by county. We developed a metric of aggregated 10-day rolling period prevalence, calculated as a total case count from the preceding 10 days, divided by county-level population from 2018 American Community Survey (ACS) estimates.

Fig 1. Period prevalence in Philadelphia County

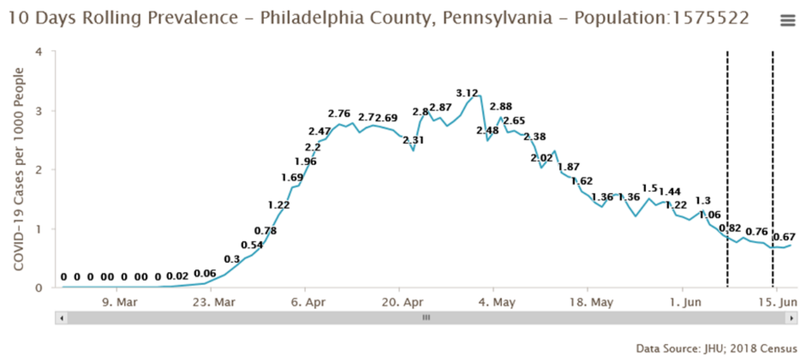


Fig 2. Period prevalence in Westchester County, NY

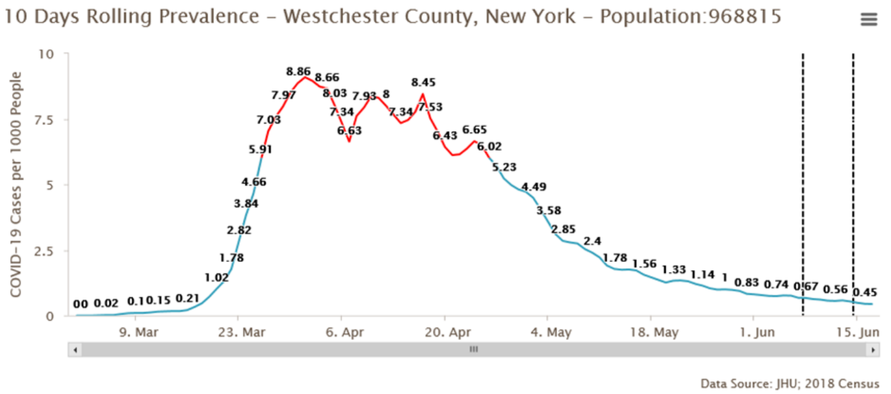
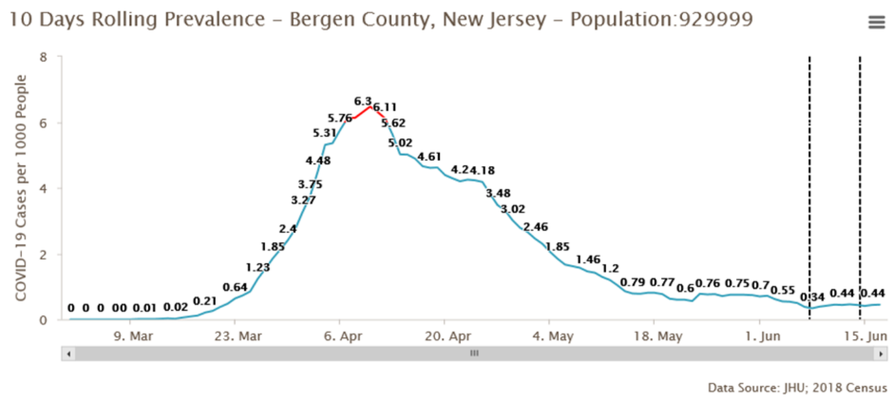


Fig 3. Period prevalence in Bergen County, NJ



Results

- Benchmarking against our local prevalence (peaking at 3.12 cases per 1,000 persons on a 10-day period prevalence), as well as information on peak prevalences in our region, we used 6 cases per 1,000 persons population over a 10-day cumulative period as the threshold for describing a county as a Geographic Region with Widespread Community Transmission (**GReWCoT**).
- Such counties would have to reach this threshold for at least 4 out of a total 7 days, starting 3 days prior to the assessment, extending to 10 days prior.
- 4 out of 7 days were required to bolster against intermittent short increases in positive test results, possibly reflective of bulking of test reporting. Likewise, the assessment period started 3 days prior to the index day, to prevent against delayed reporting of cases leading to inaccurately low prevalence assessments.
- Based on this benchmarking, our initial assessment identified a number of counties in the New York City (NYC) metropolitan area which had sustained period prevalences which were elevated above our threshold, consistent with contemporaneous guidance from our city Department of Public Health which had advised that persons coming to our area from the NYC metro area should self-isolate for 14 days.
- We continue to monitor period prevalences as collated by our data analytics tool, and leverage this in a biweekly review of geographic regions of concern, updated across our enterprise.
- Using the above definition, we have made specific infection prevention recommendations for patients/families from those regions attending our facilities. These include use of modified enhanced precautions (including surgical mask/eye protection), as well as restricted visitation of caregivers and family.

Conclusions

- This approach provided a nuanced investigation of COVID-19 prevalence in real-time, and allowed for a differential risk stratification for patients and families attending our facilities from a large catchment area.
- Our data analytics tool is shared on an inward-facing website accessible to staff, and thus supports messaging around the needs for additional screening and support for families coming from regions with high prevalence.