Calibrating COVID-19 Community Transmission Risk Levels to Reflect Infection Prevalence

Weihsueh A. Chiu, PhD[1]; Martial L. Ndeffo-Mbah, PhD[1,2]✉

[1]Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, Texas
[2]Department of Epidemiology and Biostatistics, Texas A&M University, College Station, Texas

✉ **Corresponding authors contact information:**
Martial Ndeffo-Mbah, PhD
Address: Veterinary & Biomedical Education Complex, 660 Raymond Stotzer Pkwy, College Station, TX 77843
Phone: +1 979 845 5646
Email: mndeffo@cvm.tamu.edu,

Running Title: COVID-19 Community transmission risk level

Keywords: COVID-19; Community transmission risk level; Reported cases; Test positivity rate;

Words count: 1497 words

Conflicts of interest: None is declared

Financial disclosure: Authors have no financial disclosures

Abstract

Many organizations, including the US Centers for Disease Control and Prevention, have developed risk indexes to help determine community transmission levels for the ongoing COVID-19 pandemic. These risk indexes are largely based on newly reported cases and percentage of positive SARS-CoV-2 diagnostic nucleic acid amplification tests, which are well-established as biased estimates of COVID-19 transmission. However, transmission risk indexes should accurately and precisely communicate community risks to decision-makers and the public. Therefore, transmission risk indexes would ideally quantify actual, and not just reported, levels of disease prevalence or incidence. Here, we develop a robust data-driven framework for determining and communicating community transmission risk levels using reported cases and test positivity. We use this framework to evaluate the previous CDC community risk level metrics that were proposed as guidelines for determining COVID-19 transmission risk at community level in the US. Using two recently developed data-driven models for COVID-19 transmission in the US to compute community-level prevalence, we show that there is substantial overlap of prevalence between the different community risk levels from the previous CDC guidelines. Using our proposed framework, we redefined the risk levels and their threshold values. We show that these threshold values would have substantially reduced the overlaps of underlying community prevalence between counties/states in different community risk levels between 3/19/2020-9/9/2021. Our study demonstrates how the previous CDC community risk level indexes could have been calibrated to infection prevalence to improve their power to accurately determine levels of COVID-19 transmission in local communities

across the US. This method can be used to inform the design of future COVID-19 transmission risk indexes.

Introduction

Many organizations have developed risk indexes to help determine community transmission levels for the ongoing COVID-19 pandemic[1-3]. The US Centers for Disease Control and Prevention's "community transmission risk level" (hereafter, "CDC risk level") was recommended for use in local public health decision-making up to March 4th,2022[4]. Such transmission risk indexes should accurately and precisely communicate community risks to decision-makers and the public. Therefore, transmission risk indexes would ideally quantify actual, and not just reported, levels of disease prevalence or incidence. However, these risk indexes are largely based on newly reported cases and percentage of positive SARS-CoV-2 diagnostic nucleic acid amplification tests, both of which are well-established as highly and heterogeneously biased estimates of COVID-19 transmission[5,6]. Reported case rate and test positivity rate have been shown to provide inaccurate estimate of the magnitude and trend of COVID-19 prevalence in the US with the inaccuracy level varying between states and over time[6]. Here, we evaluate the CDC risk level as a metric for COVID-19 community transmission risk and demonstrate how this index can be calibrated to infection prevalence and redefined to improve its power to accurately determine levels of COVID-19 transmission risk in communities across the United States.

Methods

Using reported cases and test positivity time-series data from 3/19/2020-9/9/2021[3], we determined the state, metro/metropolitan, or county risk level based on CDC risk level criteria for reported cases only, test positivity only, and combined[4]. The CDC classified transmission risk level values as Low, Moderate, or High according to the following metrics: Community transmission risk level was defined as “Low” if newly reported cases per 100,000 persons in the past 7 days were less than 10 and the percentage of positive nucleic acid amplification tests in the past 7 days was less than 5%. The transmission risk level was defined as “Moderate” if newly reported cases per 100,000 persons in the past 7 days were greater than 10 and less than 50 and the percentage of positive nucleic acid amplification tests in the past 7 days was greater than 5% and less than 8%. The transmission risk level was defined as “Substantial” if newly reported cases per 100,000 persons in the past 7 days were greater than 50 and less than 100 and the percentage of positive nucleic acid amplification tests in the past 7 days was greater than 8% and less than 10%. The transmission risk level was defined as “High” if newly reported cases per 100,000 persons in the past 7 days were greater than 100 and the percentage of positive nucleic acid amplification tests in the past 7 days was greater than 10%. Finally, If the two indicators suggested different transmission levels, the higher level was selected.

To quantify the actual daily “COVID-19 risk” in each location, we used two recently develop data-driven mathematical models of COVID-19 transmission (a semi- empirical model[6] and the

IHME COVID-19 model which is SEIR-type model[7]). Though the proposed transmission risk framework can readily calibrate risk indexes using total (reported and undetected) COVID-19 prevalence estimates, here, we illustrated it using undiagnosed COVID-19 infections prevalence to better reflect the importance of undetected cases in designing community-level COVID-19 transmission risk indexes. The two transmission models[6,7] were used to calculate the prevalence of undiagnosed COVID-19 infections (I_U) over time. Details on the two transmission models are presented in the Supplemental Materials. We chose these models because they were fitted and validated against empirical data on reported COVID-19 cases, hospitalizations, and deaths[6] and seroprevalence[7] in the US and they provide daily estimates of COVID-19 prevalence at different scale.

We develop a robust data-driven framework for determining COVID-19 community transmission risk levels. We use this framework to evaluate the CDC COVID-19 community risk level and demonstrate how it can be calibrated to infection prevalence and redefined to accurately reflect levels of COVID-19 transmission in local communities across the US.

To achieve this objective, we determined the ranges of I_U that best correspond to each CDC risk level using ordered probit ordinal regression with maximum likelihood (see **Supporting Materials** for details). We then assessed the performance of the CDC risk levels in predicting the correct I_U category, summarized in a confusion matrix showing rates of predicted (CDC) and actual (I_U) categories[8]. Next, we recalibrated the risk levels by first combining the “Moderate” and “Substantial” categories, because of their extensive I_U overlap, and then determining the optimum ranges for reported cases and test positivity for predicting I_U risk levels. We have developed a Web App of our data-driven transmission risk framework which is available at https://wchiu.shinyapps.io/CDC-Risk-Level-Recalibration-alpha

Results

Our analysis shows similar results using undiagnosed COVID-19 infections (I_U) from the semi-empirical model and the IHME model provided similar results. Figure 1A and S1A show the I_U distribution for each CDC risk level, the optimized I_U breakpoints between levels, and predictive performance. The breakpoints under the recalibrated method were much greater than the CDC risk levels because both transmission models account for undetected transmission and showed that COVID-19 cases were substantially underreported in the US[6,7]. For both overall and based on cases alone, I_U distributions overlap substantially across CDC risk levels, with the poorest performance for "Low" and "Substantial" (e.g., >40% of CDC "Low" risk levels are actually "Moderate" for I_U). Test positivity alone provides very poor discriminatory power (for all levels except "High," <20% correctly categorized). To address these overlaps, we combined "Moderate" and "Substantial" risk levels and recalibrated all the ranges for reported cases and test positivity. By reducing the cases and positivity thresholds (Tables 1 & S1), this recalibration substantially improved the ability to discriminate between risk levels while also reducing the rate at which community transmission risk is underestimated (Figure 1B and Figure S1B). However, it marginally increases risk overestimation by 2% for "Low" risk level communities (with 4.3% of predicted "Moderate" risk level communities being actually "Low" risk level under the Recalibrated risk level model and 2.4% under the Modified CDC risk level) and by 0.8% for “Moderate” risk level communities (with 11.1% of predicted “High” risk level communities being actually “Moderate” risk level under the Recalibrated risk level model and 10.3% under the Modified CDC risk level).

Discussion and Conclusions

Community transmission risk indexes for the ongoing COVID-19 pandemic are an essential input to both personal and public health decision-making with respect to individual’s mitigation actions and public health intervention measures, but substantial inconsistencies in these indexes have resulted from the lack of a reliable framework for determining and communicating transmission risk levels. Here, we develop such a framework, providing a more consistent measure of transmission risk. We show that COVID-19 transmission risk indexes such as the previous CDC community risk levels can be quantified in terms of undiagnosed infection prevalence, that risk categories should be designed to minimize their overlaps, and that case and positivity criteria can be calibrated to improve accuracy in reflecting underlying disease transmission in the regions of interest. Though our proposed model improves accuracy of community transmission risk level classification relative to the CDC transmission risk indexes, it marginally increases risk overestimation for Low and Moderate risk communities. This marginal increase of risk prediction will likely result in the misclassification of a handful of Low (Moderate) transmission risk communities as Moderate (High) transmission risk level. Community transmission risk levels are provided to public health officials and healthcare facilities to help inform COVID-19 control policies and allocation of health care resources for COVID-19 patients care [1,4]. With declining COVID-19 hospitalizations and deaths in the US, this marginal increase in risk overestimation is anticipated to have minimal impact on the healthcare system.

We developed a systematic approach to determine community transmission risk indexes for infectious diseases that are calibrated to infection prevalence and provide a more accurate and precise classification of community transmission risk levels. Because disease transmission risk is a function of both reported and undetected disease cases, our approach relies on disease transmission models' estimates of undiagnosed disease prevalence. Therefore, the performance of these transmission models would likely affect the underlying accuracy of the approach. Using transmission models whose projections that have been appropriately calibrated and validated against empirical data should help improve the accuracy of the community risk level predictions of the proposed method. Though the approach was developed for COVID-19 in the US, it is applicable to other countries and infectious diseases. But for each new setting/disease, a relevant transmission model should be used to estimate disease prevalence to evaluate corresponding breakpoint values of transmission risk indexes.

Though the CDC community transmission risk levels was recently replaced by the CDC community levels, this new metric is only a measure of the impact of COVID-19 illness on healthcare systems rather than a measure of disease transmission risk. Our proposed approach remains relevant for the design of future community transmission risk level indexes in the US or other countries. The same methodology can either be applied to other existing risk indexes, or be based on independently defined ranges of infection prevalence, to inform the design of future COVID-19 transmission risk indexes.

Data availability: Our code is available at: https://github.com/wachiuphd/COVID-19-CDC-Risk-Level-Recalibration

Authors contributions: WC and MLNM designed the project, developed the model, performed the analysis, drafted the manuscript, reviewed and approved the final version

Funding: Research reported in this publication was supported by the National Science Foundation (award no. DEB 2028632) and the National Institute of Environmental Health Sciences of the National Institutes of Health (award no. P30 ES029067). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Science Foundation or the National Institutes of Health.

Human subjects review boards: NA

References:

1. CDC COVID-19 Data Tracker: COVID-19 Integrated County View. Accessed September 17, 2021 https://covid.cdc.gov/covid-data-tracker/#county-view

2. New York Times: A Guide for COVID-19 risk in your county. Accessed September 17, 2021 https://www.nytimes.com/interactive/2021/us/covid-risk-map.html
3. Covid Act Now: U.S. COVID-19 Risk & Vaccine Tracker. Accessed September 17, 2021 https://covidactnow.org/
4. Christie A, Brooks JT, Hicks LA, Sauber-Schatz EK, Yoder JS, Honein MA, CDC COVID-19 Response Team. Guidance for Implementing COVID-19 Prevention Strategies in the Context of Varying Community Transmission Levels and Vaccination Coverage. *MMWR Morb Mortal Wkly Rep* 2021; 70:1044-1047
5. National Academies of Sciences and Medicine E. Evaluating Data Types: A Guide for Decision Makers using Data to Understand the Extent and Spread of COVID-19. Washington, DC: *The National Academies Press*; 2020. Available from : https://www.nap.edu/catalog/25826/evaluating-data-types-a-guide-for-decision-makers-using-data
6. Chiu WA, Ndeffo-Mbah ML. Using test positivity and reported case rates to estimate state-level COVID-19 prevalence and seroprevalence in the United States. *PLoS Comput Biol* 2021; 17(9): e1009374. https://doi.org/10.1371/journal.pcbi.1009374
7. IHME COVID-19 Forecasting Team. Modeling COVID-19 scenarios for the United States. *Nature Medicine.* 23 October 2020. doi:10.1038/s41591-020-1132-9.
8. Kuhn, M. Building Predictive Models in R Using the caret Package. *Journal of Statistical Software, 2008; 28*(5), 1 - 26. doi:http://dx.doi.org/10.18637/jss.v028.i05

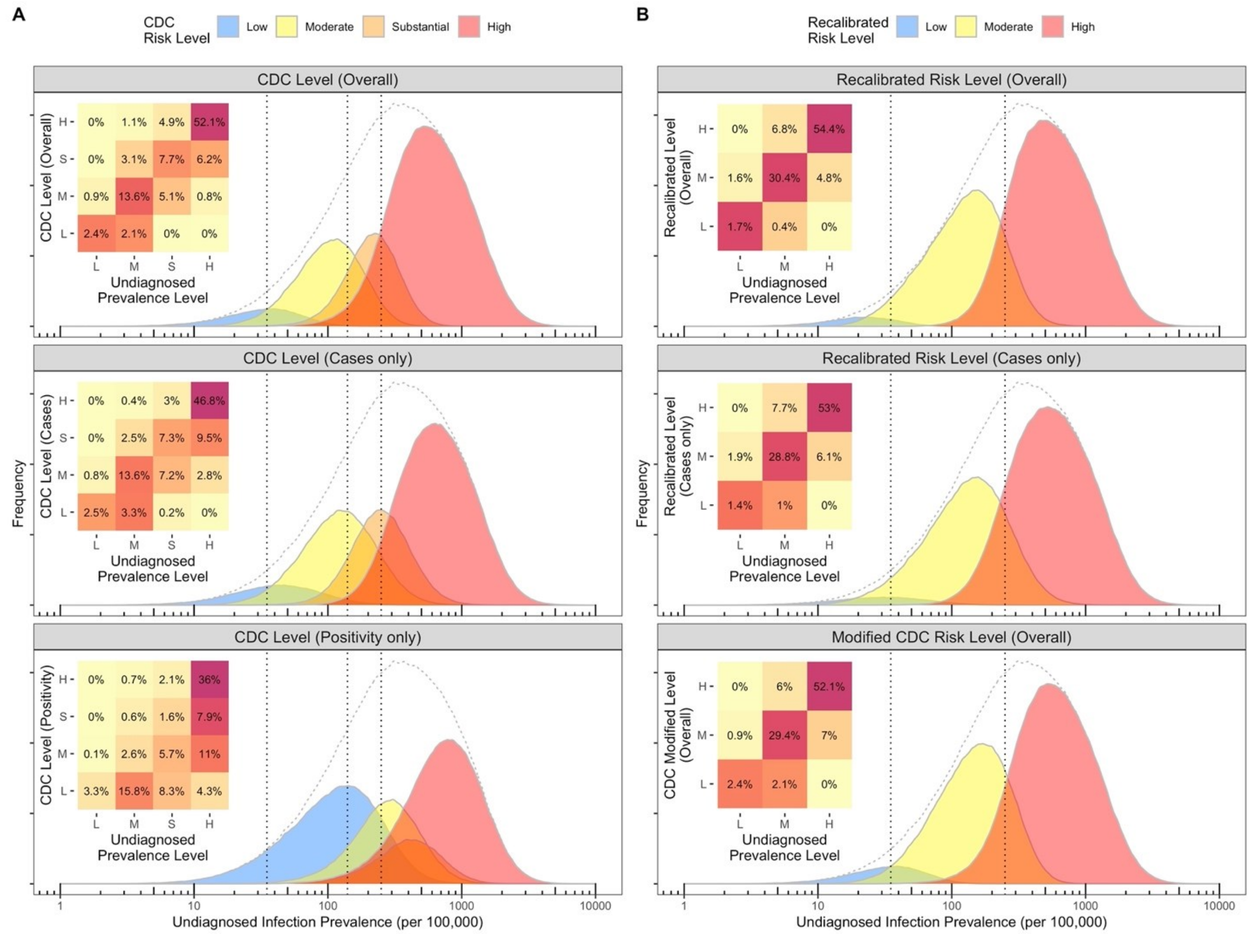

Figure 1. Comparison of performance of CDC risk levels (A: Using both cases and positivity, using cases only, using positivity only) and recalibrated risk levels (B: Using recalibrated cases and positivity criteria [Table 1], using recalibrated cases only [Table 1], using "modified" CDC criteria combining "Moderate" and "Substantial" categories with no other changes) in predicting undiagnosed infection prevalence I_U using the semi-empirical model. The frequency distribution of I_U is shown stratified by the different risk levels; dashed curve is the overall frequency distribution of I_U; dotted vertical lines are the cut-points in I_U defining the "true" categorization. The performance is summarized in terms of the "confusion matrix" which shows the "correct" categorization in each column and the "predicted" categorization in each row. Values along the diagonal are correctly predicted, values below the diagonal represent under-

predicted risk (actual risk is higher than predicted), and values above the diagonal represent over-predicted risk (actual risks are lower than predicted).

Table 1. Summary of risk level criteria based on newly reported cases per 100,000 persons and test positivity % (both during last 7 days). Recalibrated risk levels were computed using prevalence estimates from the semi-empirical model.

CDC Risk Level Criteria				
Risk level:	**Low**	**Moderate**	**Substantial**	**High**
Cases only	<10 per 100,000 persons	10 – <50 per 100,000 persons	50 – <100 per 100,000 persons	≥100 per 100,000 persons
Positivity only	<5%	5% – <8%	8% – <10%	≥10%
Recalibrated Risk Level Criteria				
Risk Level:	**Low**	**Moderate**		**High**
Undiagnosed Prevalence	<35 per 100,000 persons	35 – <250 per 100,000 persons		≥250 per 100,000 persons
Odds of Undiagnosed Infection	<1 in 2,850 people	1 in 400 to 2,850 people		≥1 in 400 people
Cases only	<5 per 100,000 persons	5 – <70 per 100,000 persons		≥70 per 100,000 persons
Low Risk Positivity	<3%	<0.5%		NA
Moderate Risk Positivity	>3%	0.5% – <10%		<4%
High Risk Positivity	NA	≥10%		≥4%