Estimated Fraction of Incidental COVID Hospitalizations in a Cohort of 250 High-Volume Hospitals Located in 164 Counties

Jeffrey E. Harris MD PhD*† January 22, 2022

Abstract: Scattered reports have suggested that as many as one-half of all hospital inpatients identified as COVID-positive are incidental cases who were admitted primarily for reasons other than their viral infections. To date, however, there are no systemic studies of a representative panel of hospitals based on pre-established criteria for determining whether an individual patient was in fact admitted as a result of the disease. To fill this gap, we developed a formula to estimate the fraction of incidental COVID hospitalizations that relies upon measurable, population-based parameters. We applied this approach to a longitudinal panel of 164 counties throughout the United States, covering a 4-week interval ending in the first week of January 2022. Within this panel, we estimated that COVID incidence has been rising exponentially at a rate of 9.34% per day (95% CI, 8.93-9.87). Assuming that only one-quarter of all recent Omicron infections have been reported by public authorities, we further estimated the aggregate prevalence of active SARS-CoV-2 infection during the first week of January to be 4.89%. During the same week, among 250 high-COVID-volume hospitals within our 164-county panel, an estimated 1 in 4 inpatients was COVID-positive. Among such COVID-positive hospitalized patients, 15.2% were estimated to be incidental infections. Across individual counties, the median fraction of incidental COVID hospitalizations was 13.7%, with an interquartile range of 9.5 to 18.4%. Incidental COVID infections appear to be a nontrivial fraction of all COVID-positive hospitalized patients. In the aggregate, however, the burden of patients admitted for complications of their viral infections appears to be far greater.

* Professor of Economics, Emeritus, Massachusetts Institute of Technology, Cambridge MA 02139; Physician, Eisner Health, Los Angeles CA 90015. Email: *jeffrey@mit.edu*

† This study relies exclusively on publicly available data that contain no individual identifiers. The author has no competing interests and no funding sources to declare. This article represents the sole opinion of its author and does not necessarily represent the opinions of the Massachusetts Institute of Technology, Eisner Health, or any other organization. Supporting programs and data have been posted at https://osf.io/7umje/.

Introduction

Scattered, anecdotal reports have suggested that up to one-half of all hospital inpatients identified as COVID-positive are incidental cases who have been admitted primarily for reasons other than their viral infections [1-3]. To date, however, there have been no broad-based, systemic studies of a representative panel of hospitals to determine what fraction of COVID-positive hospitalizations are indeed incidental. The ideal design would be to prospectively follow a cohort of individuals, evaluating each hospitalization according to pre-established criteria for determining whether or not each patient was in fact admitted as a result of the disease. Such formal criteria would not leave it to arbitrary, subjective judgment how to assess, say, a young adult with COVID-associated headache, fever, rigors, body aches, dizziness, and severe fatigue, who suffered acute brain injury while driving under the influence of Omicron.

To fill this gap, we developed a formula to estimate the fraction of incidental COVID hospitalizations that relies upon objectively measurable, population-based parameters. We applied this approach to a longitudinal panel of 164 counties throughout the United States, which contained 250 high-COVID-volume hospitals. Our analysis covered the 4-week interval ending in the first week of January 2022, during which time the Omicron variant of SARS-CoV-2 was far and away the dominant strain [4]. While this exercise requires us to make some difficult-to-verify assumptions and involves a significant attendant range of uncertainty, it still educates us as to whether there is any reasonable chance that the incidental COVID fraction is as high as a handful of sources have reported.

Methods and Data

Defining the Fraction of Incidental COVID Hospitalizations, π

We develop a model that permits us to formally define the fraction of incidental COVID hospitalizations and to express this fraction as a function of other, observable quantities.

To that end, consider a closed population at a particular point in time. We will refer to any individual within this population who has a detectable SARS-CoV-2 infection at that time as a *COVID case*. This definition does not require that the individual has in fact tested positive or even has symptoms of COVID.

Let q (where $1 > q > 0$) denote the proportion of all COVID cases that are hospitalized *because* of their illness. For shorthand, will refer to these as the *severe* cases. We refer to all other COVID cases, whether they are hospitalized or not, as *non-severe*.

While the severe cases are, by definition, hospitalized because of their COVID illness, some of the remaining non-severe cases will also be hospitalized for non-COVID reasons. We refer to the latter group as *incidental COVID hospitalizations*. Let g denote the proportion of non-severe cases that is hospitalized, where $1 > g > 0$. Then, by Bayes' rule, the proportion of all COVID hospitalizations that are incidental is:

$$\pi = \frac{g(1-q)}{g(1-q)+q} \qquad (1).$$

We refer to π as the *fraction of incidental COVID hospitalizations*. We see that π is increasing in g and decreasing in q.

The problem with the formulation of equation (1) is that it doesn't render π as a function of readily observable quantities. If we could longitudinally follow a closed cohort of individuals as they contracted COVID and then observe who got hospitalized because of their illness, we would have q. To get g as well, we would also need to track who got hospitalized for other reasons. In the absence of such a comprehensive longitudinal study, it is hardly obvious how to proceed.

Rendering π as a Function of Observables

One way to proceed is to explore the consequences of a key simplifying assumption, namely, that those individuals without COVID have the same probability g of hospitalization as those with non-severe COVID. We examine the validity of this assumption later in the Discussion.

Let p denote the proportion of individuals in the entire population who are COVID cases, where $1 > p > 0$. We refer to p as the *prevalence of COVID*. The proportion without COVID is thus $1-p$. Under our key simplifying assumption, the proportion of individuals in the population who do not have COVID *and* are hospitalized is thus $g(1-p)$. We can also express the proportion who have non-severe COVID *and* are incidentally hospitalized as $g(1-q)p$, while the proportion who have severe COVID *and* are hospitalized as qp.

Now focus solely on those individuals who are hospitalized. We test every hospitalized patient to determine who is COVID-positive and who is COVID-negative. Let c denote the fraction of all hospitalized individuals who are COVID-positive. This quantity differs from the prevalence p, which represents the fraction of all individuals in the entire population who are COVID-positive. Utilizing the foregoing expressions, we have $c = \frac{g(1-q)p+qp}{g(1-p)+(1-q)p+qp}$. We can write the odds that a hospitalized patient is COVID-positive as $\frac{c}{1-c} = \frac{g(1-q)p+qp}{g(1-p)}$, which can be rewritten as $\left(\frac{p(1-q)}{1-p}\right)\left(\frac{g(1-q)+q}{g\,(1-q)}\right)$. Utilizing the definition of π in equation (1) and rearranging terms gives:

$$\pi = (1-q)\left(\frac{p}{1-p}\right)\Big/\left(\frac{c}{1-c}\right) \qquad (2).$$

In equation (2), π is now a function of the prevalence p of COVID in the population, the fraction c of all hospitalized patients who are COVID-positive, and the proportion q of COVID-positive individuals requiring hospitalization because of their disease. If we can obtain reasonably accurate estimates of these three parameters, we can in turn estimate π. Table 1 below summarizes these points.

Table 1. Component Parameters in the Derivation of the Fraction of Incidental Hospitalizations, π

Symbol	*Definition*
p	Prevalence of COVID
c	Fraction of all hospitalized patients who are COVID-positive
q	Proportion of all infected individuals who are hospitalized because of their COVID illness
π	Fraction of incidental hospitalizations, $\pi = (1-q)\left(\frac{p}{1-p}\right)\Big/\left(\frac{c}{1-c}\right)$

Rendering the Prevalence of COVID, p, as a Function of Observables

We focus now on the estimation of the prevalence p of COVID, deferring for now the estimation of the remaining quantities c and q. To estimate the prevalence p, we would ordinarily rely upon the classic formula in epidemiology, that is, prevalence equals the incidence of SARS-CoV-2 infection per unit time multiplied by the average duration of infection. The

difficulty with this formula is that it applies only to a population with a stable incidence rate, and that is certainly not the case here [5].

Let $h(t)$ denote the incidence rate of SARS-CoV-2 infection at time $t \geq 0$. We assume that incidence is growing exponentially, that is, $h(t) = h_0 e^{\rho t}$, where $h_0, \rho > 0$. Let the duration $s > 0$ of infection have an exponential distribution with mean $1/\theta$, so that the cumulative distribution function of s is $F(s) = 1 - e^{-\theta s}$ and the corresponding survival function is $G(s) = 1 - F(s) = e^{-\theta s}$.

Given these assumptions, the prevalence of COVID at time t is the integral $p(t) = \int_0^t h(u)G(t-u)\,du$. Inserting the assumed functional forms for h and G, we get: $p(t) = \frac{h_0}{\rho+\theta}\left(e^{\rho t} - e^{-\theta t}\right)$. For sufficiently large t, the second term inside the parentheses gets small, so that we have:

$$p(t) \approx \frac{1}{\rho + \theta} h(t) \qquad (3).$$

Note that when the incidence h is stable (that is, $\rho = 0$), this formula collapses to the classic result. To estimate the prevalence p during an epidemic with exponentially growing incidence, we therefore need data on the incidence h, the growth rate ρ of infection and the mean duration $1/\theta$ of infection.

Determination of the actual incidence h of COVID is no trivial task, as it is widely acknowledged that cases have been and continue to be significantly underreported [6-9]. To address this difficulty, we follow the usual approach of incorporating an under-ascertainment factor into our analysis [10]. Let $r(t)$ denote the *reported* incidence of COVID, and let $f(t) > 0$ denote the fraction of COVID cases that are reported at time t. Then we have:

$$h(t) = r(t)/f(t) \qquad (4).$$

Accordingly, to estimate p, we need data on the parameters r, ρ, f, h, and θ, as indicated in Table 2. For clarity, we have dropped the time argument t from the functions h, r, and f listed in the table.

Table 2. Component Parameters in the Derivation of COVID Prevalence, p

Symbol	*Definition*
r	Reported COVID incidence
ρ	Exponential rate of increase of COVID incidence
f	Fraction of COVID cases reported
h	Actual COVID incidence, $h = \frac{r}{f}$
θ	Inverse of the mean duration of SARS-CoV-2 infection
p	Prevalence of COVID, $p \approx \frac{1}{\rho+\theta} h$

Data: Cohort of 250 High-COVID-Volume Hospitals

From a database of U.S. hospitals maintained by the Department of Health and Human services (HHS) [11], we identified the 250 hospitals with the highest cumulative volume of emergency department visits for COVID from the week ending June 25, 2021 through the week ending December 10, 2021. These high-volume hospitals were located in 164 counties in 41 states and territories. The map of the continental U.S. in Appendix Fig. A1 displays the locations of all but one of the hospitals.

The hospital database maintained by HHS reported data on a weekly basis. For each of the 250 high-volume hospitals, we extracted the following three variables from the week ending June 25, 2021 through the week ending January 7, 2022.

(1) *inpatient_beds_used_7_day_avg*, defined as "Average of total number of staffed inpatient beds that are occupied reported during the 7-day period."

(2) *total_adult_patients_hospitalized_confirmed_covid_7_day_avg*, defined as "Average number of patients currently hospitalized in an adult inpatient bed who have laboratory-confirmed COVID-19, including those in observation beds. This average includes patients who have both laboratory-confirmed COVID-19 and laboratory-confirmed influenza."

(3) *total_pediatric_patients_hospitalized_confirmed_and_suspected_covid_7_day_avg*, defined as "Average number of patients currently hospitalized in a pediatric inpatient bed, including NICU, PICU, newborn, and nursery, who are suspected or laboratory-confirmed-positive for COVID-19. This average includes those in observation beds reported in the 7-day period."

Estimating the Fraction of Hospitalized Patients Who Are COVID-Positive, c

For each hospital and each week, we calculated the number of COVID-positive patients as the sum of variables (2) and (3), while variable (1) gives the total number of hospital patients. For each week, summing over all 250 hospitals, we computed cohort-wide numbers of COVID-positive inpatients and total inpatients, from which we then computed c, the fraction of all inpatients who were COVID-positive. While we display the entire timeline of the fraction c in Fig. 1 below, we relied on the most recently available value of c for the week ending January 7, 2022 in our calculations of the fraction of incidental hospitalizations π.

Data: Confirmed COVID Incidence in 164 Counties

For each of the 164 counties covering our cohort of 250 high-volume hospitals, we downloaded weekly reported COVID incidence from the *Counties* tab of Excel spreadsheets regularly issued by the White House COVID-19 Team as Community Profile Reports [12]. We extracted three variables from each report: *Population*, *Cases - last 7 days*, and *Cases per 100k - last 7 days*. These variables were extracted from the reports of December 20 (covering the week from December 13-19), December 27 (covering December 20-26), January 3 (covering December 27 – January 2), and January 10 (covering January 3-9). The resulting database contained a balanced panel of 4 serial observations on each of 164 counties. We show a whiskers-on-box plot of the variable *Cases per 100k - last 7 days* in Fig. 2 below.

Estimating the Exponential Rate of Increase of COVID Incidence, ρ, and Reported COVID Incidence, r_t

We relied upon the county-specific data derived from the Community Profile Reports [12] to estimate the exponential rate of increase of COVID incidence, ρ. To that end, we ran the following log-linear fixed-effects regression model on our panel of 4 serial weekly observations (indexed t = 1,…4) on 164 counties (indexed i = 1,…,164):

$$\log y_{it} = \mu + \rho w_t + \eta_i + \varepsilon_{it} \qquad (5).$$

In equation (5), the observations y_{it} corresponded to the data variable *Cases - last 7 days*, while the observations w_t represented the ending date of each of the four weeks. The parameter μ was an overall constant term, the parameters η_i were county-specific fixed effects, and the terms ε_{it} were spherical errors. The parameter ρ was estimated by ordinary least squares.

To test whether the data adequately fit an exponential growth model, we plotted the reported incidence r_t for all 164 counties combined against w_t to check for serial correlation of residuals. The reported incidence was computed as $r_t = \sum_i n_i m_{it} / \sum_i n_i$, where m_{it} denotes *Cases per 100k - last 7 days* and n_i denotes *Population*. We show a plot of r_t versus w_t in Fig. 3 below. While we used a continuous-time model to develop our formula for prevalence in equation (3), here we use the discrete time notation to refer to computations made from the panel of 164 observations in each of 4 successive weeks.

Estimating the Fraction of COVID Cases Reported, f_t, and Actual COVID Incidence, h_t

For our estimate of the fraction of COVID cases reported, f_t, we relied upon recent estimates issued by the Institute for Health Metrics and Evaluation (IHME) [13]. The authors estimated that by late December 2021 and early January 2022, f_t, what they termed the "infection-detection rate," had fallen to 0.25. (See Figure 8-1 in [13].) By relying on reported COVID incidence r_t rather than actual COVID incidence h_t to estimate ρ above, we effectively assumed f_t to be constant during the 4 weeks covered by our panel, and we do so here as well. Thus, for each week t = 1,…,4, we compute actual incidence as $h_t = r_t / 0.25 = 4r_t$.

Estimating q, the Proportion Hospitalized

A recent technical report from the UK Health Security Agency [14] noted that 3,019 Omicron cases were hospitalized among 528,176 Omicron cases total, which gives $q \approx 0.006$. While hospitalization practices may differ in the UK's National Health Service, this source has the advantage that the denominator closely approximates all Omicron infections, and not just symptomatic cases.

Aggregate and County-Specific Estimates of the Fraction of Incidental Hospitalizations, π

We estimated the fraction of incidental hospitalizations π not only for all 164 counties combined, but also for each county individually. We focused on the first week in January, that is, for the final week t = 4 in our four-week panel. For the aggregate calculation, we estimated the reported incidence r and the proportion of hospitalized patients who are COVID-positive c for all counties combined. (Since we are focusing on a specific week, we have dropped the subscripts t.) Given the estimated parameters ρ, f, θ, and q, we then computed an aggregate value of π. For the county-specific calculations, we estimated the reported incidence r_i and the proportion of hospitalized patients who were COVID-positive c_i for each county i (again

dropping the subscript t). Assuming the parameters ρ, f, θ, and q to be constant across counties, we then computed county-specific values for each π_i.

Results

Aggregate Fraction of Hospitalized Patients Who Are COVID-Positive, c

From the week ending June 25, 2021 through the week ending January 7, 2022, Fig. 1 plots the fraction of all hospital inpatients who were COVID-positive in our cohort of 250 high-volume hospitals. During the previous summer's Delta wave, this fraction rose to 16.74 percent for the week ending August 27. More recently, during the ongoing Omicron wave, the fraction has been rapidly rising to 25.17 percent for the most recent week ending January 7. We took this most recent datum as our current estimate of the parameter c. The corresponding odds ratio was $\frac{c}{1-c}$ = 0.336. Put differently, one in four inpatients of our cohort of 250 high-COVID-volume hospitals was hospitalized.

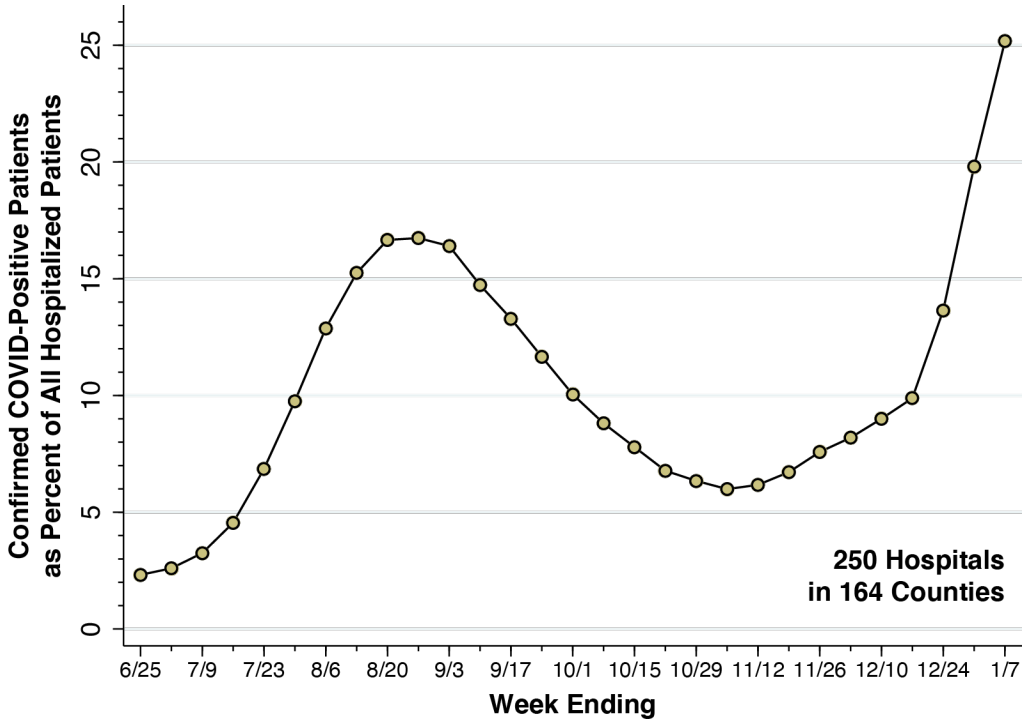

Fig. 1. Fraction of All Inpatients Who Were COVID-Positive in a Cohort of 250 High-Volume Hospitals, Weeks Ending June 25, 2021 through January 7, 2022. Source: U.S. Department of Health and Human Services [11]. For the most recent week ending January 7, 2022, the fraction of inpatients who were COVID-positive was 0.2517, which we took as the value of the parameter c. The corresponding odds ratio was $\frac{c}{1-c}$ = 0.336.

Growth Rate of COVID Incidence, 164 Counties

Fig. 2 displays a whiskers-on-box plot of reported weekly COVID incidence r_{it} in the 164 counties containing our 250-hospital cohort. The plot covers the four most recent weeks of data through the week ending January 9, 2022. During the latter week, Potter County TX, which includes Amarillo, was at the 25th percentile of reported incidence (1,069 cases per 100,000 population). Baltimore County MD was at the median (1,466 cases per 100,000 population), and Hartford County CT was at the 75th percentile (1,858 cases per 100,000 population). The two counties with the highest reported incidence were Miami-Dade County FL (4,065 per 100,000 population) and Bronx County NY (3,865 per 100,000 population).

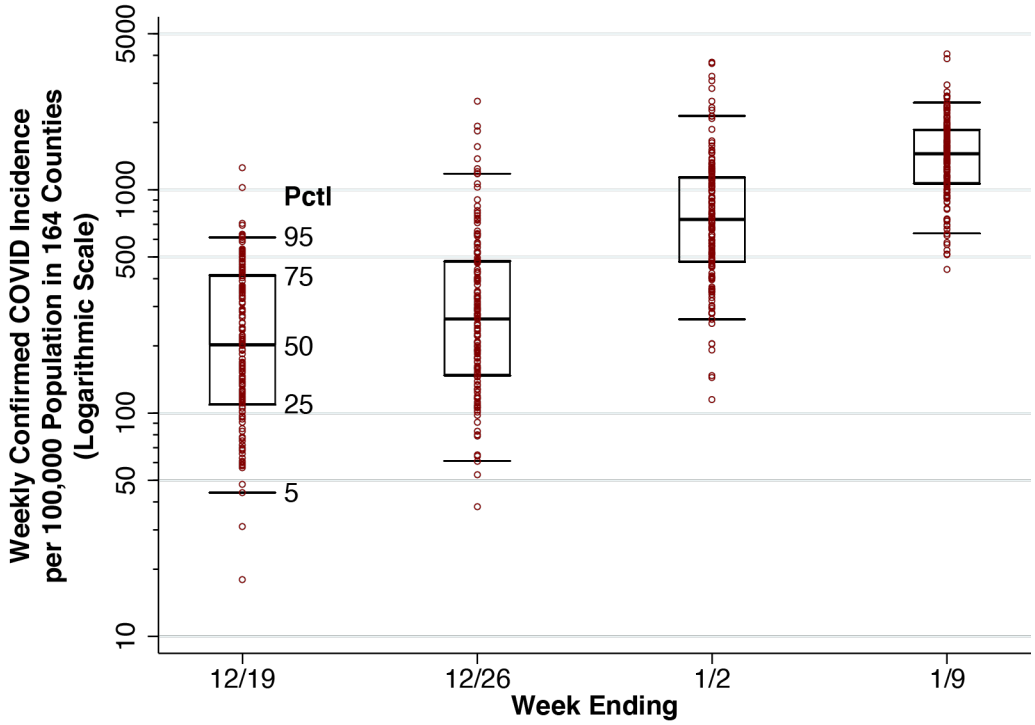

Fig. 2. Whiskers-on-Box Plots of Weekly Confirmed COVID Incidence per 100,000 Population in the 164 Counties Containing the 250 Study Hospitals, Weeks Ending December 19, 2021 Through January 9, 2022. For each week, the 5th, 25th, 50th, 75th, and 95th percentiles are superimposed upon the individual county-specific datapoints. A total of 11 datapoints with zero incident cases are omitted from the first two weeks.

For the four most recent weeks, Fig. 3 plots the population-weighted mean values of confirmed COVID incidence among all 164 counties combined. These values represent our estimates of reported incidence r_t. The fixed-effect log linear regression described on our panel of 164 counties over 4 weeks, described in equation (5), gave an estimate of ρ = 0.0934 per day

(95% confidence interval: 0.0893 – 0.0987). The fitted line in the figure with the same slope shows no evidence of serial correlation of residuals, thus supporting a model of recent exponential growth.

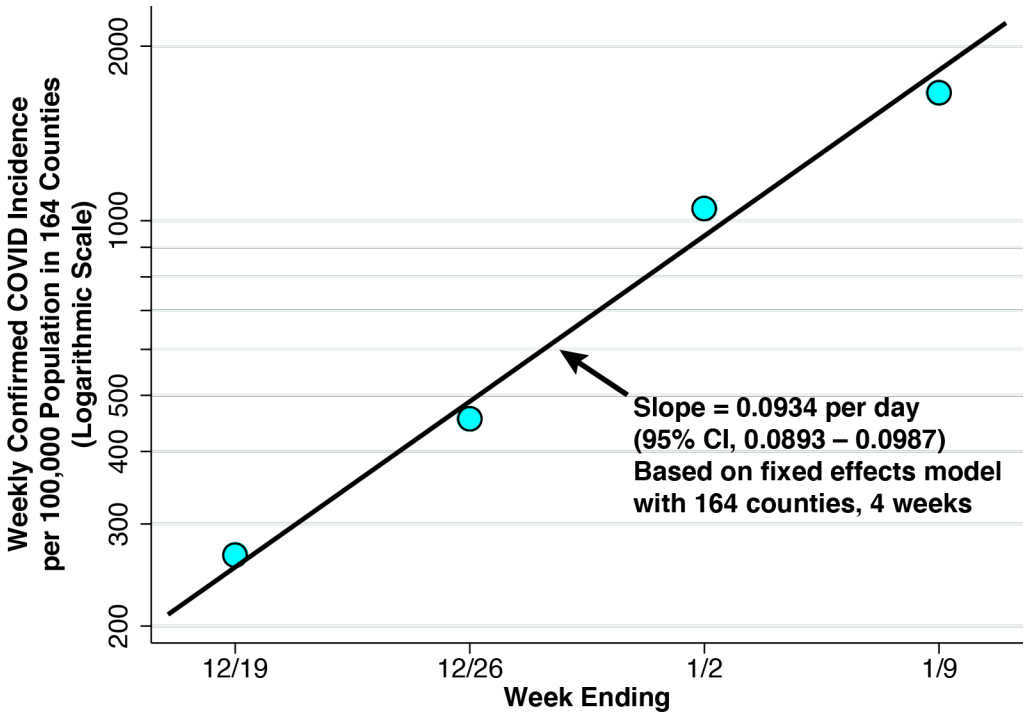

Fig. 3. Weekly Confirmed COVID Incidence per 100,000 Population in the 164 Counties Containing the 250 Study Hospitals, Weeks Ending December 19, 2021 Through January 9, 2022. The superimposed line has a slope of 0.0934 per day, derived from a linear fixed effects regression model of the logarithm of weekly COVID cases versus the date on which each week ended (equation 5). The balanced panel had 4 observations for each of 164 counties. No serial correlation of residuals is evident.

Actual COVID Incidence and Prevalence, 164 Counties

For the most recent week ending January 9, 2022, the mean weekly confirmed COVID incidence for the 164 counties in our panel was 1,663 per 100,000 population (Fig. 3), which comes to a reported incidence of r = 237.6 per 100,000 per day. Given our estimate of the fraction of cases reported at f = 0.25, we obtain an estimate of actual incidence of $h = r/f$ = 950.4 per 100,000 per day.

In a recent review of the literature, the U.S. Centers for Disease Control and Prevention (CDC) has estimated that the mean duration of infectiousness for Omicron is 5-6 days [15]. That gives us $\theta = 1/5.5 = 0.101$. From equation (3), we thus estimate actual COVID prevalence in our panel of 164 counties during the most recent week ending January 9 to be $p \approx \frac{1}{\rho+\theta} h$, which comes to 4.89 percent of the population.

Estimated Fraction of Incidental Hospitalizations

Given estimates of the parameters q = 0.006, c = 0.2517, and p = 0.0489, we rely upon equation (2) to compute $\pi = (1-q)\left(\frac{p}{1-p}\right)/\left(\frac{c}{1-c}\right)$ = 0.152. That is, we estimate 15.2 percent incidental COVID hospitalizations for the 164 counties containing our high-volume hospital cohort during the first week of January 2022. Our calculations are summarized in Table 3.

Table 3. Estimates of Component Parameters in the Derivation of the Fraction of Incidental Hospitalizations, π

Symbol	*Definition*	*Source*	*Estimate**
r	Reported COVID incidence	Community Profile Reports [12], Fig. 3	237.6 per 100,000 population per day
ρ	Exponential rate of increase of COVID incidence	Community Profile Reports [12], Eq. (5)	0.0934 per day
f	Fraction of COVID cases reported	IHME Report [13]	0.25
h	Actual COVID incidence	Eq. (4)	950.4 per 100,000 population per day
θ	Inverse of the mean duration of SARS-CoV-2 infection (days)	CDC [15].	0.101
p	Prevalence of COVID	Eq. (3)	0.0489
q	Proportion of all infected individuals who are hospitalized because of their COVID illness	UK Health Security Agency [14]	0.006
c	Fraction of hospitalized patients who are COVID-positive	HHS [11], Fig. 1	0.2517
π	Fraction of incidental hospitalizations	Eq. (2)	0.152

*Estimates apply specifically to the 164 counties containing 250 high-COVID-volume hospitals during the first week of January 2022.

Variability of the Fraction of Incidental Hospitalizations, π, Across Counties

Our overall estimate of π = 15.2 percent is an average that does not capture its variability across the 164 counties under study. When we applied the formula of equation (2) individually to each county, once again restricting the computations to the first week of January 2022, the median value of π was 13.7 percent, with an interquartile range of 9.5 to 18.4 percent.

Fig. 4 displays this variability in the form of a scatterplot of $\frac{p}{1-p}$ versus $\frac{c}{1-c}$ for 162 of 164 counties. If all counties had the same value of π, they would line up along the superimposed ray from the origin. Those datapoints above the ray have higher values of π, while those below the ray have lower values. Geometrically speaking, the value of π at each datapoint depends on the slope of the ray drawn from that point back to the origin. Thus, Allegheny County PA, which houses Pittsburgh, has the highest estimated value of 58.3 percent. By contrast, while Miami-Dade has the highest estimated prevalence of p = 11.9 percent, the estimated fraction of incidental COVID hospitalizations is 20.8 percent. As equation (2) shows, when the overall prevalence p of COVID in the county is higher, the fraction of incidental COVID hospitalizations π tends to be higher. But as hospitals in the county fill up with COVID patients, thus pushing c higher, the fraction of incidental COVID hospitalizations π tends to go lower.

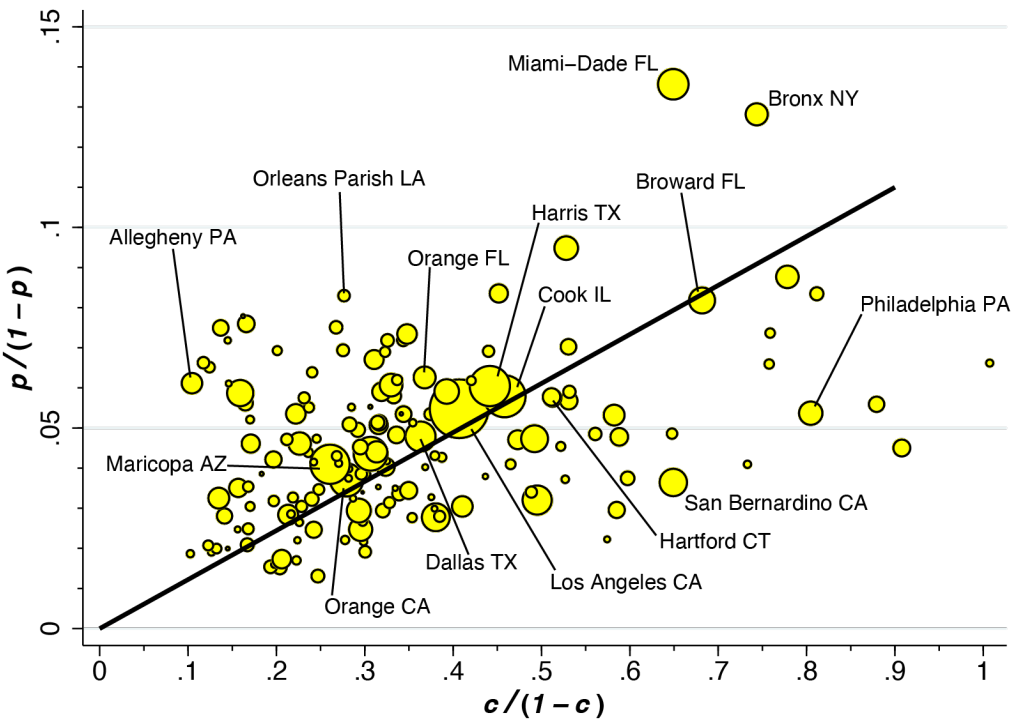

Fig. 4. Plot of $p/(1-p)$ Versus $c/(1-c)$ Among 162 or 164 Counties During the First Week of January 2022. The ray from the origin has slope 0.122 (95% CI: 0.114 – 0.131) based on a population-weighted regression.

Discussion

Fraction of Reported COVID Cases, f, as a Principal Source of Uncertainty

The principal source of uncertainty in our estimate of incidental COVID hospitalization is the fraction f of reported COVID cases. During the Omicron wave, as noted above, this fraction has declined to 25 percent of actual cases [13], in great part as a result of a surge in asymptomatic and mildly symptomatic infections, as well as an increasing proportion of rapid test results not tabulated by public health authorities. Our formula for the prevalence p of infection (equation 3) is a linear function of the actual incidence h, which is turn directly proportional to the inverse of f (equation 4). Thus, if only 20 percent of cases were reported (that is, f = 0.20), our estimate of COVID prevalence p would increase to 6.1 percent and the resulting estimate of the fraction of incidental hospitalizations π would rise to 19.2 percent. At the other extreme, if as many as 50 percent of cases were reported (that is, f = 0.50), our estimate of COVID prevalence p would drop to 2.45 percent and the resulting estimate of π would fall to 7.4 percent.

Equality of Hospitalization Rates as a Critical Assumption

In the derivation of our population-based formula, we made the key assumption that those individuals without COVID have the same probability g of hospitalization as those with non-severe COVID. Strict adherence to the notion of causality would seem to require this assumption. If a concurrent COVID infection increases the probability of hospitalization, then the infection has a causal role in the hospitalization and is thus not incidental. If a hospitalization is purely incidental, then the probability of hospitalization would be the same with or without COVID.

Consider a patient the hypertensive heart disease who, as a result of high fevers and dehydration from an Omicron infection, develops atrial fibrillation, a cardiac arrhythmia. He needs to be hospitalized to get his heart rate down and convert his heart rhythm back to normal. His would not be an incidental COVID hospitalization. Consider instead a patient with a history of extreme myopia since childhood who suffers a spontaneous retinal detachment and is hospitalized for eye surgery. During her admission workup, she is found to be COVID-positive. Since the infection did not apparently affect her probability of hospitalization, hers would be an incidental COVID hospitalization. (This is not to deny that the COVID epidemic has delayed the

time when patients have sought treatment and thus affected the severity of vision loss upon initial presentation [16].)

The problem with the foregoing logic is that individuals in the two groups – those without COVID and those with non-severe COVID – are different people. Thus, one individual with a white-collar occupation may be able to work remotely and does not contract COVID. Another individual with a blue-collar occupation cannot work remotely and comes down with a non-severe infection at work. The latter individual may also have a higher risk of hospitalization from an onsite workplace accident.

Generalizing our formula for π, we suppose that those without a COVID infection have a hospitalization probability g' that is not necessarily equal to the hospitalization probability g of those with non-severe COVID. Then the fraction of incidental COVID hospitalizations becomes $R\pi$, where $R = \frac{g'}{g}$ is the relative risk of hospitalization between the two groups and π is as already defined in equation (2). This does not mean that our strong assumption that $R = 1$ necessarily understates π. In the foregoing example comparing white-collar and blue-collar workers, $R < 1$ and thus π would be overstated. More generally, when those individuals who take precautions to reduce the risk of infection, such as getting vaccinated [17], also tend to adopt other preventive measures to reduce the risk of hospitalization generally, such as not smoking, our strong assumption that $R = 1$ will tend to overstate π.

Reliability of the Estimate of q from UK Data

Relying on data reported by the UK Health Security Agency [14], we took q, the probability that an infected individual would have a severe case requiring hospitalization, as 0.006. It is worth inquiring whether there are any U.S.-based sources that might provide a more reliable estimate. The difficulty is that computation of q needs to be based upon a population denominator that includes all cases of COVID, even asymptomatic and unreported cases.

A study of symptomatic patients infected with the Omicron variant in the Houston Methodist hospital system [18] revealed that out of 2,232 symptomatic patients, 313 were admitted to the hospital. This source, it would seem, yields an estimate of $q = 0.14$, which is an order of magnitude greater than the estimate derived from the UK data. The problem with relying upon this alternative data source is that the denominator reflects symptomatic patients primarily presenting to the emergency department, rather than all COVID cases. In fact, the percentage

reported by Houston Methodist quite close to the rate of 15.1 hospital admissions per 100 emergency department visits for COVID that we computed from our 250-hospital cohort, as shown in Appendix Fig. A2.

Variability of π Across Counties

While our aggregate measure of the incidental COVID fraction for all 164 counties is on the order of 15 percent, Fig. 4 shows significant variability across individual counties. Some of this variability is no doubt due to sampling errors in the measurement of the parameters c_i, the proportions of inpatients who are COVID-positive, especially in counties with only one high-volume hospital. Our assumption that the fraction of COVID cases reported f was uniform across counties may have introduced errors in the county-specific incidence estimates $h_i = r_i/f$. Likewise, our assumption that the exponential rate of increase of COVID incidence ρ was uniform across counties may have introduced errors in the county-specific prevalence estimates $p_i = \frac{h_i}{\rho+\theta}$.

Still, some of the variability of the variability may be due to systemic differences between hospitals in their range of services and patient populations. These hospitals may be genuine outliers. For example, a hospital may have a large, specialized transplant service with many immunosuppressed patients who are persistently COVID-positive.

Concluding Comments

Assuming that only one-quarter (that is, f = 0.25) of all recent Omicron infections have been reported by public authorities, we estimated the aggregate prevalence of active SARS-CoV-2 infection during the first week of January to be 4.89% (that is, p = 0.0489). During the same week, among 250 high-COVID-volume hospitals within our 164-county panel, an estimated 1 in 4 inpatients were found to be COVID-positive (more precisely, c = 0.2517). Among all COVID-positive hospitalized patients in all 164 counties combined, an estimated 15.2% were incidental infections (that is, π = 0.152). Across individual counties, the median fraction of incidental COVID hospitalizations was 13.7%, with an interquartile range of 9.5 to 18.4%.

Scattered, anecdotal reports have suggested that as many as half of COVID-positive hospital inpatients are merely incidental cases [1-3]. In the absence of a sufficiently large, longitudinal study of a representative sample of hospitalized patients, and without uniformly

established clinical criteria for determining when a COVID infection is incidental, it is hardly obvious how such fragmentary evidence is supposed to be interpreted.

In this study, we have inquired whether such estimates of the fraction of incidental COVID infections are consistent with available data on COVID cases in hospitals and in the population generally. Our population-based estimates suggest that incidental COVID infections are indeed a nontrivial fraction of all COVID-positive hospitalized patients. In the aggregate, however, the burden of patients admitted for complications of their viral infections appears to be far greater.

Appendix

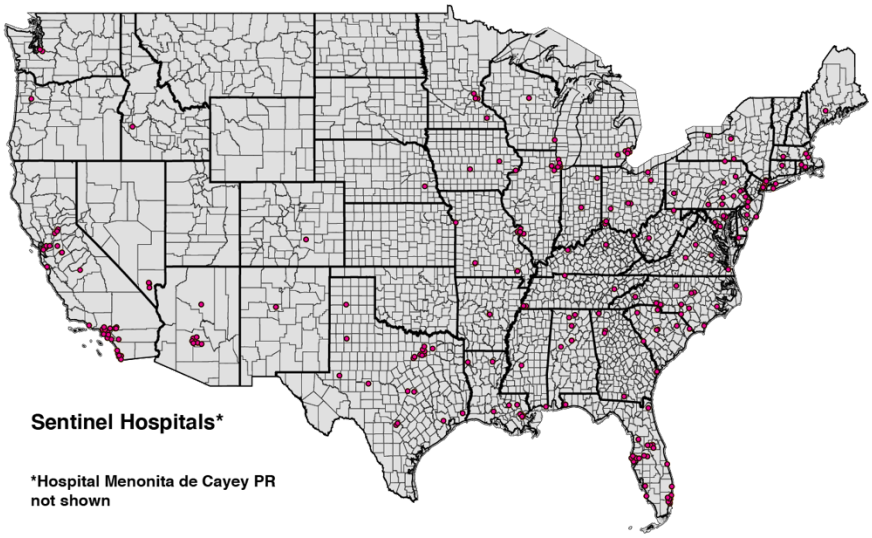

Fig. A1. U.S. Continental Map Showing Locations of 249 of the 250 High-Volume Hospitals. Hospital Menonita de Cayey, Cayey, Puerto Rico, is not shown. State and county boundaries are indicated.

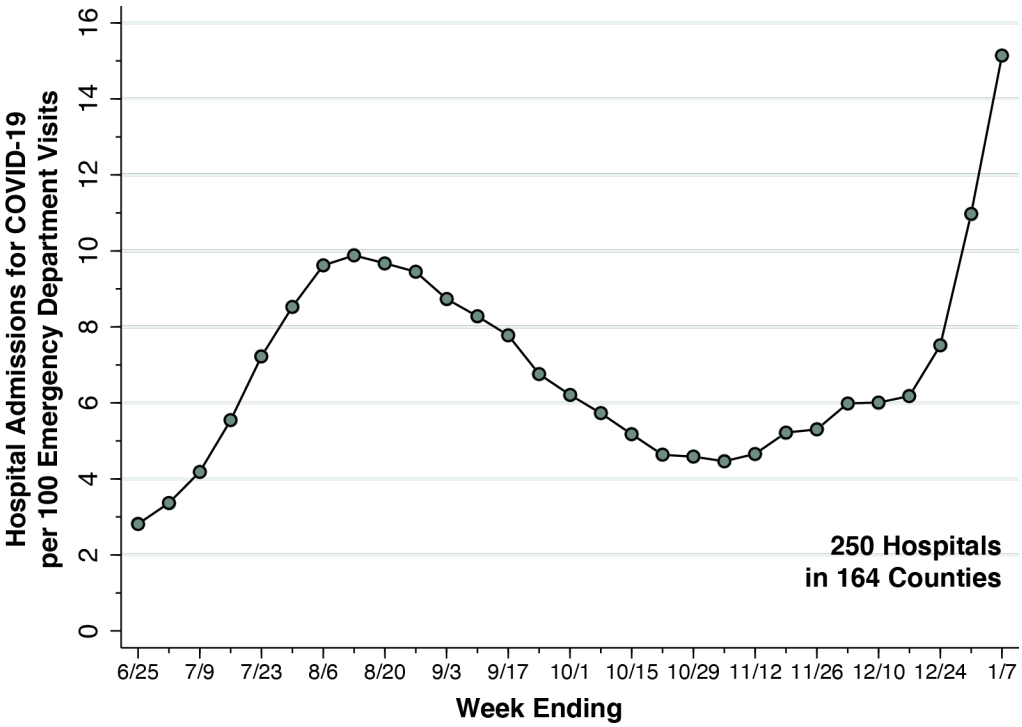

Fig. A2. Hospital Admissions for COVID-19 per 100 Emergency Department Visits in a Cohort of 250 High-Volume Hospitals, Weeks Ending June 25, 2021 through January 7, 2022. See Technical Notes to Fig. A2 below.

Technical Notes to Fig. A2.

The data source was U.S. Department of Health and Human Services [11], as described under *Data: Cohort of 250 High-Volume Hospitals* in the section on Methods and Data in the main text.

Emergency department visits to each of the cohort hospitals was derived from the variable *previous_day_covid_ED_visits_7_day_sum*, defined as "Sum of total number of ED visits who were seen on the previous calendar day who had a visit related to COVID-19 (meets suspected or confirmed definition or presents for COVID diagnostic testing – do not count patients who present for pre-procedure screening) reported in 7-day period."

Hospital admissions were determined as the sum of two variables:

a) *previous_day_admission_adult_covid_confirmed_7_day_sum*, defined as "Sum of number of patients who were admitted to an adult inpatient bed on the previous calendar day who had confirmed COVID-19 at the time of admission reported in the 7-day period."
b) *previous_day_admission_pediatric_covid_confirmed_7_day_sum*, defined as "Sum of number of pediatric patients who were admitted to an inpatient bed, including NICU, PICU, newborn, and nursery, on the previous calendar day who had confirmed COVID-19 at the time of admission."

References

1. Murray, S.G., R. Croci, and R.M. Wachter, *Is a patient hospitalized 'with' covid or 'for' covid? It can be hard to tell.* 2022, https://www.washingtonpost.com/outlook/2022/01/07/hospitalization-covid-statistics-incidental/: Washington Post, January 7.
2. Lovelace, B.J., *Many patients hospitalized for other ailments are also testing positive for Covid*. 2022, https://www.nbcnews.com/health/health-news/omicron-hospital-many-patients-hospitalized-ailments-also-test-positiv-rcna11247: NBC News, January 8.
3. Radcliffe, S., *The Difference Between Being Hospitalized 'For COVID' and 'with COVID'*. 2022, https://www.healthline.com/health-news/the-difference-between-being-hospitalized-for-covid-and-with-covid: Healthline, January 13.
4. U.S. Centers for Disease Control and Prevenetion, *COVID Data Tracker: Variant Proportions*. 2022, https://covid.cdc.gov/covid-data-tracker/#variant-proportions: Accessed January 20.
5. Alho, J.M., *On prevalence, incidence, and duration in general stable populations.* Biometrics, 1992. **48**(2): p. 587-92.
6. Noh, J. and G. Danuser, *Estimation of the fraction of COVID-19 infected people in U.S. states and countries worldwide.* PLoS One, 2021. **16**(2): p. e0246772.
7. Havers, F.P., et al., *Seroprevalence of Antibodies to SARS-CoV-2 in Six Sites in the United States, March 23-May 3, 2020*. 2020, https://www.medrxiv.org/content/10.1101/2020.06.25.20140384v1: MedRxiv June 26, 2020.
8. Moghadas, S.M., et al., *The implications of silent transmission for the control of COVID-19 outbreaks.* Proc Natl Acad Sci U S A, 2020. **117**(30): p. 17513-17515.
9. Harris, J.E., *Critical Role of the Subways in the Initial Spread of SARS-CoV-2 in New York City.* Front Public Health, 2021. **9**: p. 754767.
10. Harris, J.E., *Los Angeles County SARS-CoV-2 Epidemic: Critical Role of Multi-generational Intra-household Transmission.* Journal of Bioeconomics, 2021. **23**: p. 55-83.
11. U.S. Department of Health and Human Services, *COVID-19 Reported Patient Impact and Hospital Capacity by Facility*. 2022, https://healthdata.gov/Hospital/COVID-19-Reported-Patient-Impact-and-Hospital-Capa/anag-cw7u: HealthData.gov, Updated January 17, 9:13AM PST.
12. White House COVID-19 Team, *COVID-19 Community Profile Report*. 2022, https://healthdata.gov/Health/COVID-19-Community-Profile-Report/gqxm-d9w9: HealthData.gov, Excel Files for 12/20/2021, 12/27/2021, 1/3/2022,1/10//2022,.
13. Institute for Health Metrics and Evaluation (IHME), *COVID-19 Results Briefing: United States of America, January 8*. 2022, https://www.healthdata.org/sites/default/files/files/102_briefing_United_States_of_America_3.pdf: Accessed January 17.
14. UK Health Security Agency, *SARS-CoV-2 variants of concern and variants under investigation in England – Technical briefing: Update on hospitalisation and vaccine effectiveness for Omicron VOC-21NOV-01*

(B.1.1.529). 2021, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1045619/Technical-Briefing-31-Dec-2021-Omicron_severity_update.pdf: December 31.

15. U.S. Centers for Disease Control and Prevenetion, *Ending Isolation and Precautions for People with COVID-19: Interim Guidance*. 2022, https://www.cdc.gov/coronavirus/2019-ncov/hcp/duration-isolation.html: Updated January 14.
16. Patel, L.G., et al., *Clinical Presentation of Rhegmatogenous Retinal Detachment during the COVID-19 Pandemic: A Historical Cohort Study.* Ophthalmology, 2021. **128**(5): p. 686-692.
17. Harris, J.E., *COVID-19 Incidence and Hospitalization During the Delta Surge Were Inversely Related to Vaccination Coverage Among the Most Populous U.S. Counties.* Health Policy Technol, 2021: p. 100583.
18. Christensen, P.A., et al., *Signals of significantly increased vaccine breakthrough, decreased hospitalization rates, and less severe disease in patients with COVID-19 caused by the Omicron variant of SARS-CoV-2 in Houston, Texas*. 2022, https://www.medrxiv.org/content/10.1101/2021.12.30.21268560v3: medRxiv, January 12.