# Population-Based Estimation of the Fraction of Incidental COVID-19 Hospitalizations During the Omicron Wave in the United States

Jeffrey E. Harris MD PhD* June 19, 2022

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

**Abstract**

*Background*

Scattered reports have suggested that as many as one-half of all hospital inpatients identified as COVID-19-positive during the Omicron variant-driven wave were incidental cases admitted primarily for reasons other than their viral infections. To date, however, there are no systematic 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.

*Methods*

To fill this gap, we developed a formula to estimate the fraction of incidental COVID-19 hospitalizations that relies upon measurable, population-based parameters. We applied our 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.

*Results*

Within this panel, we estimated that COVID-19 incidence was rising exponentially at a rate of 9.34% per day (95% CI, 8.93-9.87). Assuming that only one-quarter of all Omicron infections had 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. Based upon these estimates, we computed that 15.2% of such COVID-19-positive hospitalized patients were incidental infections. Across individual counties, the median fraction of incidental COVID-19 hospitalizations was 13.7%, with an interquartile range of 9.5 to 18.4%.

*Conclusion*

Incidental COVID-19 infections appear to have been a nontrivial fraction of all COVID-19-positive hospitalized patients during the Omicron wave. In the aggregate, however, the burden of patients admitted for complications of their viral infections was far greater.

**Keywords**

SARS-CoV-2; Omicron variant; incidental hospitalizations

**Background**

During the massive wave of Omicron variant-driven SARS-CoV-2 infections in the United States during the winter of 2021–2022, total hospitalizations among infected persons reached 161,876 on January 19, 2022, a nationwide census substantially exceeding the peaks recorded during the previous winter of 2020–2021 and the Delta variant-driven wave of July–October 2021 [1]. Yet some observers cautioned that more than half of all hospital inpatients identified as COVID-19-positive during the Omicron wave were so-called incidental cases admitted primarily for reasons other than their viral infections [2]. Reports by individual hospitals and state health departments did not generally detail their methodology for determining whether a hospitalization was incidental [3, 4]. A notable exception was the Massachusetts Department of Public Health, which confined non-incidental admissions to patients who had been administered dexamethasone [5], a treatment principally indicated for severe cases requiring high-flow oxygen but contraindicated in some hospitalized COVID-19 patients [6, 7].

To date, there have been no broad-based, systematic studies of a representative panel of hospitals to determine what fraction of COVID-19-positive hospitalizations are indeed incidental. The ideal design would be to prospectively follow a cohort of individuals, evaluating each hospitalization according to pre-established, objective criteria for determining whether each patient was in fact admitted as a result of the disease. At issue here are more than technical matters of disease classification, but also fundamental questions of ultimate causation [8]. Consider a young adult recently infected with SARS-CoV-2 who continues to drive while suffering from viremia-induced headache, fever, body aches, dizziness and severe fatigue and, as a result, gets into a major car crash. His admission diagnoses might list such proximate causes as acute brain injury, spleen laceration and a pelvic fracture. But the ultimate cause was COVID-19.

To fill this gap, we developed a formula to estimate the fraction of incidental COVID-19 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 [9]. While this exercise required us to make some difficult-to-verify assumptions, it still educates us as to whether there was any reasonable possibility that the incidental COVID-19 fraction was as high some sources reported.

## Methods

*Defining the Fraction of Incidental COVID Hospitalizations, $\pi$*

We developed a model to formally define the fraction of incidental COVID-19 hospitalizations and to express this fraction as a function of observable quantities. To that end, consider a closed population at a particular point in time. We refer to any individual within this population who has a detectable SARS-CoV-2 infection at that time as a *COVID-19 case*. This definition does not require that the individual has tested positive or has symptoms of COVID-19.

Let $q$ (where $1 > q > 0$) denote the proportion of all COVID-19 cases that are hospitalized *because* of their illness. For shorthand, we refer to these as the *severe* cases. We refer to all other COVID-19 cases, whether they are hospitalized or not, as *non-severe*. While the severe cases are, by definition, hospitalized because of their COVID-19 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-19 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-19 hospitalizations that are incidental is:

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

We refer to $\pi$ as the *fraction of incidental COVID-19 hospitalizations*. We see that $\pi$ is increasing in $g$ and decreasing in $q$.

*Rendering $\pi$, the Fraction of Incidental Hospitalizations, as a Function of Observables*

Ideally, one would estimate the quantity $q$ in equation (1) by longitudinally following a closed cohort of individuals as they contracted COVID-19 and then observing who was hospitalized because of their illness. Within this longitudinal cohort, we could also estimate $g$ by tracking who was hospitalized for other reasons. In the absence of such a longitudinal study, we can still proceed provided we make a key simplifying assumption, namely, that those individuals without COVID-19 have the same probability of hospitalization as those with non-severe COVID-19. We examine the validity of this key assumption later in the Discussion.

Let $p$ (where $1 > p > 0$) denote the proportion of individuals in the entire population who are COVID-19 cases. We refer to $p$ as the *prevalence of COVID-19*. Now focus more sharply on just those individuals who are hospitalized. We test every hospitalized patient to determine who

is COVID-19-positive and who is COVID-19-negative. Let $c$ (where $1 > c > 0$) denote the fraction of all hospitalized individuals who are COVID-19-positive. In the Supplement, we derive the following expression for $\pi$:

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

In equation (2), $\pi$ is now a function of the prevalence $p$ of COVID-19 in the population, the fraction $c$ of all hospitalized patients who are COVID-19-positive, and the proportion $q$ of COVID-19-positive individuals requiring hospitalization because of their disease. Table 1 below summarizes these points.

**Table 1. Component Parameters in the Derivation of the Fraction of Incidental Hospitalizations, $\pi$**

| *Symbol* | *Definition* |
|---|---|
| $p$ | Prevalence of COVID-19 |
| $c$ | Fraction of all hospitalized patients who are COVID-19-positive |
| $q$ | Proportion of all infected individuals who are hospitalized because of their COVID-19 illness |
| $\pi$ | Fraction of incidental hospitalizations, $\pi = (1-q)\left(\frac{p}{1-p}\right)\Big/\left(\frac{c}{1-c}\right)$ |

*Rendering p, the Prevalence of COVID-19, as a Function of Observables*

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 [10].

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 $1/\theta$, where $\theta > 0$. In the Supplement, we show that the prevalence of infection can be approximated as:

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

When the incidence $h$ is stable (that is, $\rho = 0$), this formula collapses to the classic result $p(t) = h(t)/\theta$. To estimate the prevalence $p$ during an exponentially growing epidemic, we therefore

need data on the incidence $h$, the growth rate $\rho$ of infection and the mean duration $1/\theta$ of infection.

It is widely acknowledged that COVID-19 cases have been and continue to be significantly underreported [11-14]. Accordingly, to estimate the incidence $h$ from data on reported cases, we follow the usual approach of incorporating an under-ascertainment factor into our analysis [15]. Let $r(t)$ denote the *reported* incidence of COVID-19 and let $f(t) > 0$ denote the fraction of COVID cases that are reported at time $t$. Then we have:

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

Accordingly, to estimate $p$, we need data on the parameters $r$, $\rho$, $f, h$, and $\theta$, as indicated in Table 2. For clarity, we have dropped the time argument $t$ from the functions $h$, $r$, and $f$.

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

| *Symbol* | *Definition* |
|---|---|
| $r$ | Reported COVID incidence |
| $\rho$ | Exponential rate of increase of COVID incidence |
| $f$ | Fraction of COVID cases reported |
| $h$ | Actual COVID incidence, $h = \frac{r}{f}$ |
| $\theta$ | 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) [16], we identified the 250 hospitals with the highest cumulative volume of emergency department visits for COVID-19 from the week ending June 25, 2021 through the week ending December 10, 2021. These high-volume hospitals occupied 164 counties in 41 states and territories. Fig. A1 in the Supplement displays a map of hospital locations within the continental United States.

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.

i. *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."

ii. *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."

iii. *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 $c$, the Fraction of Hospitalized Patients Who Are COVID-Positive*

For each hospital and each week, we calculated the number of COVID-19-positive patients as the sum of variables (ii) and (iii) above, while variable (i) gave the total number of hospital patients. For each week, summing over all 250 hospitals, we computed cohort-wide numbers of COVID-19-positive inpatients and total inpatients, from which we then computed $c$, the fraction of all inpatients who were COVID-19-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 $\pi$.

*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-19 incidence from the *Counties* tab of Excel spreadsheets regularly issued by the White House COVID-19 Team as Community Profile Reports [17]. 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, $\rho$, and Reported COVID Incidence, $r_t$*

We relied upon the county-specific data derived from the Community Profile Reports [17] to estimate the exponential rate of increase of COVID-19 incidence, $\rho$. 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):

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

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 $\mu$ was an overall constant term, the parameters $\eta_i$ were county-specific fixed effects, and the terms $\varepsilon_{it}$ were spherical errors. The parameter $\rho$ 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-19 Cases Reported, $f_t$, and Actual COVID-19 Incidence, $h_t$*

For our estimate of the fraction of COVID-19 cases reported, $f_t$, we relied upon estimates issued by the Institute for Health Metrics and Evaluation (IHME) [18]. 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 [18].) By relying on reported COVID-19 incidence $r_t$ rather than actual COVID-19 incidence $h_t$ to estimate $\rho$ 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 contemporaneous technical report from the UK Health Security Agency [19] 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 approximated all Omicron infections, and not just symptomatic cases.

*Aggregate and County-Specific Estimates of the Fraction of Incidental Hospitalizations, $\pi$*

We estimated the fraction of incidental hospitalizations $\pi$ 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-19-positive $c$ for all counties combined. (Since we are focusing on a specific week, we have dropped the subscripts $t$.) Given the estimated parameters $\rho$, $f$, $\theta$, and $q$, we then computed an aggregate value of $\pi$. 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 $\rho$, $f$, $\theta$, and $q$ to be constant across counties, we then computed county-specific values for each $\pi_i$.

## Results

*Aggregate Fraction of Hospitalized Patients Who Are COVID-19-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-19-positive in our cohort of 250 high-volume hospitals. During the summer's Delta wave, this fraction rose to 16.74 percent for the week ending August 27, 2021. During the Omicron wave, the fraction rose rapidly rising to 25.17 percent for the week ending January 7, 2022. We took the latter datum as our estimate of the parameter $c$. Accordingly, about one in four inpatients of our cohort of 250 high-COVID-volume hospitals was hospitalized.

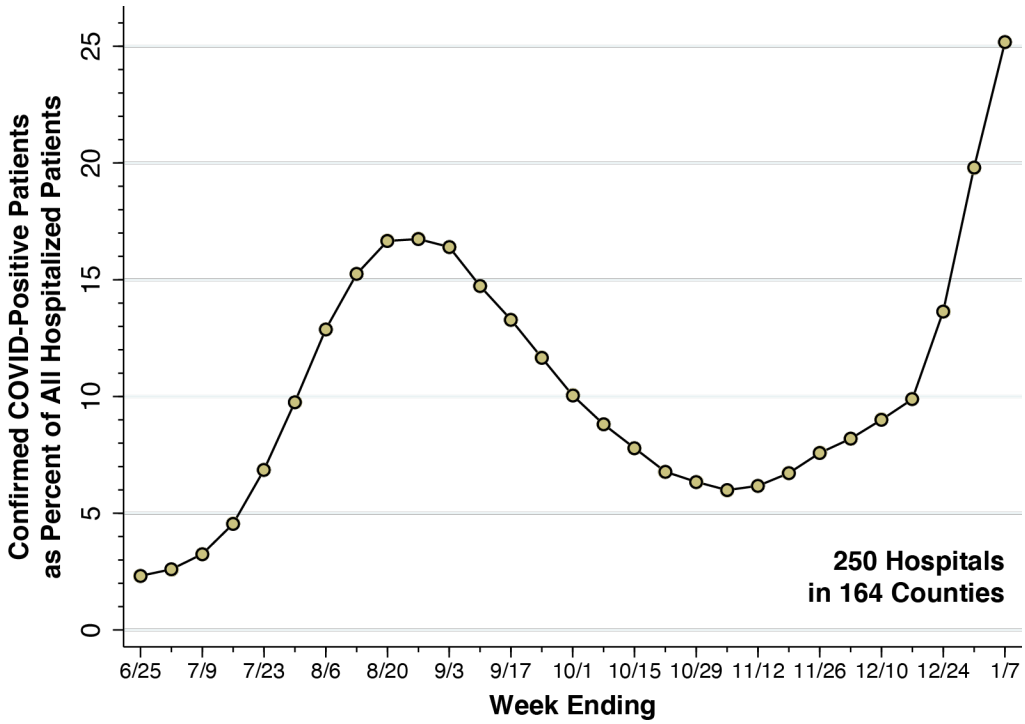


**Fig. 1. Fraction of All Inpatients Who Were COVID-Positive in a Cohort of 250 High-Volume Hospitals, Weeks Ending June 25, 2021, through January 7, 2022.** Source: U.S. Department of Health and Human Services [16]. For the most recent week ending January 7, 2022, the fraction of inpatients who were 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-19 Incidence, 164 Counties*

Fig. 2 displays a whiskers-on-box plot of reported weekly COVID-19 incidence $r_{it}$ in the 164 counties containing our 250-hospital cohort. The plot covers the four 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).

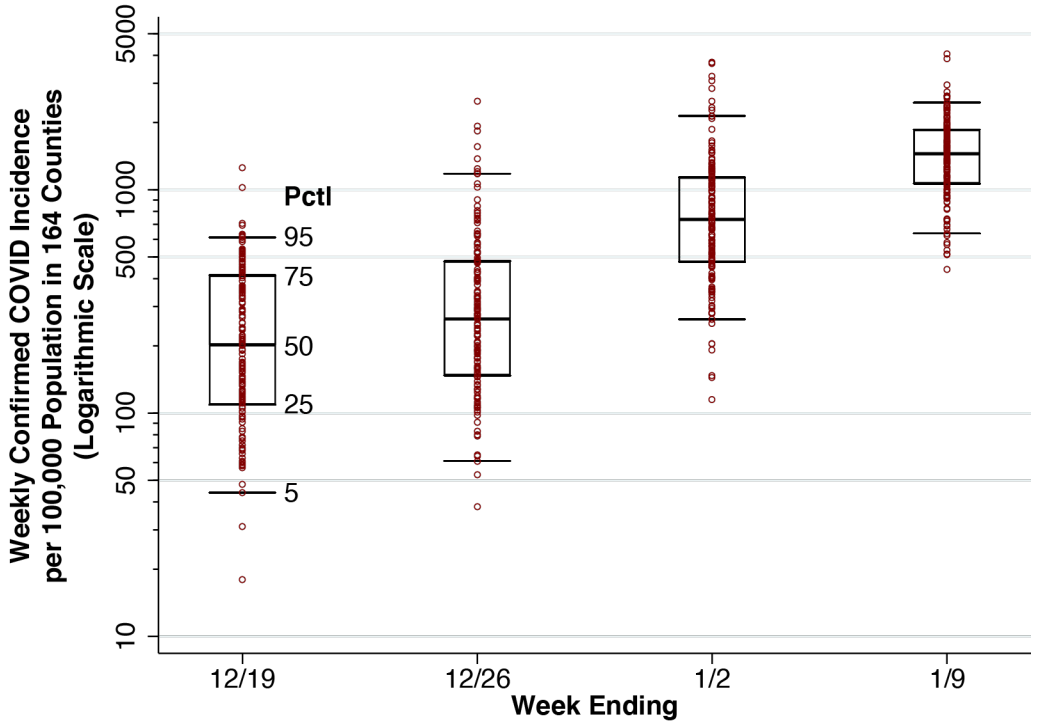


**Fig. 2. Whiskers-on-Box Plots of Weekly Confirmed COVID-19 Incidence per 100,000 Population in the 164 Counties Containing the 250 Study Hospitals, Weeks Ending December 19, 2021, Through January 9, 2022.** For each week, the 5th, 25th, 50th, 75th, and 95th percentiles are superimposed upon the individual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

For the same four weeks, Fig. 3 plots the population-weighted mean values of confirmed COVID-19 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 $\rho$ = 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.

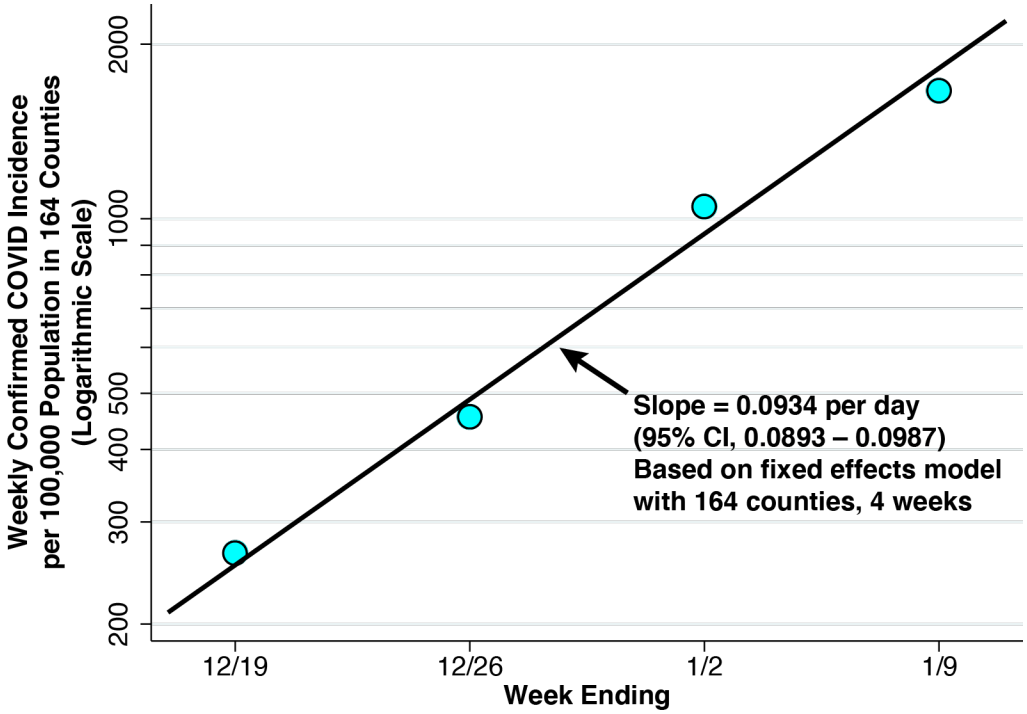


**Fig. 3. Weekly Confirmed COVID-19 Incidence per 100,000 Population in the 164 Counties Containing the 250 Study Hospitals, Weeks Ending December 19, 2021, Through January 9, 2022.** The superimposed line has a slope of 0.0934 per day, derived from a linear fixed effects regression model of the logarithm of weekly COVID-19 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-19 Incidence and Prevalence, 164 Counties*

For the most recent week ending January 9, 2022, the mean weekly confirmed COVID-19 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.

The U.S. Centers for Disease Control and Prevention (CDC) has estimated that the mean duration of infectiousness for Omicron is 5-6 days [20]. 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 relied 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, $\pi$**

| *Symbol* | *Definition* | *Source* | *Estimate** |
|---|---|---|---|
| $r$ | Reported COVID-19 incidence | Community Profile Reports [17], Fig. 3 | 237.6 per 100,000 population per day |
| $\rho$ | Exponential rate of increase of COVID-19 incidence | Community Profile Reports [17], Eq. (5) | 0.0934 per day |
| $f$ | Fraction of COVID-19 cases reported | IHME Report [18] | 0.25 |
| $h$ | Actual COVID-19 incidence | Eq. (4) | 950.4 per 100,000 population per day |
| $\theta$ | Inverse of the mean duration of SARS-CoV-2 infection (days) | CDC [20]. | 0.101 |
| $p$ | Prevalence of COVID-19 | Eq. (3) | 0.0489 |
| $q$ | Proportion of all infected individuals who are hospitalized because of their COVID-19 illness | UK Health Security Agency [19] | 0.006 |
| $c$ | Fraction of hospitalized patients who are COVID-19-positive | HHS [16], Fig. 1 | 0.2517 |
| $\pi$ | 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 $\pi$ = 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 $\pi$ 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 $\pi$, they would line up along the superimposed ray from the origin. Those datapoints above the ray have higher values of $\pi$, while those below the ray have lower values. Geometrically speaking, the value of $\pi$ 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-19 in the county is higher, the fraction of incidental COVID-19 hospitalizations $\pi$ tends to be higher. But as hospitals in the county fill up with COVID-19 patients, thus pushing $c$ higher, the fraction of incidental COVID hospitalizations $\pi$ tends to go lower.

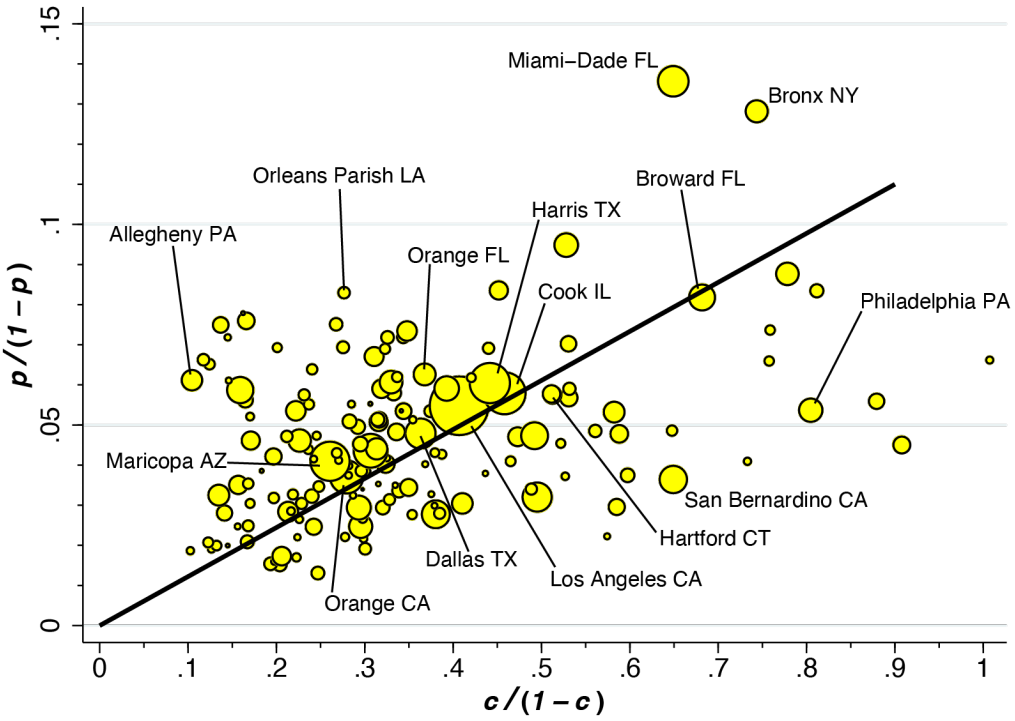


**Fig. 4. Plot of $p/(1-p)$ Versus $c/(1-c)$ Among 162 or 164 Counties During the First Week of January 2022.** The ray from the origin has slope 0.122 (95% CI: 0.114 – 0.131) based on a population-weighted regression.

## Discussion

*Fraction of Reported COVID-19 Cases, $f$, as a Principal Source of Uncertainty*

The principal source of uncertainty in our estimate of incidental COVID-19 hospitalization is the fraction $f$ of reported COVID cases. During the Omicron wave, as noted above, this fraction declined to 25 percent of actual cases [18], 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-19 prevalence $p$ would increase to 6.1 percent and the resulting estimate of the fraction of incidental hospitalizations $\pi$ 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-19 prevalence $p$ would drop to 2.45 percent and the resulting estimate of $\pi$ 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-19 have the same probability $g$ of hospitalization as those with non-severe COVID-19. Strict adherence to the notion of causality would seem to require this assumption. If a concurrent COVID-19 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-19.

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-19 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-19-positive. Since the infection did not apparently affect her probability of hospitalization, hers would be an incidental COVID-19 hospitalization. (This is not to deny that the COVID-19

epidemic has delayed the time when patients have sought treatment and thus affected the severity of vision loss upon initial presentation [21].)

The problem with the foregoing logic is that individuals in the two groups – those without COVID-19 and those with non-severe COVID-19 – are different people. Thus, one individual with a white-collar occupation may be able to work remotely and does not contract COVID-19. 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 $\pi$, we suppose that those without a COVID-19 infection have a hospitalization probability $g'$ that is not necessarily equal to the hospitalization probability $g$ of those with non-severe COVID-19. Then the fraction of incidental COVID-19 hospitalizations becomes $R\pi$, where $R = \frac{g'}{g}$ is the relative risk of hospitalization between the two groups and $\pi$ is as already defined in equation (2). This does not mean that our strong assumption that $R = 1$ necessarily understates $\pi$. In the foregoing example comparing white-collar and blue-collar workers, $R < 1$ and thus $\pi$ would be overstated. More generally, when those individuals who take precautions to reduce the risk of infection, such as getting vaccinated [22], 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 $\pi$.

*Reliability of the Estimate of q from UK Data*

Relying on data reported by the UK Health Security Agency [19], 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-19, even asymptomatic and unreported cases.

A study of symptomatic patients infected with the Omicron variant in the Houston Methodist hospital system [23] 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-19 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 Supplement Fig. A2.

*Variability of π Across Counties*

While our aggregate measure of the incidental COVID-19 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 proportions $c_i$ of inpatients who were COVID-19-positive, especially in counties with only one high-volume hospital. Our assumption that the fraction of COVID-19 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-19 incidence $\rho$ 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. For example, a hospital may have a large, specialized transplant service with many immunosuppressed patients who are persistently COVID-19-positive.

**Conclusion**

Assuming only one-quarter (that is, $f$ = 0.25) of all Omicron infections were 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-19-positive (more precisely, $c$ = 0.2517). Applying these estimates to our population-based formula, we found that among all COVID-19-positive hospitalized patients in all 164 counties combined, an estimated 15.2% were incidental infections (that is, $\pi$ = 0.152). Across individual counties, the median fraction of incidental COVID-19 hospitalizations was 13.7%, with an interquartile range of 9.5 to 18.4%.

Our population-based estimates suggest that incidental COVID-19 infections have indeed been a nontrivial fraction of all COVID-positive hospitalized patients, but they fall far short of the proportions suggested by some sources [2-5]. Even under our conservative assumption that there were four times as many COVID-19 infections as cases reported by public authorities,

there were not nearly enough prevalent COVID-19 cases to push the fraction of incidental hospitalizations even close to the one-half mark suggested by those sources. With the exception of one source [5], the methodology underlying such claims has not been clearly elucidated and, accordingly, we cannot resolve the discrepancy. Still, our examples of the patient with hypertensive heart disease who developed a cardiac arrythmia and the young man who got into a car crash as a result of their viral infections suggest that such methodologies have been overly restrictive in their assessment of COVID-19-caused admissions.

In the aggregate, the burden of patients admitted for complications of their COVID-19 infections appears to be far greater than the number of incidental cases.

## Author Declarations

*Authorship*

JEH is the sole author of this work. He is responsible for the conceptualization of the study, the data analysis, the drafting of the manuscript, and the creation of the figures.

*Funding*

The author received no specific funding for this work.

*Competing Interests*

The author has no competing interests to declare.

*Acknowledgments*

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.

*Human Studies Approval*

This study relies exclusively on publicly available data that contain no individual identifiers. No human studies approval was required.

*Public Posting of Data and Programs*

Supporting programs and data have been posted at https://osf.io/7umje/.

*List of Abbreviations*

| | |
|---|---|
| IHME | Institute for Health Metrics and Evaluation |
| CDC | U.S. Centers for Disease Control and Prevention |
| HHS | U.S. Department of Health and Human Services |

**References**

1. Johns Hopkins University, *United States. Data Timeline: Daily COVID-19 Hospitalizations*. 2022, https://coronavirus.jhu.edu/region/united-states: Johns Hopkins Coronavirus Resource Center, last accessed June 15.
2. 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.
3. Jackson Health System, *Jackson Health System hospitals currently have 564 patients who have tested positive for COVID-19 (Tweet)*. 2022, Twitter.com, January 12: https://twitter.com/JacksonHealth/status/1481268904970358790.
4. New York Governor, *Governor Hochul Updates New Yorkers on State's Progress Combating COVID-19* 2022, https://www.governor.ny.gov/news/governor-hochul-updates-new-yorkers-states-progress-combating-covid-19-131: Press Release, January 7.
5. Massachusetts Department of Public Health, *COVID-19 Hospitalizations (Powerpoint Presentation)*. 2022, https://www.mass.gov/doc/incidental-covid-19-report-january-25-2022/download: Commonwealth of Massachusetts, Executive Office of Health and Human Services, January 25.
6. U.S. National Institutes of Health, *Therapeutic Management of Hospitalized Adults With COVID-19*. 2022, https://www.covid19treatmentguidelines.nih.gov/management/clinical-management/hospitalized-adults--therapeutic-management/: NIH COVID-19 Treatment Guidelines, Last updated February 24.
7. Chen, F., et al., *Potential Adverse Effects of Dexamethasone Therapy on COVID-19 Patients: Review and Recommendations.* Infect Dis Ther, 2021. **10**(4): p. 1907-1931.
8. Mayr, E., *Cause and effect in biology.* Science, 1961. **134**(3489): p. 1501-6.
9. 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.
10. Alho, J.M., *On prevalence, incidence, and duration in general stable populations.* Biometrics, 1992. **48**(2): p. 587-92.

11. 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.
12. 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.
13. 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.
14. 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.
15. 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.
16. 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.
17. 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,.
18. 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.
19. 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.

20. 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.

21. 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.

22. 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.

23. 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.