Estimated Average Probabilities of COVID-19 Infection, Hospitalization, and Death From Community Contact in the United States

Rajiv Bhatia MD, MPH *

Assistant Professor of Medicine (affiliated), Stanford University
drajiv@stanford.edu

Jeffrey Klausner, MD, MPH
Professor of Medicine and Public Health, UCLA
JDKlausner@mednet.ucla.edu

Abstract: - 325 words
Manuscript: 2938 words
References: 899 words
Tables: 1
Figures: 3
Supplemental Materials

* Corresponding Author
**Key Points**

**Question:** What is the average probability of acquiring COVID-19 infection, being hospitalized, or dying from an unprotected community-level contact in US?

**Findings:** Among the 100 most populous US Counties, for the week ending May 30, 2020, the median probability of COVID-19 infection transmission is 1 infection per 3836 unprotected community-level contacts. For a 50 to 64 year old individual, the estimated median probability of hospitalization is 1 hospitalization per 852,000 community level person-contacts and the median probability of a fatality is 1 fatality per 19.1 million community-level person-contacts.

**Meaning:** Estimates of individual level probability for COVID-19 infection may inform more accurate risk perceptions and facilitate re-engagement with social activity.
Abstract

**Importance:** Case and death counts, government rules limiting social contact, media attention, peer behaviors, and analogies to war have all influenced public perceptions of risk to COVID-19. During an epidemic, the general public may not be accurately interpreting their individual risks from COVID19. U.S. public sentiment surveys indicate a high level of apprehension about routine community activities. Sufficient data allow estimating individual-level probabilities of infection, hospitalization, and death from community level contacts.

**Objective:** Our objective is to estimate the average probability of acquiring COVID19 infection, being hospitalized or dying from a single unprotected and substantive community level contact with an individual of unknown infection status.

**Design:** We estimate individual level probability of a contact resulting in infection using established principles of infection transmission, published data on the secondary attack rate, the current case incidence, the infectious period, the proportion of asymptomatic infection, and the case hospitalization and infection fatality ratios.

**Setting:** U.S. Counties

**Outcomes:** Probabilities of COVID19 infection, hospitalization, and death from a single unprotected person-contact in a community setting.

**Results:** Among the 100 most populous US Counties, the median probability of COVID19 infection transmission at the end of May 2020 is 1 infection per 3836 unprotected community-level contacts (Range: 626 to 31,800). For a 50 to 64 year old individual, the estimated median
probability of hospitalization is 1 hospitalization per 852,000 community level person contacts (Range: 139,000 to 7,080,000). For a 50 to 64 year old individual, the median probability of a fatality is 1 fatality per 19.1 million community-level person-contacts (Range: 3.13 million to 159,000,000 million).

**Conclusions and Relevance:** Across the country, current probabilities of infection transmission, hospitalization, and death from COVID19 vary substantially, yet severe outcomes are still rare events. Individuals may be overestimating their risks of hospitalization and death and a moderate number and frequency of community contacts is unlikely to overwhelm hospital capacity in most U.S. settings. Systematic public reporting of COVID19 case incidence and prevalence of seropositivity by age, risk group, locality and setting would improve individual-level risk estimation.
Introduction and Purpose

During epidemics, perceptions of risk are active modifiers of disease transmission, motivating, or not, protective behaviors such as hand hygiene, wearing masks and social distancing at the individual level and quarantine, travel restrictions, and restrictions on gatherings at the societal level. Novel infectious agents such as the COVID-19 virus, where the understanding of susceptibility, transmission, and lethality is immature, challenges accurate risk estimation and risk communication. In addition, the uncertainty of knowledge itself, along with the invisibility of viral transmission and the perceived lack of control over risk distort and amplify risk perception.

For COVID-19, case fatality rates, reports of the counts of cases and deaths, stories and pictures in the media, government actions, and peer behavior have all influenced the perceptions of risk to COVID-19 and human behavioral responses. Metaphors of war and declarations of emergency also influence risk perception. Under those circumstances, neither policy makers nor the general public may be interpreting risk accurately at all times during an epidemic.

Predictive models, government surveillance, and epidemiological studies have largely characterized the risk of COVID-19, in terms of aggregate outcomes such as population-level case and death counts and hospitalizations. (1) Studies have also identified risk factors for adverse outcomes from COVID-19 disease, such as age and chronic disease status. (2) We are not aware of any published research that estimates individual level probabilities of infection, hospitalization, and death.
Absent systematically collected and reported, robust case risk factors and contact tracing data, gauging relative risks among individual types of settings is speculative. Still, society has taken dramatic and unprecedented steps to control COVID-19, choosing to apply universal contact reductions through home confinement, limits on travel, closures of schools and businesses and limits on gatherings. While heightened perception of risk (fear) motivated those proscriptions on social contact at the outset of the epidemic, ongoing restrictions on community activity may be mediating ongoing risk perceptions. Today, even with transmission falling in many places and States re-opening, public sentiment surveys indicate a high level of apprehension about returning to everyday community activities. As of May 20\textsuperscript{th}, 2020, over half of the US population fears getting a haircut, going shopping, or visiting a friend. (3)

While data does not permit estimating setting specific risks of COVID-19 transmission, sufficient data do allow an estimate of the average individual-level probability of infection across all community settings. Here, we contribute to COVID-19 risk perception by estimating the individual probabilities of acquiring infection, being hospitalized, and dying from community-level contacts in large U.S. Counties. Our findings may inform both the public as well as policy makers less familiar with epidemiologic metrics. Equally important, we identify areas of available and future knowledge that could make risk assessment more precise and context specific.

**Methods and Data**
Our objective is to estimate the probability of acquiring COVID-19 infection from a contact with an individual who has an unknown infection status including a household member. We conceptualize that probability using standard epidemiological principles as the product of the proportion of the population susceptible, the probability of transmission per contact, and the opportunities for contact with individuals who are currently infectious. The latter term is the product of the prevalence of active infections, the duration of infectiousness, and the proportion of time infectious people circulate in the community, accounting for symptomatic infections that do not become detected cases, pre-symptomatic transmission and for asymptomatic infections. The formula used to compute probability is described below. Table 1 provides the parameters we use to estimate the probability of infection.

\[ P = S \beta \lambda I_{\text{Reported}} \left\{ (\sigma D_{\text{Infectious}} + \eta \lambda (1-\sigma) D_{\text{Infectious}}) \left[1 + \alpha/(1-\alpha)\right]\right\} \]

Where

- \( S \): Proportion of the population susceptible
- \( \beta \): Secondary attack rate for non-household contacts
- \( I_{\text{Reported}} \): Incidence of reported cases
- \( \lambda \): Proportion of infectious persons circulating in the community
- \( r \): Ratio of total symptomatic infections to reported cases
- \( D_{\text{Infectious}} \): Total duration of infectiousness (days)
- \( \sigma \): Proportion of infectious time pre-symptomatic
- \( \eta \): Proportion of symptomatic infections not complying with isolation
- \( \alpha \): Proportion of infections that are asymptomatic or unreported
In this risk assessment, the probability of being a reported case, being hospitalized and dying from COVID-19 disease follow fixed ratios related to the probability of infection. For example, for an individual in the 50-64 year old age group the ratios between infections; reported cases, hospitalizations, and deaths are: 10:1:0.045:0.002. (Table 1)

Community-level contact includes those that happen in households and community settings such as workplaces, and group living situations. Because public data do not yet provide data on case transmission by setting, we are unable to differentiate non-household contact from the habitual close physical contact that occurs within households and among family members, understanding that attack rates are likely larger within households and group living environments. Examples of substantive contacts outside households might include dining with a friend or business contact, working in a shared office space or having close or physical contact without the types of precautions now recommended for prevention of infection transmission (e.g. avoiding handshaking, embraces, wearing a mask or indoor ventilation). Without alternative data, we postulate that 100% of infections are among individuals in the community as a whole versus separating out the fraction that might be within isolated facilities, such as residential care facilities and detention centers.

The prevalence of susceptibility to COVID-19 is unknown. Pre-existing immunity and cross immunity is plausible but speculative. Reliable estimates for the proportion of the population who have acquired immunity is unknown. Most risk assessments have conservatively estimated the prevalence of susceptibility to be 100%. We do the same.
We acquired COVID-19 case incidence data from publicly reported statistics compiled by The New York Times. (4) Reported case rates underestimate the true incidence of infection in the community both because of unreported symptomatic infections and because of undetected asymptomatic infections. Several studies have attempted to estimate the asymptomatic fraction. (5-7) In one meta-analytic review, the proportion of asymptomatic cases ranged from 6% to 41%. (8) A weighted mean estimate from those studies suggests that about 1 out of 6 persons, or 16%, may be asymptomatic. On the other hand, seroprevalence studies capture the reported, unreported and asymptomatic fractions. Some seroprevalence studies suggest that up to 90% of infections may be unreported. (9) The US CDC recently estimated the proportion of asymptomatic proportion as 35% for planning purposes. (10) We use a proportion of 90% for infections that are unreported including both those symptomatic and asymptomatic.

The true attack rate for day-to-day community exposure is unknown but likely varies by exposure context, proximity, and duration. Further work will be needed to accurately apply setting specific attack rates. Nevertheless, for the purpose of risk assessment, we can assume an average community-wide plausible attack rate based on the range of estimates reported in published contact tracing studies. Overall, secondary attack rates in those studies range from 0.7% to 16.3%. One study in Taiwan estimated a mean attack rate of 0.7% with an attack rate of ~5% among household and non-household family contacts. (11) A Hong Kong study of the quarantined contacts of visitors from China estimated a secondary attack rate of 11.7%. (12) Two published study within China found a household attack rate of 16.3% and 11.2% respectively. (13-14) Reports of published investigations of “super spreading” events yield higher attack rates but these are not representative of typical community-level spread. (15-18) The above contact
tracing studies include widely different exposure situations; still contacts are not characterized as simple proximity with short duration, such as passing by a person on the street. Our analysis applies a secondary attack rate of 10%, acknowledging that this may overestimate the attack rate for a non-household community-level encounter.

We estimate the total duration of infectiousness as 8 days. Research suggests that individuals who develop symptoms may be infectious two to three days before the onset of symptoms. (19) We apply the US CDC’s estimates that the proportion of infectiousness before symptom onset is 40% of the total duration. (10) Conservatively, we treat asymptomatic and unreported fraction as infectious for the same duration as those with symptoms.

Because COVID-19 can result in mild illness, compliance with self-isolation affects the number of infectious people circulating with infection after symptoms develop. Current research within the context of the COVID-19 pandemic finds that compliance with isolation ranges from 57% without financial compensation to 94% with compensation. (20) Given the current US context and the availability of sick leave compensation, we assume that 75% of symptomatic individuals will voluntarily self-isolate after symptoms develop. We do not alter the duration of infectiousness for the unreported and asymptomatic fraction.

To estimate the probability of hospitalization and death we apply symptomatic case hospitalization ratios and symptomatic case fatality ratios proposed by the US CDC for the purpose of COVID-19 pandemic planning. (10) We used the estimates for scenario five (Table 1), which represents their current best estimate. The estimated fatality ratios are based on CDC’s
internal analysis of case and fatality data and are within the range of published infection fatality rates in the United States at the early part of the epidemic. For example, one study of deaths through the early part of the epidemic estimated the infection fatality ratio for symptomatic cases to be 1.3% (95% CI: 0.6% to 2.1%). (2)

To validate our estimates, we compared our estimates of the probability of hospitalization per contact in individual counties against the reported incidence of laboratory confirmed COVID-19 hospitalization from the US CDC’s COVID-NET active surveillance system. (21) COVID-NET has active hospital case surveillance in 14 sub-state regions. To make this comparison, we assumed a modest number of daily contacts equal to the number of other household members plus one.

Results

As of May 30, 2020, among the 100 most populous US Counties, the median daily case incidence is 5.92 per 100,000 (Range, 0.65 - 35). In these counties, the median probability of COVID-19 infection transmission is 1 infection per 3836 unprotected (e.g., without social distancing, wearing of masks, hand hygiene, etc.) community-level contacts (Range, 626 – 31,800). That estimate does not differentiate between contacts within and outside of households.

Figure 1 illustrates estimates of the probability of transmission of one COVID-19 infection and the probability of being a reported case per community-level person contact as a function of daily case incidence for the 100 most populous US counties during the same period.
Using age-specific case hospitalizations ratios from the CDC, we estimate that for a 50 to 64 year-old individual, the estimated median probability of hospitalization is 1 hospitalization per 852,000 community level person-contacts (Range, 139,000 – 7,080,000). We found good concordance between our estimates and US CDC hospitalization surveillance under the assumption that average daily contacts equaled the average number of other household members plus one. (eTable 1) Figure 2 shows the estimated probability of a COVID-19 hospitalization per person-contact for each age group as a function of daily case incidence. As an alternate way to illustrate risk, eFigure 1 illustrates the estimated number of COVID-19 hospitalizations per million person-contacts for the 20 most populous US counties.

Using age-specific case fatality ratios from the CDC, we estimate that for a 50 to 64 year old individual, the median probability of a fatality is 1 fatality per 19.1million community-level person-contacts (Range, 3.13 million – 159,000,000 million). Figure 4 shows the probability of a COVID-19 fatality per person contact for each age group as a function of daily case incidence.

**Discussion**

Using principles of infectious disease transmission, we provide estimates of the average individual-level probabilities of COVID-19 transmission, hospitalizations, and deaths in U.S. community settings, as a function the number of community-level contacts. The estimates are based on current reported COVID-19 case incidence in the most populous 100 US counties. Probabilities vary across a wide range reflecting varying case incidences across counties.
Our estimated probabilities may be considerably lower than popular risk perception might suggest. This may be due to challenges to perceiving risks specific to particular locations and in different time points in the epidemic. As discussed above, governmental action and media attention as well as subjective feelings of uncertainty and perceived lack of control may be influencing risk perception.

We use an array plot as an alternate way visually communicate the risk of acquiring COVID-19 infection, being hospitalized, and dying under a scenario using the example of the County of Los Angeles, with twice the current reported case incidence and 6 contacts per day for 14 days. (Figure 5) In this scenario, in a population of 10,000 50-64 year olds, we expect three COVID-19 hospitalizations. Using visualizations to communicate risk and exploring the concordance of perceived risk and risk probabilities would be an appropriate subject for further work.

The infection transmission parameters chosen are either from published data or research on COVID-19 transmission or our own conservative assumptions. Rates of hospitalization for laboratory confirmed COVID-19 disease in several US CDC active surveillance areas generally corroborate our estimates under the assumption of a modest level of social contact.

Our estimates of average probabilities may overestimate those experienced by most people in community settings. First, estimates of the secondary attack rate arise from observations at the onset of the pandemic, prior to normalization of behaviors that would reduce the risk of
transmission, such as increased hand washing, observing physical distance, forgoing physical greetings, and wearing masks.

The applied secondary attack rates do not differentiate between community and household transmission. Most studies indicate that household attack rates are higher than for contacts in the community. Systematic public reporting of anonymized contact tracing data would provide information to assess context specific attack rates.

Many infections occur within geographically and socially constrained chains of transmission, for example, within clusters of related or socially connected individuals or among those in congregate living facilities such as nursing homes. Clusters of COVID-19 have been reported associated with prisons, workers dormitories, religious services, nightclubs, schools, cruise ships, sporting events, and professional conferences. (22) Identifying and excluding all of the cases that occur within non-community congregate settings would reduce our probability estimates. Systematic public reporting of the likely setting of infection would allow for more precise estimates of the circulating infectious individuals.

We do not yet have a way to account for intra-individual variation in the secondary attack rate. With respiratory viruses, the number of secondary cases generated by each index case can vary significantly. (23-24) One recent estimate suggests that 80% of COVID-19 infections are due to a small fraction (10%) of infectious individuals. (25)
The prevalence and significance of measured antibodies to COVID-19 is not known. However, there may well be factors conferring protection to infection or responsible for differences in susceptibility. For example, cross immunity with other coronaviruses may be occurring. (26) Researchers have also observed cellular immune system responses to COVID-19 among unexposed individuals likely due to prior exposure to related coronaviruses. (27) More data will be required before adjusting risk assessment for population susceptibility or immunity.

Becoming fearful and avoiding unnecessary human contact in the setting of an uncertain and lethal epidemic threat is an expected and self-protective human behavior. Prevalent beliefs today about the sources of COVID-19 infection include ‘contact with infected persons’, ‘people coming from abroad’ or ‘foreign nationals.’ (28) Notably, many people may not consider contact with a family member to be a threat, although, the threat from a family member may be no more or less than that from a stranger.

Policy to control the COVID-19 epidemic in the United States has emphasized limiting human-to-human contact in community non-household settings (e.g. schools, workplaces, restaurants, etc), versus, for example, testing and isolating infected individuals from others. Government restrictions on gatherings with friends, school and non-essential business closures in combination with daily reports on deaths and the remaining scientific uncertainties may have re-enforced people’s beliefs about the relative unimportance of the settings of risk.

Notably, in Wuhan, the interventions that occurred before the peak of the epidemic included isolation and quarantine, mandatory mask wearing, canceled New Year’s celebrations, and the
curtailment of intra-city and intercity transportation and extension of the New Year’s holiday period. The set of interventions included notable differences from the ones applied in the western economies. (29) Other interventions, which occurred after the peak of the epidemic in Wuhan, included public screening points and mandatory isolation of the infected from family households. Analysis of the epidemic in South Korea suggests that scaled up testing and public awareness alone was sufficient alone to control the epidemic. (30) Insufficient understanding of how various countries have brought their epidemics under control also maybe an important factor influencing the perception of risk to COVID-19 in community settings in the US.

A public perception of a high level of COVID-19 transmission risk from social contact in the community now co-exists with the dominant infection control policies in the US. Reopening society will require individuals to be comfortable with their personal risk of acquiring COVID-19 infection. Evolving personal knowledge about the location and magnitude or risks along with personal assessments of knowledge uncertainties and personal control will no doubt influence people’s choices to return to public life. Estimates on the individual probabilities of infection, hospitalization and death may contribute to a more accurate risk perception.

Overall, our estimates suggest that current risks may be, at least, quantitatively lower than those perceived and a moderate number and frequency of community non-household contacts will not undermine the policy aim of preserving hospital capacity in most U.S. settings. Systematic public reporting of contact tracing data by setting, reporting of case incidence and prevalence of seropositivity by age, risk group, locality and setting, and prevalence of immunity would improve individual-level risk estimation.
Contributions: Rajiv Bhatia and Jeffrey Klausner equally contributed to the design of the study and writing of the manuscript. Rajiv Bhatia conducted the analysis.

Conflicts of interest: None

Funding: None
References


13. Wei Li, Bo Zhang, Jianhua Lu, Shihua Liu, Zhiqiang Chang, Peng Cao, Xinhua Liu, Peng Zhang, Yan Ling, Kaixiong Tao, Jianying Chen, The characteristics of household transmission of COVID-19, Clinical Infectious Diseases. doi.org/10.1093/cid/ciaa450


22. Leclerc QJ, Fuller NM, Knight LE et al. What settings have been linked to SARS-CoV-2 transmission clusters? Wellcome Open Res 2020, 5:83 doi.org/10.12688/wellcomeopenres.15889.1


https://doi.org/10.1038/nature04153


Table 1. Parameters Used For Risk Assessment

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportion of infectious individuals circulating in the community</td>
<td>100%</td>
</tr>
<tr>
<td>Proportion of the population susceptible</td>
<td>100%</td>
</tr>
<tr>
<td>Proportion of infections either asymptomatic or unreported</td>
<td>90%</td>
</tr>
<tr>
<td>Average community-level secondary attack rate</td>
<td>10%</td>
</tr>
<tr>
<td>Average days infectious per infection</td>
<td>8</td>
</tr>
<tr>
<td>Proportion of infectious days asymptomatic</td>
<td>40%</td>
</tr>
<tr>
<td>Proportion non-compliant with isolation</td>
<td>25%</td>
</tr>
<tr>
<td>Symptomatic Case Hospitalization Ratio, 0 to 49 years</td>
<td>0.017</td>
</tr>
<tr>
<td>Symptomatic Case Hospitalization Ratio, 50 to 64 years</td>
<td>0.045</td>
</tr>
<tr>
<td>Symptomatic Case Hospitalization Ratio, Over 65 years</td>
<td>0.074</td>
</tr>
<tr>
<td>Symptomatic Case Fatality Ratio, 0 to 49 years</td>
<td>0.0005</td>
</tr>
<tr>
<td>Symptomatic Case Fatality Ratio, 50 to 64 years</td>
<td>0.002</td>
</tr>
<tr>
<td>Symptomatic Case Fatality Ratio, Over 65 years</td>
<td>0.013</td>
</tr>
</tbody>
</table>
Figure 1. Risk of Covid-19 Infection Per Person-Contact (US Counties)
Figure 3. Risk of Covid-19 Fatality per Person-Contact (USA)
Figure 4. Los Angeles Covid-19 Average Community Risks at Twice Current Case Incidence and with 6 Daily Contacts for 2 Weeks

Risks for 10000 50-64 Year Olds: Deaths: 0 Hospitalized: 3 Reported Cases: 80 Infected: 755 Uninfected: 9162

- uninfected
  9162 (91.62%)