UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title

Role of environmental variability, individual behavior, and public health policy in the transmission dynamics of emerging infectious disease

Permalink

https://escholarship.org/uc/item/3fp6w185

Author

Head, Jennifer Renee

Publication Date

2022

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Role of environmental variability, individual behavior, and public health policy in the transmission dynamics of emerging infectious disease

by

Jennifer Renee Head

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Epidemiology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Justin V Remais, Co-chair
Professor Ellen Eisen, Co-chair
Professor John Taylor

Spring 2022

Abstract

Role of environmental variability, individual behavior, and public health policy in the transmission dynamics of emerging infectious disease

by

Jennifer Renee Head

Doctor of Philosophy in Epidemiology

University of California, Berkeley

Professor Justin V Remais, Co-chair; Professor Ellen Eisen, Co-chair

Emerging infectious diseases (EIDs) are infectious diseases that have newly appeared in the population, or have existed but are rapidly increasing in incidence, severity or geographic range. Emergence and re-emergence are attributed to a variety of interacting factors, including global climate change and shifting social behaviors. EIDs represent unique challenges to public health due to the unpredictable nature of epidemics, and lack of prior estimates about which public health policies best mitigate disease transmission. In this dissertation, I examine elements of EID transmission using three diseases: coccidioidomycosis, COVID-19, and hand, foot and mouth disease (HFMD). First, using distributed-lag non-linear modeling and an ensemble modeling approach, I examine how climate variability and drought have contributed to the rise of coccidioidomycosis, an emerging fungal infection, in California, and how underlying regional factors contribute to disproportionate increases in certain geographies. I find that drought temporarily displaces cases, reducing cases during drought, and amplifying cases after, and that the relative increase is highest in areas not typically considered endemic for coccidioidomycosis. The majority of EIDs are zoonotic in nature, with rodents as common reservoir species; yet it remains highly debated whether rodents serve as zoonotic reservoirs for *Coccidioides*. In Chapter 2, I describe results of an experimental field study to ascertain the role that rodents and burrows have on *Coccidioides* presence in the soil, finding that both rodents and burrows independently are associated with higher probability of detection of *Coccidioides* DNA. Overall, the presence of rodents is associated with >15 times higher odds of *Coccidioides* detection, with nearly 75% of the association mediated via burrows. Beyond coccidioidomycosis, the emergence of SARS-CoV-2 demonstrates the continued threat of emerging infectious diseases with pandemic potential, as well as the challenges in anticipating epidemic dynamics and inferring natural history parameters. Amidst uncertain epidemic predictions, transmission modeling has been a critical decision-making tool for identifying effective public health policy. In Chapter 3, I develop and apply an agent based transmission model to understand the retrospective impact of school closures on COVID-19 outcomes, and project how various within-school policies and individual behaviors will affect transmission

within schools. I document evidence of reduced social contact patterns following school closures, and estimate that spring 2020 closures of elementary schools averted 2,167 cases in the Bay Area. I also document benefits of minimizing contacts to within classroom cohorts, classroom, masking and immunization of both children and adults. Finally, interactions between pathogens may affect emergence and re-emergence of other pathogens. Serotype-specific vaccination has, in some cases, led to the emergence of non-vaccine serotypes. In Chapter 4, I estimate the early impact of vaccination of the EV71 vaccine on EV71-associated HFMD in Sichuan Province, China, and explore whether we detect significant rises in non-vaccine serotypes that may be indicative of early serotype emergence using an interrupted time series approach and a change point model. I find that introduction of the EV71 vaccine is associated with a decline of EV71-associated HFMD by 60%, but do not find evidence for substantial serotype replacement. Overall, the social, demographic, and environmental factors that influence disease emergence are expected to increase over the coming decades. Public health will need strong surveillance platforms and disease modeling tools to contend with these emerging public health threats. This dissertation demonstrates a variety of approaches towards contending with such challenges.

For Mom, Dad, Freddie, and Yuan

Contents

Acknowledgements

I am incredibly thankful to everyone who has supported me throughout graduate school. I have benefitted tremendously from the generous and kind mentorship of my PhD adviser and dissertation chair, Prof. Justin Remais. I came to UC Berkeley to pursue doctoral study in large part due to confidence in the level of mentorship I expected to receive, and in this I was not disappointed. To Justin, thank you for teaching me how to write grants, to question which critical research gaps are holding back the field, for providing resources for learning and collecting data, for gentle prodding to master more sophisticated methods, for providing new audiences to elevate our results, for connecting me to an incredibly interdisciplinary research community, for modeling a commitment to gender equality in academia, and for your effusive support that built my confidence as a public health researcher.

I am grateful that our entry into the field of coccidioidomycosis led to collaboration with Profs. John Taylor and Ellen Eisen. Thank you to John for teaching me so much about *Coccidioides*, for being a force in challenging the status quo on *Coccidioides* environmental biology, and for welcoming me into the laboratory environment. I am immensely glad that my research involved zoonotic elements, a direct result of your compelling theories. Thank you to Ellen for being an absolute joy to talk through research ideas with and your enthusiasm for our work. We have learned so much from your epidemiology expertise.

Many UC Berkeley faculty have been instrumental in advancing my research ideas, furthering my career developing, and supporting my graduate student teaching. I am grateful for the expertise and encouragement of Profs. Joseph Lewnard, Patrick Bradshaw, Sandra McCoy, Alan Hubbard, Jim Patton, and Art Reingold.

This work stems from collaborative partnerships with the California Department of Public Health, the Carrizo Plain Ecology Project (CPEP), and the China Centers for Disease Control and Prevention. Thank you to Dr. Seema Jain, Gail Sondermeyer-Cookey, and Dr. Alex Yu for taking a chance on a public health-academic partnership, for your patience in translating our results into meaningful messaging, and your support. Thank you to Prof. William (Tim) Bean for being enthusiastically willing to let us dig holes in your CPEP plots.

I acknowledge support for this work from NIH NIAID under Grant Nos. R01AI148336, R01AI125842, and F31AI152430, as well as the MIDAS COVID-19 Modeling Urgent Grant. I am particularly grateful for Dr. Alex Heaney and Kristin Andrejko for being great teammates in working together on grants.

Being able to call co-workers friends is an immense privilege. I am beyond grateful to have worked with the incredibly kind and smart team that is the Remais lab – Chris, Nick, Isabel, Qu, Alex, Phil, Mandy, Whitney, Simon, Nicole, Erika, Sophie, Katia, Eve, Catherin, and Kate, as well as other non-Remais lab members with whom I've worked closely –Kristin, Abi, Rob, and Lily. I am also grateful for my inspiring epi PhD cohort – Shelly, Stephanie, Veronica, Shalika,

Cheyenne, and Yuan. Your collective skills and knowledge expanded my own expertise, your passion for research and health disparities inspired me, and your friendship made daily work fun.

I am grateful for the members of the Berkeley community who made Berkeley feel like home. To Heather and Chris, thank you for giving Elba and I our beloved first cottage in Berkeley. To Anna Johnson, thank you for being my first California friend. To my Run Club – Sarah, Candice, Kristen, Barb, Nick, Nelson, Jeremy, Ralphie, and Yuan – thank you for giving me a sense of community. To my book club – Sarah, Michelle, Corey, Anita, Aaron, and Yuan – thank you for giving me the opportunity to learn about a great number of topics, from Berkeley-based murder mysteries to mycelial networks to the benefits of various operating systems. To my CASA youths, thank you for the weekly joys of football in the rain, slime making, and ice cream sprinkles. Chapter 3 is a result of you. Without all of you, the research may have been completed, but it would have involved a lot less joy.

I am immensely grateful for Yuan Zhu. I met Yuan Zhu on day two of being a UC Berkeley student, which means my entire doctoral program has been filled with good days, laughter, puns, and local adventures. It also means that my entire doctoral program I have had a continual supply of soap, toothpaste, and toilet paper. A great number of my research ideas were conceptualized while hiking with Yuan, and during grant writing periods, Yuan quietly ensured Elba, Mewshky, and I stayed fed, clean, and relatively stress-free. To Yuan, thank you for being my running partner, my friend, and finally, my husband. Without you, the research may have been completed, but it would have involved a lot more tears, periods of total chaos, a distinct absence of bobcats, and 75% fewer eggs.

Finally, I am grateful for the support of my family, and my friends who are like family. To Emily and Lauren, thank you for inspiring me to pursue engineering and health. To Noah, thank you for making me an aunt. To Michael and Jason, thank you for being very helpful brother-in-laws. To Erin Humphries, thank you for being my *dada rafiki*, and inspiring me since college days. To Kevin Cheung, thank you for driving with me from Atlanta to Berkeley, planning the entire trip, and ultimately joining me in the Bay Area. To Elba and Mewshky, thank you for your pure and unbridled affection, loyalty, and love of life. And finally, to mom and dad, to whom this dissertation is devoted, thank you for loving me without question. Your confidence in me has given me the courage to do so many things.

INTRODUCTION

> "Emerging infectious diseases can be defined as infections that have newly appeared in the population or have existed but are rapidly increasing in incidence or geographic range"
> - *National Institutes of Allergy and Infectious Diseases [1]*

The global problem of emerging infectious diseases

The occurrence of emerging infectious diseases (EIDs) globally has increased since the 1940's, even controlling for reporting biases. There were five times more documented EID events in the 1980's compared to the 1940's [2]. Emergence results from dynamic interactions between rapidly evolving infectious agents and changes in the environment and host behavior and health status that provide agents with favorable new ecological and immunological niches [3]. What is more, the social drivers of emergence (e.g., increasing social inequities, global travel [4]), along with demographic (e.g., aging population), immunological (e.g., rise of immunosuppressive conditions [5]), and environmental drivers (e.g., global climate change [6], population encroachment into wildlife habitats [7]) are expected to intensify in the coming decades [8].

The emergence and re-emergence of infectious diseases creates challenges for the prevention and control of disease, as shifting transmission dynamics result in unpredictable epidemics.[3] Routine surveillance is essential for rapid detection of EIDs [9], and mathematical modeling is considered a useful tool for predicting the impact of emerging epidemics [3]. In this dissertation, I leverage surveillance data and develop and apply statistical and mathematical modeling approaches to increase our understanding of major aspects of EID emergence and transmission. First, I use routine surveillance data from California to examine the role of climate variability and wildlife in the transmission dynamics of coccidioidomycosis, an emerging fungal disease in the southwestern United States. Next, I demonstrate the value of mathematical modeling in quantifying the benefit of individual and population-level public health interventions in controlling the COVID-19 epidemic. Finally, using surveillance data on hand, foot, and mouth disease, I examine whether shifts in transmission dynamics as a result of serotype-specific public health interventions could lead to unintended emergence of disease serotypes.

Climate variability and wildlife dynamics as drivers of pathogen emergence: examples using coccidioidomycosis

It is hypothesized that global climate change may contribute to disease emergence, particularly for diseases that have vectors that are sensitive to environmental conditions or have

environmental reservoirs [2, 3, 6, 10]. Changing climatic conditions can alter species range and density [6]. For instance, mosquitos may be particularly sensitive to thermal changes across a narrow range, so small shifts in temperature may yield a pronounced increase in the probability of detection of mosquitos that carry West Nile Virus within a new geographic region [11]. Climate change may also play an important role in the spillover of pathogens from zoonotic reservoirs into human hosts [6, 7]. An estimated 60.3% of EIDs are zoonoses, with over 70% originating in wildlife populations [2]. The number of EID events caused by pathogens with a wildlife origin is increasing over time [2]. Climate-induced shifts in the geographic range of pathogens may lead to novel interactions between wildlife and domestic hosts. For instance, shifts in the range of Hendra virus may have led to increases of virus among horse populations, with subsequent spillover into humans [10]. Anthropogenic environmental disturbances, such as intensification of agriculture, has also lead to spillover of pathogens – including Nipah virus and Influenza A – between wildlife, domestic animals, and human hosts [7]. In Chapters 1 and 2, I examine the role of climate variability and wildlife dynamics in the emergence and environmental epidemiology of an emerging fungal infection, coccidioidomycosis.

Coccidioidomycosis is caused by inhalation of infectious fungal spores of *Coccidioides* species in dust and can cause progressive pulmonary disease and influenza-symptoms that last months in previously healthy individuals [12]. It is a major cause of community-acquired pneumonia in the southwestern U.S., especially in the endemic areas of California's San Joaquin Valley and Central Coast [12-14]. In California, annual coccidioidomycosis cases have doubled since 2014 [15]. It remains unclear what has driven this precipitous increase, but changing climatic factors that influence the distribution of suitable *Coccidioides* habitat may play a major role [16].

Key gaps remain in our understanding of the environmental transmission of coccidioidomycosis, limiting development of mitigation strategies. These include poor understanding of how inter-annual dynamics of rainfall and drought influence pathogen growth. In **Chapter 1**, I leverage >60,000 surveillance records of all reported cases of coccidioidomycosis in the California since 2000 in order to establish a comprehensive understanding of the relationships between temperature, precipitation and coccidioidomycosis incidence. I also estimate the causal effect of the California droughts of 2007-2009 and 2012-2015 on coccidioidomycosis incidence across geographic areas, applying a non-parametric substitution-estimator (G-computation [17]) approach. This chapter fills critical knowledge gaps about the region-specific roles of seasonal climate and drought on coccidioidomycosis, and in doing so, strengthens our understanding of why California has observed substantial increase and geographic expansion of incidence over the past two decades.

Moreover, fundamental questions remain concerning the role of small mammal reservoir hosts in sustaining *Coccidioides* spp. survival and growth in the soil. The prevailing historical hypothesis is that dry, sandy soils are the primary reservoir for *Coccidioides*, with rodents as non-reservoir hosts [18]. However, there is evidence to suggest that rodents play a more substantial role in the pathogen's environmental biology and transmission. *Coccidioides* has been detected in wild rodents since 1942 [19, 20], and is found in rodent burrows (where rodents contain their dead[21]) at concentrations over four times greater than that of other

soils [22-24]. Moreover, comparative genome analyses demonstrate that the fungus evolved to obtain nutrients from animal substrates rather than plant matter [25], challenging the notion that *Coccidioides* is a soil saprobe and suggesting instead that it requires association with keratin-rich animals both during infection of a living host and after the host's death [24, 26]. If so, prevention measures and epidemiologic risk analyses could benefit substantially from consideration of rodent host abundance, distribution, and infection. **Chapter 2** described results from an experimental field study that compares the probability of detecting *Coccidioides* DNA in soils across a factorial design crossing burrows and surface soils with rodent presence and absence. The study leverages an experimental design that has excluded rodents from certain areas for the past 15 years, permitting, for the first time, disentanglement of the role of rodents and the role of burrows in the environmental distribution of *Coccidioides*.

Designing effective interventions for emerging pathogens in the midst of uncertain transmission parameters: examples using COVID-19

The emergence of the Severe Acute Respiratory Coronavirus 2 (SARS-CoV-2) demonstrates the continued threat of emerging infectious diseases globally, as well as the challenges in anticipating epidemic dynamics and inferring natural history parameters. Physical distancing measures intended to reduce close contacts between infectious and susceptible individuals have been enacted globally to mitigate the transmission of SARS-CoV-2 [27]. The San Francisco Bay Area (California) was the first region in the United States to adopt physical distancing measures on March 17, 2020, including the closure of schools and non-essential services [28]. School closures present a grave threat to healthy child development [29-31] and may exacerbate existing racial and socioeconomic gaps in school achievement [32] and nutrition [33]. As such, there is an urgent need to assess the effectiveness of school closures on reducing SARS-CoV-2 transmission and weigh risks of school reopening policies [34].

Amidst uncertain epidemic predictions, transmission modeling has been a critical decision-making tool for the COVID-19 pandemic. In **Chapter 3**, I develop and apply an agent based transmission model to understand the retrospective impact of school closures on COVID-19 outcomes, and project how various within-school policies and individual behaviors will affect transmission within schools.

Serotype-specific immunization as a potential driver of non-vaccine serotype emergence: examples using hand, foot, and mouth disease

Interactions between pathogens also affect emergence and re-emergence of other pathogens. For instance, vaccination with PCV7 has significantly reduced the burden of pneumococcal disease; however, the vaccine targets only 7 of the 92 pneumococcal serotypes, and in some populations has led to emergence of non-vaccine serotypes [35]. Similarly, the incidence of Haemophilus influenza type a meningitis increased 8-fold following introduction of Haemophilus influenza type b vaccine in Brazil [36]. In **Chapter 4**, I examine the shifting

dynamics of serotypes that cause hand, foot, and mouth disease (HFMD) in China and assess potential for serotype replacement.

HFMD causes substantial burden in the Asia-Pacific region, including China, which reported over 20 million cases between 2008-2018 [37]. The most prevalent enterovirus serotypes causing HFMD in China are EV71 and coxsackievirus A16 (CA16), with EV71 implicated in 70% of severe cases and 92% of deaths [38-41]. In 2016, a vaccine against EV71 was made available for children under five years of age. Infection with EV71, CA16, or another HFMD-association enterovirus is thought to confer lifelong serotype-specific immunity [38], as well as transient (~7 weeks) cross-serotype protection [42]. In contrast, the vaccine is not expected to confer short-term protection against non-EV71 serotypes [42, 43], which could increase risk of serotype replacement from CA16 and other HFMD-causing enteroviruses. In **Chapter 4**, I estimate the early impact of vaccination of the EV71 vaccine on EV71-associated HFMD in Sichuan Province, China, using an interrupted time series approach and a change point model. I explore whether we detect significant rises in non-vaccine serotypes that may be indicative of early serotype emergence.

CHAPTER 1. Influence of meteorological factors and drought on coccidioidomycosis incidence in California, 2000–20201

[1] Chapter 1 is included with the permission of coauthors, Gail Sondermeyer-Cooksey, Alexandra K. Heaney, Alexander T. Yu, Isabel Jones, Abinash Bhattachan, Simon Campo, Robert Wagner,, Whitney Mgbara, Sophie Phillips, Nicole Keeney, John Taylor, Ellen Eisen, Dennis P. Lettenmaier, Alan Hubbard, Gregory S. Okin, Duc J. Vugia, Seema Jain, Justin V. Remais

1.1 Abstract

Background
Coccidioidomycosis is an emerging infection in the southwestern United States. We examined the effects of precipitation and temperature on the incidence of coccidioidomycosis in California during 2000-2020, and estimated incident cases attributable to the California droughts of 2007-09 and 2012-15.

Methods
We analyzed monthly California coccidioidomycosis surveillance data from 2000–2020 at the census tract-level using generalized additive models. Models included distributed lags of precipitation and temperature within each endemic county, pooled using fixed-effects meta-analysis. An ensemble prediction algorithm of incident cases per census tract was developed to estimate the impact of drought on expected cases.

Results
Across 14 counties examined, coccidioidomycosis was strongly suppressed during, and amplified following, the 2007-2009 and 2012-2015 droughts. An estimated excess of 1,358 and 2,461 drought-attributable cases were observed in California in the two years following the 2007-2009 and 2012-2015 droughts, respectively. These post-drought excess cases more than offset the drought-attributable declines of 1,126 and 2,192 cases, respectively, that occurred during the 2007-2009 and 2012-2015 droughts. Across counties, a temperature increase from the 25th to 75th percentile (interquartile range) in the summer was associated with a doubling of incidence in the following fall (incidence rate ratio (IRR): 2.02, 95% CI: 1.84, 2.22), and a one IQR increase in precipitation in the winter was associated with 1.45 (95% CI: 1.36, 1.55) times higher incidence in the fall. The effect of winter precipitation was stronger (interaction coefficient representing ratio of IRRs: 1.36, 95% CI: 1.25, 1.48) when preceded by two dry rather than average winters. Incidence in arid lower San Joaquin Valley counties was most sensitive to winter precipitation fluctuations, while incidence in wetter coastal counties was most sensitive to summer temperature fluctuations.

Conclusions
In California, wet winters along with hot summers, particularly those following previous dry years, increased risk of coccidioidomycosis in California. Drought conditions may suppress incidence, then amplify incidence in subsequent years. With anticipated increasing frequency of drought in California, continued expansion of incidence, particularly in wetter, coastal regions, is expected.

1.2 Introduction

Coccidioidomycosis is a major cause of community-acquired pneumonia in the southwestern U.S. [12-14]. Infection can lead to a primarily respiratory illness that can last months and may progress to a chronic state in 5-10% of individuals, that can last years or be lifelong [12, 44]. Infection occurs through inhalation of spores of the soil-dwelling fungus belonging to the *Coccidioides* genus, that can become airborne through wind erosion or soil disturbance [12]. The dominant species of *Coccidioides* varies geographically, with *C. immitis* most dominant in California and *C. posadasii* in Arizona and other parts of the southwest. Coccidioidomycosis has seen dramatic increases in incidence and an expansion in geographic range over the past two decades. In California, state-wide age-adjusted incidence rates of coccidioidomycosis increased nearly 800% from 2000 to 2018, and over 300% from 2014 to 2018 [15]. The highest disease burden in California occurs in the southern San Joaquin Valley, but the largest proportional increases in incidence rates are seen among counties that lie outside this hot and arid region where cases were historically concentrated [14, 15, 45]. For instance, incidence in counties within the northern San Joaquin Valley increased by over 1,500% between 2000 and 2018, and incidence in the counties within the central coast increased 800% between 2014 and 2018 [15]. These regions are not uniformly hot and arid, with some exhibiting average precipitation twice that of the historically endemic southern San Joaquin Valley [46]. Changing climatic factors that influence the distribution of suitable *Coccidioides* habitat may play a major role in the recent rise of coccidioidomycosis in California.

Among the most concerning climatic changes recently observed in California is the increase in drought frequency and severity, a trend which may continue under anthropogenic warming [47]. A drought is defined as a period of anomalously dry conditions that results in water-related problems, and can be classified into types [48]. California experienced one of its most severe droughts in recorded history between May 2012 and October 2015, receiving less precipitation in 2013 than in any previous calendar year since records began [47, 49]. The drought caused by anomalously low precipitation during 2012-2015 was exacerbated by record high temperatures [50, 51]. The 2012-2015 drought was preceded by a less severe drought spanning March 2007 to November 2009 [52]. While public records show that statewide coccidioidomycosis incidence was lowest during drought and highest in years immediately following drought [53], the change in coccidioidomycosis incidence attributable to drought has yet to be estimated formally.

Droughts result from anomalously low precipitation and may be exacerbated by high temperatures [50, 51]. Precipitation and temperature are two climatic factors that covary with the geographic range of *Coccidioides* spp. in soil [16]. In California, precipitation and temperature are highly seasonal, with wet and cool winters, and dry and warm summers. Coccidioidomycosis incidence also tends to be seasonal, with incidence rising from spring to a peak in late fall. Periods of precipitation facilitate *Coccidioides* hyphal growth and sporulation [54, 55], and hot and dry periods cause the hyphae to autolyze, liberating infectious, heat-tolerant spores termed ‘arthroconidia’, and permitting dispersal of spores from desiccated soils via wind erosion or soil disturbance [56-60]. While prior studies generally support that the

alternating wet and dry periods enhance transmission [57, 59, 61, 62], specific details concerning these wet and dry periods—such as the timing and duration for which they are associated with amplified risk, and the magnitude of this effect—are less clear, and results have been found to vary by model structures, geographic foci, and approaches to disaggregate seasonal trends and account for lagged effects of climate [60, 63-65]. Divergence in results of prior work may stem from an underlying nonlinear relationship between coccidioidomycosis incidence and ambient temperature and precipitation, which would limit inference made from traditional linear models and prevent generalization of results to geographies with differing climates. For instance, while a wet period may support hyphal growth, *Coccidioides* is typically found in areas with <600 mm of annual precipitation, suggesting that excess moisture may limit *Coccidioides* presence [16]. Moreover, sequences of antecedent precipitation and temperature extending not only over the year prior to infection, but over several years, may regulate coccidioidomycosis transmission. Previous studies have reported an effect of seasonal precipitation on transmission of *C. posadasii* in Arizona delayed by as much as 2-3 years [57, 58, 65], but none have examined these delayed effects for *C. immitis* transmission in California, nor whether precipitation in prior years modifies the influence of more recent precipitation effects in the months leading up to transmission.

In this study, we establish a comprehensive understanding of the complex and nonlinear relationships between temperature, precipitation and coccidioidomycosis incidence, examining how these relationships vary across different time periods and geographic areas, and the degree to which the effects of intra-annual climatic factors are modified by inter-annual climatic factors. We estimate the causal effect of the California droughts of 2007-2009 and 2012-2015 on coccidioidomycosis incidence across geographic areas, applying a non-parametric substitution-estimator (G-computation [17]) approach to simultaneously describe space-varying, delayed, and nonlinear effects. By filling critical knowledge gaps about the region-specific roles of seasonal climate and drought on coccidioidomycosis, we strengthen our understanding of why California has observed substantial increase and geographic expansion of incidence over the past two decades.

1.3 Methods

Data

We obtained California Department of Public Health (CDPH) reportable disease surveillance data on confirmed coccidioidomycosis cases reported among California residents with estimated date of disease onset from April 1, 2000 – March 31, 2020. Diagnosed coccidioidomycosis cases are required to be reported by health care providers and laboratories to local health departments and then the CDPH [13, 66, 67]. Prior to January 1, 2019, a confirmed coccidioidomycosis case in California met both laboratory and clinical criteria of the Council of State and Territorial Epidemiologists (CSTE) 2011 coccidioidomycosis case definition, although case definition compliance varied by local health jurisdiction [68]. Following January 1, 2019, confirmed cases were only required to meet coccidioidomycosis laboratory criteria as defined by CDPH [69]. Per CDPH protocols, a reported coccidioidomycosis case can only be

reported once per person, so patients with reactivation of infection were included only for their initial diagnosis. Using an offline geocoding routine in ArcGIS [70], we determined the census tract of each case's residence based on reported patient address. Where street address could not be identified (11% of patients) we used the centroid of the zip code. Overall, we matched 95% of patients to a location.

Coccidioidomycosis case data was summarized as the total number of cases recorded each month, within each census tract. Environmental and climate data was also summarized as monthly vales at the census-tract level; for each census tract, we calculated the total monthly precipitation, the mean daily average temperature. We averaged time-invariant landscape features including elevation and soil texture across census tract. To maintain comparability of temporal trends with published records, patients were assigned to the month in which the surveillance record indicated estimated disease onset. However, for many patients, this was also the date of specimen collection, which may have occurred weeks following symptom onset. We obtained daily minimum and maximum temperature and daily precipitation from PRISM from 2000–2020 at 4 km resolution [46]. We obtained information on soil texture (e.g., percent sand, silt, and clay) from POLARIS, a 30-meter probabilistic soil series dataset for the contiguous United States [71]. We obtained information on the fractional coverage of impervious surface from the National Land Cover Database [72], and on elevation from United States Geologic Service (USGS) National Elevation Database [73].

We focused our analyses on geographic areas with notable incidence rates and case counts. Because the Sierra Nevada and San Emigdio-Tehachapi mountains produce strong gradients in temperature and precipitation within the boundaries of some counties with high incidence, we split the counties of Kern, Fresno, Madera, and Tulare into eastern and western regions, and Los Angeles into northern and southern regions, along a 500-meter elevation isocline. We defined the study region to include counties or sub-counties where cumulative cases exceeded 500 cases over the study period and mean annual incidence rate exceeded 10 cases per 100,000 population (Figure 1A). Fourteen counties or sub-counties were included in the final analysis (Figure 1; Table 1), which we hereinafter refer to as "counties".

Distributed lag nonlinear regression models to estimate associations between climatic factors and incidence

Associations between coccidioidomycosis incidence and temperature and precipitation were estimated using a meta-analytic approach for estimating nonlinear, delayed effects across spatial locations [74-76]. We restricted this stage of our analysis to patients with an estimated disease onset (as described by the surveillance record) between September through November—when most cases are typically reported in in California. As the effect of seasonal and lagged climatic factors may vary by season of disease onset [9,13], doing so enables more clear identification of the effect of climate in distinct seasons while improving comparability of our results with prior results. Moreover, incidence during September through November is strongly correlated with total incidence in a transmission year.

We first estimated county-specific associations of lagged monthly average temperature and total precipitation and monthly incidence using distributed-lag generalized additive models [75]. Full model details are included in the Supplement. We used monthly cases per census tract as the outcome variable and the log of each census tract's population as an offset term so that model coefficients reflected the log incidence rate ratio. The primary exposure variables were lagged total precipitation and mean temperature. These were modeled with natural cubic spline functions of smoothed three-month averages, with lags spanning 1 to 36 months prior to estimated date of onset. We also included a natural cubic spline for soil type (percent sand) and year. To determine the location of knots for the cubic spline, we systematically varied the location of internal knots placed at precipitation or temperatures corresponding to average percentiles across counties [75], selecting the model that minimized the sum of Q-AIC across all counties, where the Q-AIC is a modification of the Akaike information criterion (AIC) for quasi-likelihood models [77]. We identified the 25^{th} and 75^{th} percentiles of precipitation or temperature for the lagged months included in a 12-month cycle (e.g., a lag of one month corresponds to August - October). We calculated the incidence rate ratio per interquartile range (IQR) increase in temperature or precipitation as the incidence rate at the 75^{th} percentile compared to the incidence rate at the 25^{th} percentile, keeping other factors fixed.

Then, we used a fixed-effects meta-analysis to pool estimates of county-specific incidence rate ratios [75]. We examined the overall shape of the spline relating precipitation and temperature to incidence at each of the 36 monthly lags, and calculated the pooled IRR associated with an increase of one IQR in precipitation or temperature over each lag. We assessed factors explaining heterogeneity in the temperature-incidence and precipitation-incidence relationships across counties by extending the fixed-effects meta-analysis to include meta-predictors of county-level information (e.g., mean county total precipitation, mean county temperature). We determined the significance of the meta-predictors using Wald tests [75] and plotted separate exposure-response relationships across our set of significant meta-predictor values.

To examine potential modification of the effect of the wet winter period—the strongest predictor of fall coccidioidomycosis incidence as determined by the fixed-effects meta-analysis—by antecedent conditions, we created a nonlinear interaction term between total monthly winter precipitation per census tract and antecedent conditions by multiplying the basis function for the cubic spline on 9-month lagged precipitation (i.e. recent winter precipitation) by a binary indicator for whether or not the census tract had a drier than average winter in the two years prior to estimated date of disease onset, three years prior to date of onset, or both. The target parameter, the exponentiated coefficient on the modelled interaction term, is thus expressed as the ratio of the IRR for an IQR increase in total monthly winter precipitation following a dry year to the IRR for winter precipitation following a wet year. All statistical analyses were conducted in R, version 3.3.1 (R Foundation for Statistical Computing), using the splines and dlnm package for fitting distributed lag generalized additive models and the mvmeta package for performing fixed-effect meta-regression modeling [75, 78].

Ensemble model to estimate changes in incidence attributable to drought in California

We estimated cases attributable to—or averted because of—major droughts in California between April 1, 2000 and March 31, 2020 using an ensemble modelling approach that predicts incidence under counterfactual scenarios corresponding to the presence or absence of drought [17, 79]. For this stage of the analysis, we examined incidence throughout the entire year, rather than restricting incident cases to September through November. For each county, we modelled monthly cases at the census tract-level using parametric and non-parametric prediction algorithms (see Supplement). These included a set of generalized linear models with increasing complexity with respect to variables and interaction terms; generalized additive models; and Random Forest [80]. Model predictors varied by candidate algorithm, but could include: season; year; soil texture; elevation; percent impervious surface; total lagged monthly precipitation; and lagged mean temperature. We calculated the sum of squared errors for each algorithm using leave-out-one-year cross-validation whereby the model was fit for all but one year of the time period, and then used to predict the out-of-sample cases. We then generated an ensemble prediction by generating a weighted average of candidate model predictions where the weights were derived using non-negative least squares and were inversely associated with their out-of-sample prediction error.

Drought can be classified in different ways. Here, we examine agricultural drought, which is defined by lack of soil moisture [48], and thus thought to be most relevant for *Coccidioides*. Soil moisture varies with both precipitation and temperature, both precipitation and temperature are considered to be primary exposures of interest. We used a simple substitution estimator (G-computation [17]) to calculate the expected incident cases in census tract *i* in month *t* with covariates, $W_{i,t}$, as observed and the primary exposure, $A_{i,t}$, set to either observed or counterfactual values for lagged rainfall and temperature (equation 1). For counterfactual, "no-drought" scenario, we deterministically set any monthly average temperature higher than the historical average and any total monthly precipitation below the historical average to their monthly county-level means during the two droughts (March 2007 - November 2009; May 2012 - October 2015). We summed across specific time periods and across all census tracts in a county to estimate the number of expected cases in a county over a time period (equation 1).

$$\hat{E}(Y) = \sum_{t=t0}^{t=T} \sum_{i=1}^{N} \hat{E}(Y_{i,t} | A_{i,t} = a_{i,t}, W_{i,t}) \quad [1]$$

The incident cases attributable to—or averted by—the drought were estimated as the difference between predicted cases under the observed conditions, $\hat{E}(Y)$, and those predicted under the counterfactual "average climate" scenario, $\hat{E}(Y_0)$ (equation 2).

$$\hat{\psi} = \hat{E}(Y) - \hat{E}(Y_0) \quad [2]$$

Because antecedent conditions as far back as three years may carry influence, we examined the attributable incidence separately for the 2007-2009 and 2012-2015 droughts in the two years following the end of each drought. Because seasonality of coccidioidomycosis in California is such that incidence is lowest in March-April, with peaks occurring in the fall, we considered the change in incident cases "during drought" to include the period starting at the onset of the

drought and extending until the end of the transmission season following the drought (e.g., March 1, 2007 - March 31, 2010; May 1, 2012 - March 31, 2016). The two years post drought encompassed the full epidemiological seasons following the drought (e.g., April 1, 2010 – March 31, 2012; April 1, 2016 – March 31, 2018).

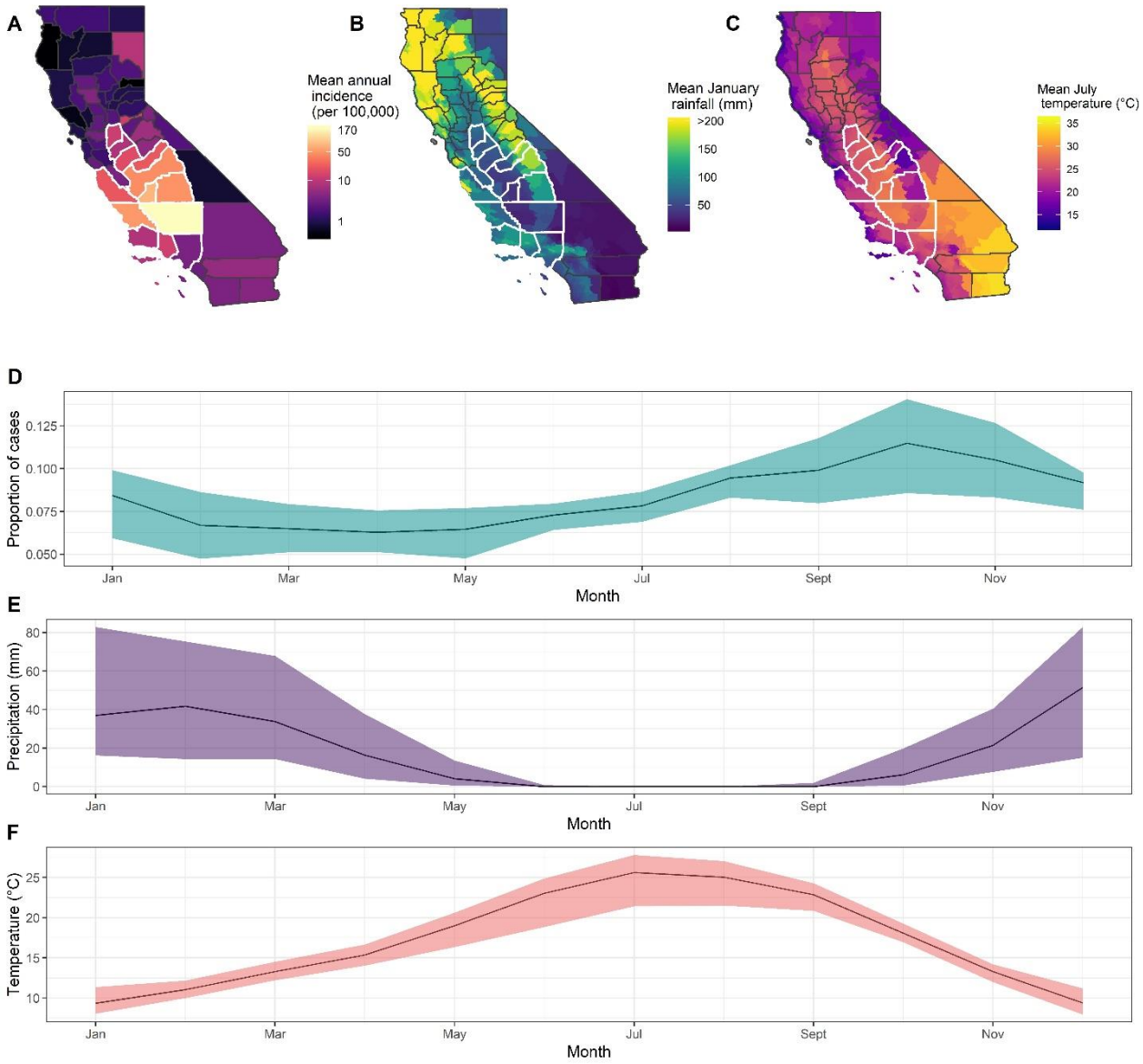


Figure 1. Average annual incidence of coccidioidomycosis (A) between 2000 and 2020. Mean total monthly precipitation (B) during January between 2000–2020. Average daily temperature (C) during July between 2000–2020. Counties outlined in white were included in analyses. Mean (dark line) and interquartile range (shaded area) of the proportion of annual cases (D) with an estimated date of onset per month, between 2000 and 2020 in the study region. Median (dark line) and interquartile range (shaded area) of precipitation (E), in millimeters, per month, between 2000 and 2020 in the study region. Median (dark line) and interquartile range (shaded area) of temperature (F), in degrees Celsius, per month, between 2000 and 2020 in the study region.

1.4 Results

Descriptive analyses

Between April 1, 2000 and March 31, 2020, there were 81,448 reported coccidioidomycosis cases throughout California reported via California Department of Public Health (CDPH) reportable disease surveillance. There were 62,002 (76.1%) cases among residents of the examined counties (Figure 1A), of which 33% of patients had an estimated onset in September – November, 25% in December – February, 24% in June – August, and 18% in March – May (Figure 1D). The counties with the highest average annual incidence rate were Kern (170 cases per 100,000 individuals), Kings (104 cases per 100,000), and San Luis Obispo (43 cases per 100,000) (Figure 1A).

Among the counties analyzed, precipitation and temperature exhibited strong seasonal patterns. From summer lows, precipitation increased starting in September and October and typically peaked between December and February (Figure 1E), exhibiting substantial spatial variation with the lowest mean annual precipitation observed among the southern San Joaquin Valley counties, and the highest in the north and west (Figure 1B). Air temperature typically peaked in July and was lowest in December and January (Figure 1C; 1F).

Association between precipitation, temperature and fall (September – November) incidence across time and space

Effect of recent precipitation and temperature (1-4 months lag)

When analyzing data on cases with estimated onset from September through November, interquartile range (IQR) increases in precipitation in the one to four months prior to the estimated date of disease onset (i.e., during the typically dry summer and early fall), were negatively associated with fall incidence (IRRs by lag: 1 month prior: 0.87, 95% CI: 0.80, 0.94; 2 prior: 0.89, 95% CI: 0.85, 0.94; 3 prior: 0.95, 95% CI: 0.92, 0.98; 4 prior: 0.87, 95% CI: 0.81, 0.94; Figure 2A; Table S1). That is, increasing precipitation in the month prior to estimated disease onset from the 25^{th} percentile to the 75^{th} percentile was associated with a 13% reduction in incidence rates in September – November. The shape of the relationship between recent precipitation and incidence (Figure 3A) indicated a suppression of incidence as precipitation increased from 0 mm (no summer precipitation) to 6 mm, with subsequent increases in precipitation exhibiting lower marginal effect. Exposure-response relationships between precipitation and incidence at all lags are shown in Figure S1.

Average daily mean temperature in the one to three months prior to estimated disease onset, during the typically hot summer months, was positively associated with coccidioidomycosis incidence (IRRs by lag: 1 month prior: 1.29, 95% CI: 1.16, 1.44; 2 prior: 1.55, 95% CI: 1.32, 1.82; 3 prior: 2.02, 95% CI: 1.84, 2.22; Figure 2B, Table S1). The exposure-response relationship at a 3-month lag (Figure 3D) showed that fall incidence increased with increasing summer temperature monotonically, with no apparent maximum beyond which temperatures are too hot. Exposure-response relationships for all lags are shown in Figure S2.

Effect of lagged precipitation and temperature (5-10 months lag)

Positive, significant pooled associations were detected between precipitation lagged 5-10 months and coccidioidomycosis incidence during September through November. The association peaked for precipitation in the winter prior to estimated date of disease onset (i.e., precipitation lagged 9 months; Figure 2A; Table S1). An increase of total monthly winter precipitation from the 25th percentile (27.1 mm) to the 75th percentile (73.2 mm) in the 9 months preceding disease onset was associated with a 45% (IRR:1.45, 95% CI: 1.36, 1.55) increase in coccidioidomycosis incidence in the fall. Pooled exposure–response relationships for both winter and spring precipitation showed a unimodal response whereby an increase in coccidioidomycosis incidence was observed with incremental increases in precipitation until an optimal value was achieved (around 40-65 mm during spring, 80-105 mm during winter; Figure 3B, 3C), after which additional precipitation was associated with lower incidence as compared to the optimal.

Higher temperatures in the winter and spring prior to estimated date of disease onset (i.e., 5-10 months prior) were associated with suppressed incidence. In pooled analyses, an increase of one IQR in average monthly temperature in the winter prior to estimated date of disease onset (from 9.5ºC to 12ºC) was associated with a 26% (IRR: 0.74, 95% CI: 0.69%, 79%) decrease in incidence rates in the fall. Fall incidence increased as winter temperatures (i.e., 9-month lagged temperatures) dropped, until around 8ºC, below which the effect of temperature was uncertain due to sparse data (Figure 3F).

Antecedent precipitation and temperature (>12 months lag)

Antecedent precipitation and temperature occurring in the two to three years prior to estimated disease onset were associated with coccidioidomycosis incidence (Figures 2B and 2C). While total monthly precipitation in the winter immediately prior to estimated date of disease onset was positively associated with incidence, precipitation in the winters two and three years prior to estimated date of disease onset were negatively associated with incidence (Figure 2A). For temperature, warmer summers and cooler springs occurring two to three years prior to disease onset were associated with higher incidence, with a dampening of this association seen as the lag increased (Figure 2B).

Antecedent conditions were important effect modifiers of the effect of more recent meteorological conditions. While precipitation in the most recent winter was the strongest predictor of coccidioidomycosis incidence when compared to other lags, precipitation in the winters 2-3 years prior was found to be a significant effect modifier of this relationship, with antecedent dry conditions amplifying the effect of wet winters on incidence. Specifically, when following a year with low winter precipitation (i.e., a year with winter precipitation falling below the 50th percentile), an IQR increase in current year winter precipitation was associated with an IRR 1.19 (95% CI: 1.10, 1.30) times larger than the IRR when the same IQR increase in current year winter precipitation was experienced following a year with high winter precipitation (i.e., a year with winter precipitation above the 50th percentile). When following two years of below median precipitation, the effect of a one IQR increase in current year winter precipitation was

1.36 (95% CI: 1.25, 1.48) times the effect of the same increase in winter precipitation following two years in which precipitation for both years was not drier than average. Put another way, the incident cases attributable to precipitation in a wet winter prior to estimated date of disease onset was highest for wet winters that followed two consecutive dry years.

Heterogeneity in effects by precipitation and temperature gradients

In multivariate meta-regression models, we found that the typical winter precipitation (i.e., the median winter precipitation) in counties explained a significant amount of heterogeneity in county-specific IRRs representing the effect of winter precipitation on incidence (Figure 4A). The effect of a one IQR increase in winter precipitation (from 27 mm to 73 mm) was most pronounced among counties where the median monthly winter precipitation was low. For instance, in western Kern, which experiences only 22.7 mm of total precipitation in a typical (50^{th} percentile) winter month, an increase from 27 to 73 mm of precipitation was associated with an IRR of 1.67 (95% CI: 1.42, 1.96). While the pooled exposure-response relationship was strongly nonlinear, very dry regions, such as western Kern county, rarely if ever exhibit winter precipitation high enough to see the maximum amplification of incidence attributable to precipitation (indicated by the 'x' in Figure 3C). As the median precipitation per county increased, the effect of precipitation on incidence attenuated (Figure 4A). Precipitation in the wettest counties, such as Monterey, which typically receives 70 mm of precipitation in a single winter month, had a non-significant negative effect on incidence. In contrast, there was little heterogeneity in the county-specific effects of summer precipitation on incidence, as most counties examined experienced little to no precipitation during these months. Thus, spatial variation in winter precipitation drives heterogeneity in the delayed effect of rainfall on coccidioidomycosis incidence, with dry counties most sensitive to fluctuations in precipitation.

For temperature, multivariate meta-regression results likewise suggested that typical summer temperature (i.e., the median summer temperature) in counties explained a significant amount of heterogeneity in county-specific effects of summer temperature on incidence (Figure 4B). The effect of a one IQR increase in summer temperature (from 20.3°C to 25.8°C) was most pronounced among counties where the median summer temperature was coolest. For instance, in Monterey, which experienced a mean monthly temperature of 16.2°C in a typical summer month, an increase from 20.3°C to 25.8°C in temperature was associated with an IRR of 12.7 (95% CI: 3.07, 53.3). An increase in mean summer temperature in the hottest 4 of 14 counties, such as western Kern, which typically experiences a mean month temperature of 27.0°C, was non-significant (IRR in Kern: 0.79, 95% CI: 0.48, 1.30). There was little heterogeneity in the county-specific effects of winter temperature on incidence. In summary, spatial variation in summer temperature drives heterogeneity in the delayed effect of temperature on coccidioidomycosis incidence, with cooler counties most sensitive to fluctuations in temperature (Figure 4B).

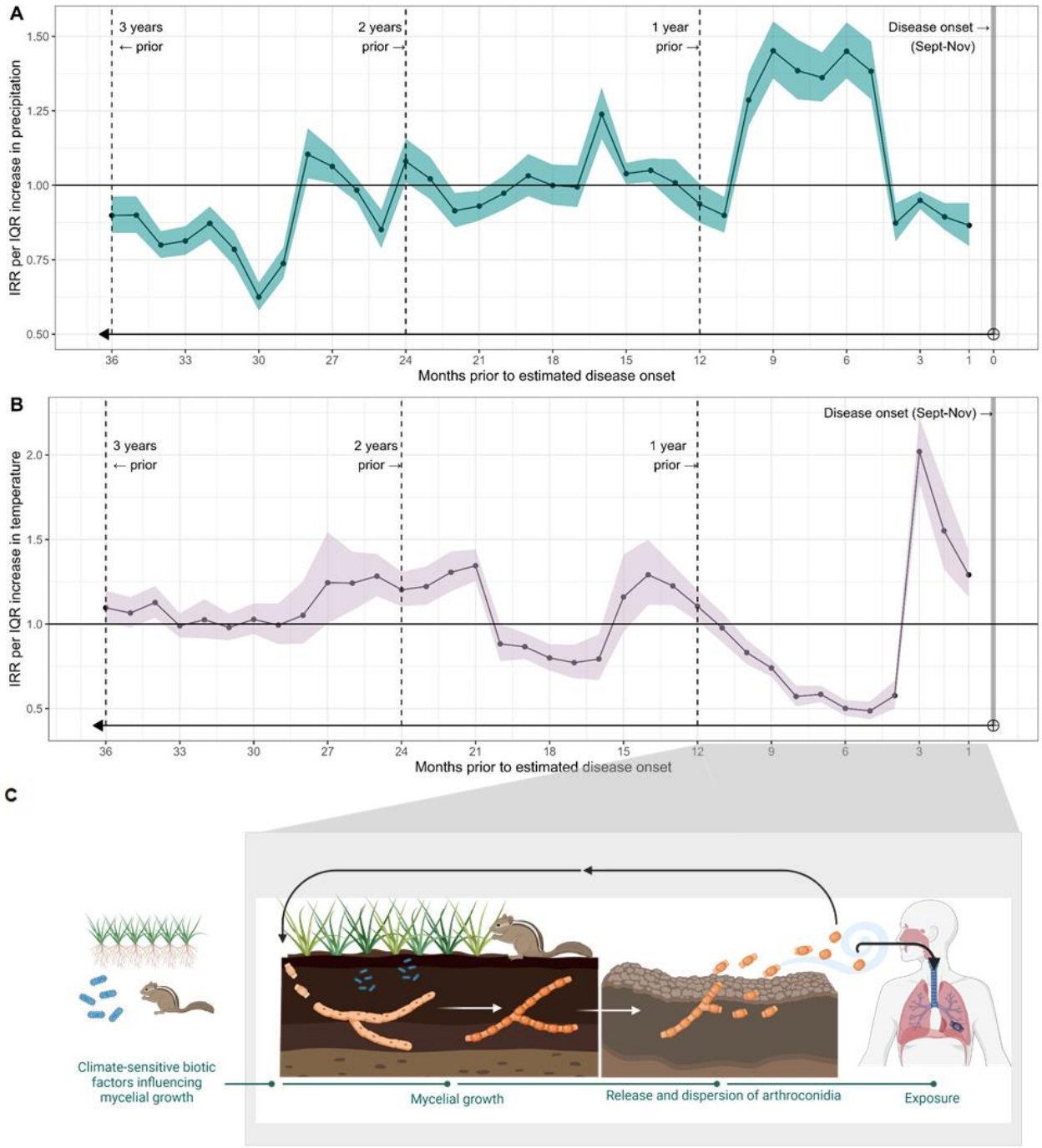


Figure 2. Results of distributed lag, generalized additive model testing the association between fall (September – November) incidence rates and lagged meteorological variables. Incidence rate ratios (IRRs) express the effect of an IQR increase in precipitation (A) or temperature (B) in months prior to the estimated date of disease onset, with confidence intervals shown by shading. The horizontal line at one indicates null association (IRR=1). Panel C displays the saprobic lifecycle of *Coccidioides* and maps the hypothesized grow and blow cycle to the intra-annual wet-dry patterns. Inter-annual influences affecting mycelial growth might include biota, small mammals, and soil dwelling microbial competitors (in blue). These factors may be influenced by climate across inter-annual time scales.

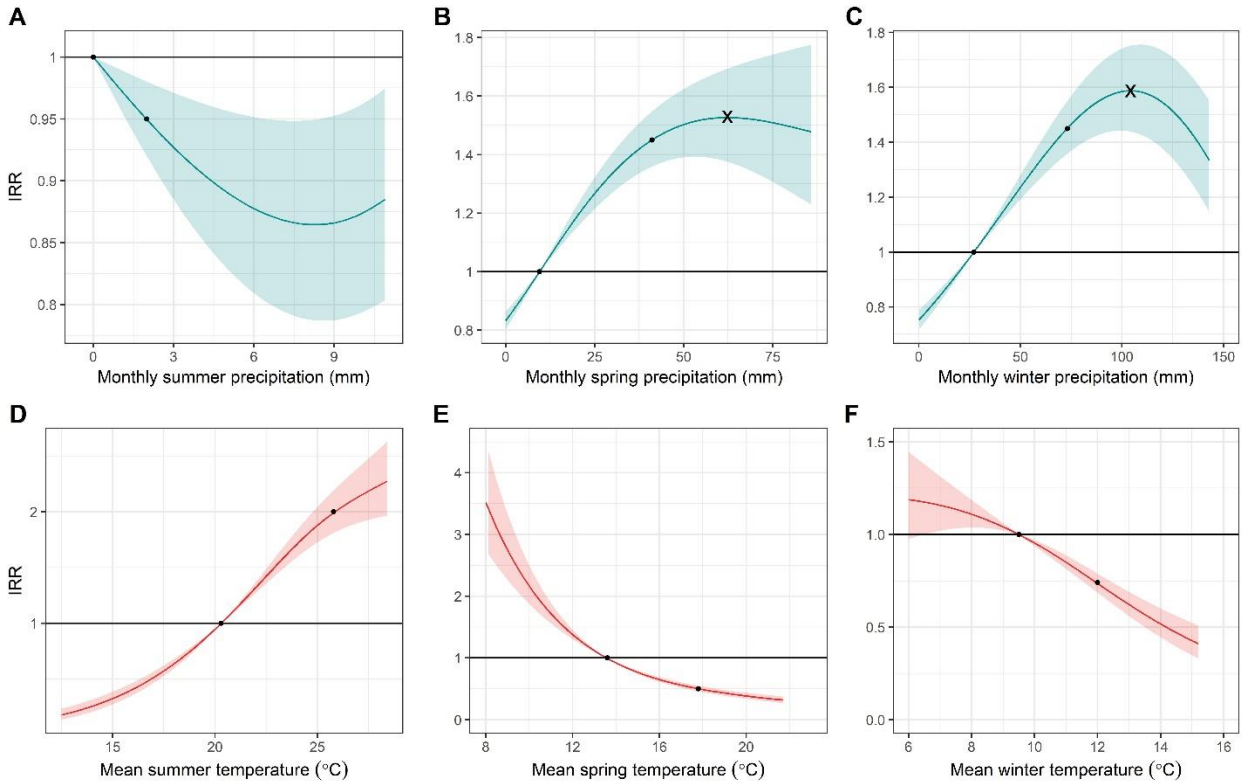


Figure 3. Pooled incidence rate ratios (IRR; solid lines) corresponding to the effect on fall (Sept – Nov) incidence of changing precipitation (A-C) and temperature (D-F) at lags corresponding to summer (A, D), spring (B, E), and winter (C, F) conditions from a reference level to the value shown. The reference level for the IRR is the 25th percentile for the study region (for which IRR=1), such that the solid curved line indicates the incidence rate ratio for fall incidence at a given temperature or precipitation value compared to the incidence rate at the 25th percentile condition for the study region in that time period. Shaded regions reflect the 95% confidence interval. Precipitation corresponding to the maximum IRR is indicated by an 'x' in panels B and C. Black circle markers indicate the IRR corresponding to the 25th and the 75th percentile of the exposure values.

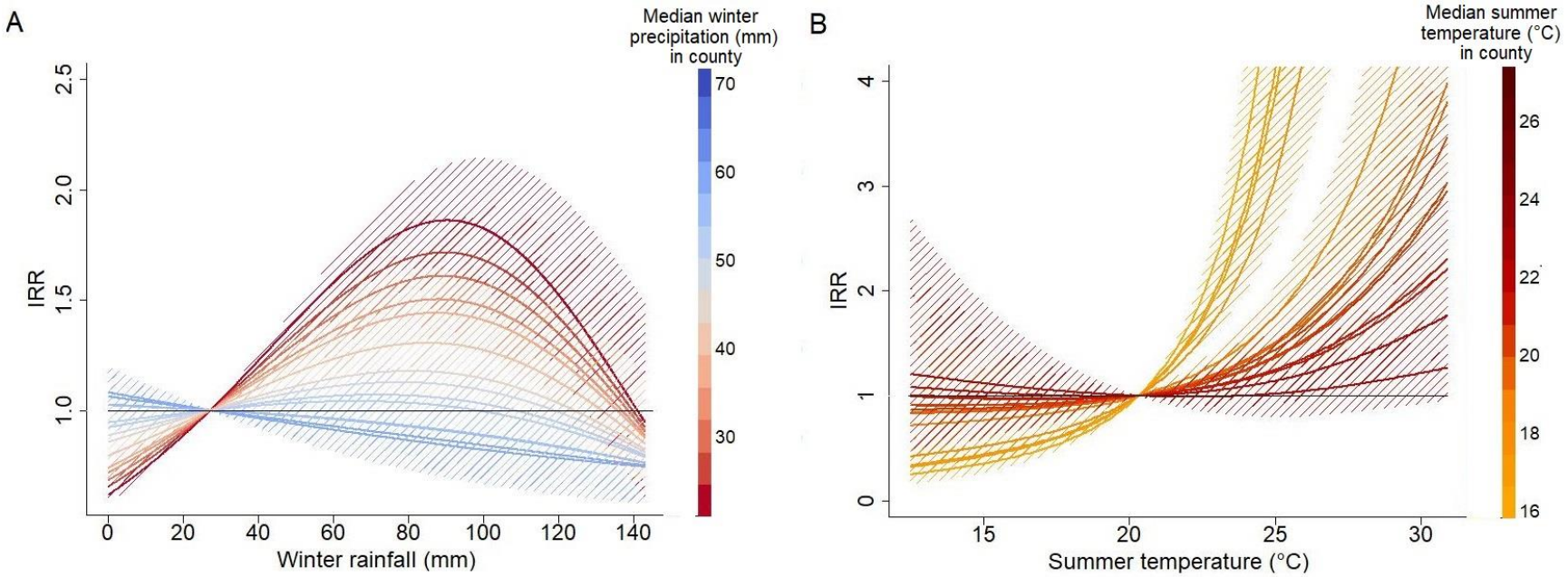


Figure 4. Increases in winter precipitation or summer temperature had the greatest effect in regions where rain was scarce or temperatures were low, respectively. (A) Estimated exposure-response relationships expressed as incidence rate ratios (IRR; colored lines) corresponding to the effect of changing winter precipitation and (B) summer temperature from a reference level to the value shown. The reference level for the IRR is the 25th percentile for the study region (for which IRR=1). Each line indicates an exposure-response relationship expressed as the incidence rate ratio for a given temperature or precipitation value compared to the incidence rate at the 25th percentile condition for a given county, based on the county's median winter precipitation (A) or median mean summer temperature (B). Dashed regions around the solid lines indicate 95% confidence intervals.

Coccidioidomycosis incidence attributable to drought in California (ensemble models)

Ensemble models were able to explain over 90% of the variation in coccidioidomycosis incidence in the study region. Counterfactual model predictions aggregated across the study period (Figure 5A; with county-specific model fits shown in Figures S5-17) revealed lower than expected transmission during periods of drought, with higher than expected transmission across the two years following (Table 1). Higher excess incidence attributable to drought were seen following the more severe 2012-2015 drought (Figure 5C) as compared to the 2007-2009 drought (Figure 5B).

In 11 of the 14 counties examined, the 2012-2015 drought was associated with a decline of cases below expected during the drought, followed by an increase in cases above expected in the two years following the drought (Table 1; Figure 6 for relative changes, Figure S4 for absolute changes). In all counties except for western Kern, the increase in cases following the drought exceeded the decrease in cases during the drought. In the counties examined, drought was associated with 2,172 fewer cases between May 1, 2012 – March 31, 2016, and 2,460 excess cases between April 1, 2016 – March 31, 2018.

Kern County west of the Sierra-Nevada Mountains has the largest burden of coccidioidomycosis in the state. Over the 47 months spanning May 1, 2012 – March 31, 2016 drought period, 3,390 cases of coccidioidomycosis were reported among residents of western Kern County. Had conditions not been drier or hotter than average (i.e., absent the drought), there would have been an estimated 1,965 more cases of coccidioidomycosis during this period. Thus, drought conditions were associated with an estimated 36.7% reduction from expected (counterfactual) incidence over May 1, 2012 – March 31, 2016 in Kern County. However, in the 24 months following the drought—when Kern County observed 6,394 cases—we estimated that if the drought had not taken place, the county would have seen 445 fewer cases. Thus, drought was associated with a 7.5% increase from expected incidence in the 24 months following the drought in Kern County.

The 2007-2009 drought followed similar patterns as the 2012-2015 drought but was associated with fewer averted cases during the drought period, while being associated with a similar number of excess cases in the two epidemiological years (April - March) that followed. As a result, this drought—which was shorter in duration and experienced lower temperatures than the later 2012-2015 drought—also resulted in a net increase in cases. Across all counties examined, the drought was associated with 1,125 fewer cases between March 1, 2007 – March 31, 2010, and 1,358 excess cases between April 1, 2010 – March 31, 2012 in the study region. The decline in cases during the drought was most prominent in western Kern County and least pronounced in the counties in the northern San Joaquin Valley, while the increase following the drought was most prominent among the coastal counties and those in the northern San Joaquin Valley (Figure 6; Figure S4). An estimated 1,145 cases were averted in western Kern county during the 37 months spanning March 1, 2007 – March 31, 2010, corresponding to a decline of 40.4% from expected. There was an estimated excess of 273 cases during the 24 months

following the drought from April 1, 2010 – March 31, 2012 in Kern County, corresponding to an increase of 6.1% from expected.

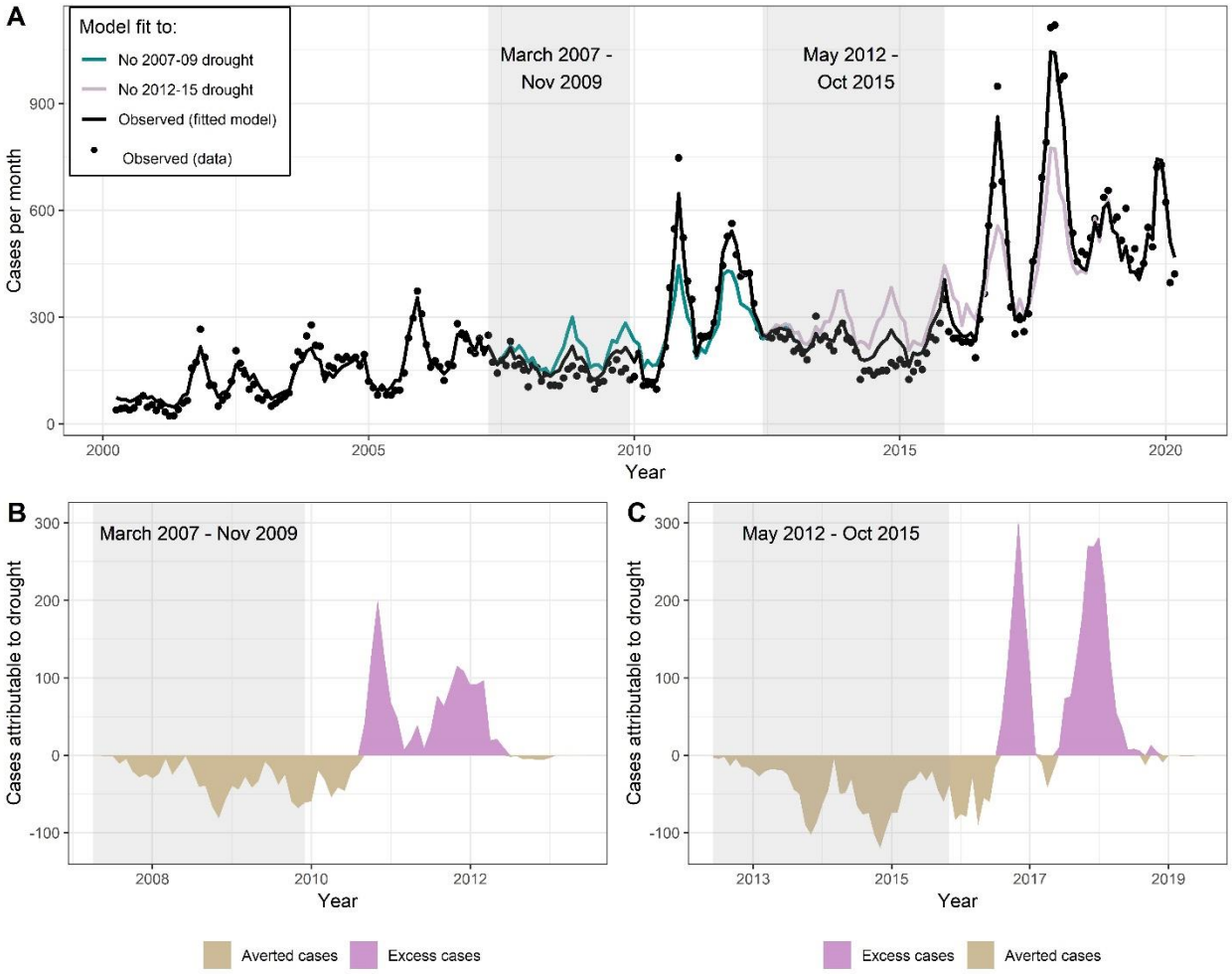


Figure 5. Droughts were associated with reduced incidence during the drought, and excess incidence following the drought. (A) Observed incidence (black dots) by month within the study region. Black line is the model fit under the observed environmental conditions. Color lines represent the expected incidence under the counterfactual intervention if the 2007-09 drought did not occur (cyan) or the 2012-15 drought did not occur (pink). Counterfactual scenarios were generated by setting temperatures observed to be higher than historical averages, and precipitation values observed to be below historical averages, deterministically to their average values. Gray boxes indicate the drought period. (B and C) Difference between expected cases and counterfactual cases if the 2007-09 (B) and the 2012-15 (C) droughts had not occurred, respectively.

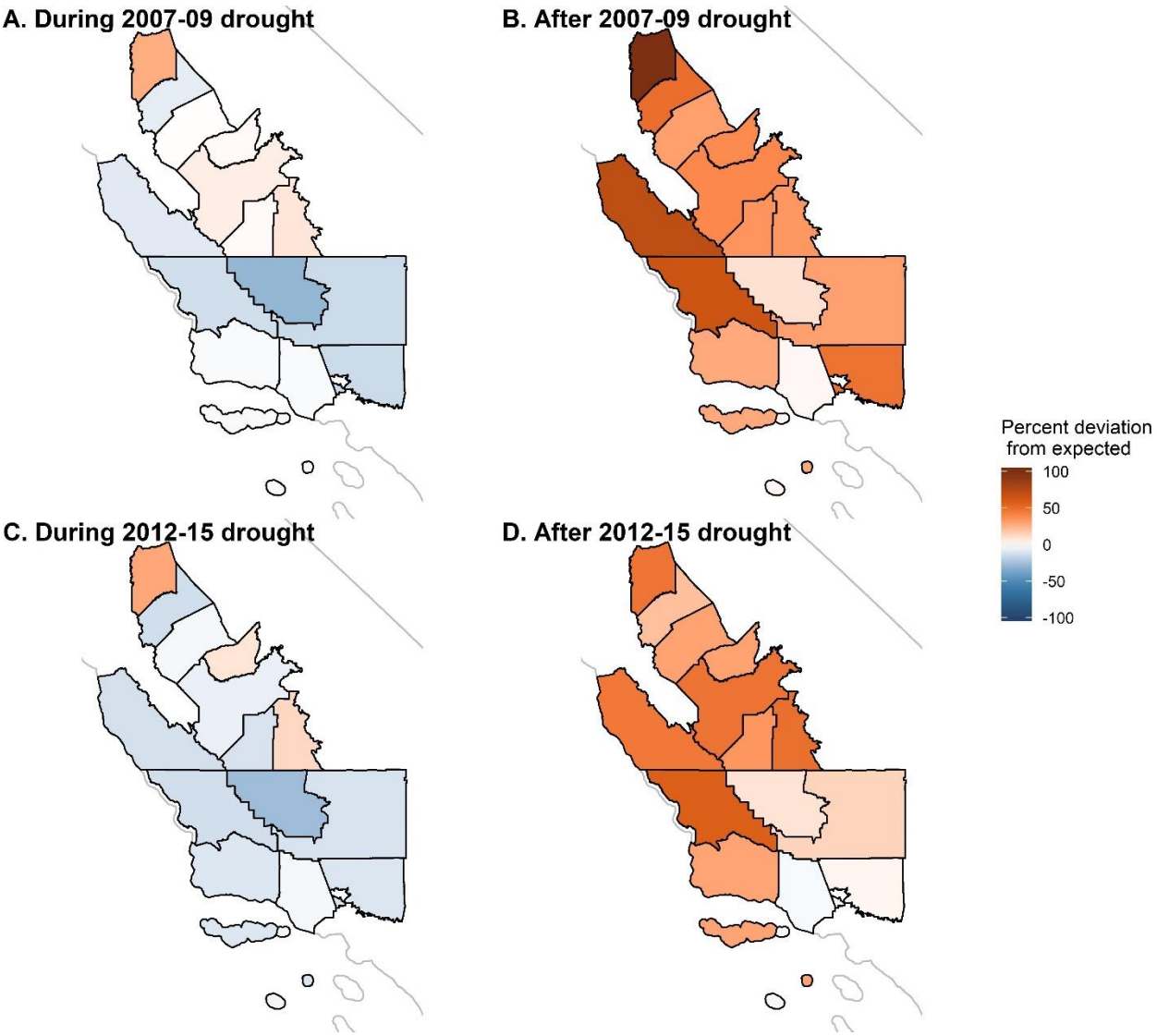


Figure 6. Estimated percent deviation in incident cases compared to the number expected in the absence of drought during (A and C) and in the two years following (B and D) the 2007-09 (A, B) and 2012-15 (C, D) droughts across the 14 counties in the study region. California state outline shown in light gray.

Table 1. Excess cases associated with California droughts spanning 2007-2009 and 2012-2015. 'Expected' represents expected incident cases if below average precipitation and hotter than average temperatures during the drought period had been deterministically set to average levels. 'Difference' represents difference between observed incidence and expected incidence, here interpreted as the incident cases attributable to the drought.

	May 2012 – October 2015 drought					
County	**During drought**		**2-yrs following**		**Total**	
	Expected	**Difference**	**Expected**	**Difference**	**Expected**	**Difference**
Western Kern	5,355	-1,965	5,949	446	11,304	-1,608
Western Fresno	1,638	-115	2,054	674	3,692	591
Western Tulare	808	92	924	315	1,732	430
San Luis Obispo	383	-68	811	301	1,195	257
Ventura	496	-8	627	-19	1,123	-17
Kings	458	-63	509	132	966	77
Monterey	308	-33	642	205	950	181
San Joaquin	376	95	556	181	932	291
Northern Los Angeles	419	-48	427	7	846	-32
Eastern Kern	370	-48	324	42	694	-7
Merced	266	-7	219	55	485	48
Stanislaus	194	-25	252	41	446	14
Santa Barbara	151	-13	234	51	384	43
West Madera	146	12	127	30	272	44
Total	11,367	-2,192	13,654	2,461	25,021	313
	March 2007 – November 2009 drought					
	During drought		**2-yrs following**		**Total**	
	Expected	**Difference**	**Expected**	**Difference**	**Expected**	**Difference**
Western Kern	2,829	-1,145	4,511	274	7,341	-960
Western Fresno	1,203	71	1,326	345	,2529	434
Western Tulare	588	45	439	106	1,027	167
San Luis Obispo	274	-44	379	144	653	114
Ventura	189	-3	143	2	332	2
Kings	295	2	410	102	705	111
Monterey	122	-9	174	67	296	60
San Joaquin	165	34	192	93	357	132
Northern Los Angeles	190	-35	222	67	412	48
Eastern Kern	198	-33	245	52	443	18
Merced	72	0	119	27	191	26
Stanislaus	77	-7	132	42	209	33
Santa Barbara	69	-3	78	17	147	18
West Madera	78	2	77	20	155	22
Total	6,349	-1,126	8,446	1,358	14,795	228

1.5 Discussion

This work finds that drought temporarily displaces coccidioidomycosis transmission, suppressing cases in years characterized by drought conditions but amplifying cases in years immediately following drought conditions. Across 14 counties examined, we estimated that an excess of 1,358 and 2,461 drought-attributable cases were reported in California following the 2007-2009 and 2012-2015 droughts, respectively, more than offsetting the drought-attributable decline of 1,126 cases during the 2007-2009 drought and 2,192 during 2012-2015 drought. This represents a previously underexplored health consequence of drought in California. Meanwhile, the impact of seasonal variation in climatic factors on coccidioidomycosis transmission, irrespective of drought, underscore this trend: in general, cases increased following wetter than average winters and hotter than average summers. The magnitude of this effect was mediated by the underlying average climate regime of a given region. Overall, our findings motivate inclusion of drought and winter rainfall in incidence projections, anticipation of different effects of drought and precipitation across geographic regions, and intensification of prevention measures in wet years that follow drought.

Our findings can be interpreted in the context of specific lifecycle stages of the fungus, as well as the major hypotheses that link coccidioidomycosis transmission to environmental processes (Figure 2C; Table 2). First, associations detected between recent climatic conditions and coccidioidomycosis incidence likely signal the suppression or amplification of mycelial fragmentation into spores, and spore dispersal, while associations detected between coccidioidomycosis incidence and climatic conditions at longer time lags (e.g., 5 months or more) likely signal mechanisms involving long-term growth of mycelia and survival of spores in the environment (Figure 2C). The “grow and blow” hypothesis is among the most widely accepted mechanistic theories regarding the influence of climate conditions occurring within the past year on seasonal coccidioidomycosis [57, 59, 61, 62], and other soil-dwelling fungal pathogens may take advantage of this strategy [81]. We support and enhance our understanding of this hypothesis by suggesting that, within an annual cycle, hyphal growth for *C. immitis* may be most important during winter and spring months when there is adequate rainfall to promote growth [55]. Pooled regression estimates show that the strongest effect of precipitation on coccidioidomycosis incidence is observed at a 9-month lag before disease onset in the fall (Figure 2A), corresponding to winter precipitation. Meanwhile, a higher temperature and lower precipitation during summer and fall was associated with increased incidence, suggesting that lysis of hyphae into singular spores and wind dispersion of spores is most important in the summer and fall.

Associations detected between coccidioidomycosis incidence and climatic conditions that occurred more than a year prior may involve the influence of upstream factors on pathogen proliferation, such as nutrient availability and presence of other soil microbes that interact with *C. immitis* (Figure 2C). Both our ensemble model and regression results suggested that an increase in the mycelial growth that occurs during wet periods is induced by prior dry conditions. Wet conditions nine months prior to disease onset yielded substantial increased incidence (Figure 2C), but the effect of this wet period was 36% stronger when it followed a

two-year dry period compared to when it did not. There are several hypotheses that may explain these acyclical inter-annual patterns observed over a multi-year time period. First, these findings could lend support to the "soil sterilization" hypothesis. Laboratory studies suggest that *Coccidioides* spp. are poor competitors for nutrients when compared with certain other soil fungi and bacteria [22, 82], but are resilient and can survive climatological extremes, including low precipitation and intense heat, to which other species may be more susceptible [83]. Thus extreme hot or dry periods may suppress the relative fitness populations of microbial competitors in the soil [57, 62, 84, 85], allowing *Coccidioides* spp. populations to grow uninhibited by competition when more favorable conditions (i.e., rainfall) return. Other plant-microbe interactions may be at play. Genomic studies of *Coccidioides* indicates a lack of plant-metabolizing enzymes, raising the possibility that *Coccidioides* lives mutualistically in soil with microbes that produce exoenzymes that permit *Coccidioides* to access nutrients [25]. Many bacteria may inhibit fungal growth [86]; bacteria such as *Bascillus subtilis* and *Streptomyces* spp. produce antimicrobial compounds that act as antifungal agents [87, 88]. Certain plants exhibit selective forces for the establishment of bacterial and other microbial populations [89].

Other explanations may exist for why antecedent dry conditions influence *Coccidioides* growth once favorable conditions return. An emerging hypothesis is that small mammals are reservoirs, harboring inactive *Coccidioides* granulomas which transform into hyphae following host death [24, 26]. *Coccidioides* has been detected in wild rodents since 1942 [19, 20], and is found in rodent burrows at concentrations over four times greater than that of other soils [22-24, 90]. Comparative genome analyses demonstrate that the fungus evolved to obtain nutrients from animal substrates more efficiently than plant matter [25]. Rodent death rate is highest during drought [91], which may lead to an accumulation of desiccated keratin in the soil that can be utilized by *Coccidioides* for hyphal growth once more favorable conditions return. In addition, physical and chemical soil properties influenced by precipitation and temperature, including soil pH, salinity, and the presence of flora and fauna, are established over long geological and ecological timescales, and may influence growth of *Coccidioides* spp. [18, 24, 54, 90]. Rodents may influence in other ways. Rodents have been shown to disperse spores of other mycorrhizal fungi [92], and may travel further during drought years to find food. Beyond biological hypotheses, human behavior is an understudied aspect of disease transmission [93]. Human adaptive behaviors to drought, like reduced in agricultural intensity, may play a role in reducing exposure to pathogens [94].

In examining how the underlying temperature and precipitation regime of a geographic region explains heterogeneity in incidence rates given climatic anomalies, we found that increasing summer temperatures were associated with especially pronounced relative increase in incidences among cooler counties as compared to already hot counties. At the same time, declines in winter precipitation were associated with strongly suppressed incidence in already very dry counties, but had a smaller effect on overall incidence in wetter counties. This finding, along with our estimates of drought-attributable cases across counties, suggests the presence of "limiting factors" in the lifecycle of *Coccidioides* that vary by region. In arid regions, limited precipitation may restrict growth. In contrast, in cooler and wetter regions, the limiting factors may be insufficient heat to lyse the mycelia into individual arthroconidia and/or desiccate the

soil to facilitate dust emissions, or excessive moisture for growth. This may explain why incidence rates have increased most dramatically in wetter and cooler counties, like the central coast counties, compared to the arid southern San Joaquin Valley counties [15].

Anticipated changes in California's climate may continue to change the spatiotemporal distribution of coccidioidomycosis in the decades to come. Average precipitation over winter months is projected to see a modest increase [95] while precipitation in autumn and spring is projected to decrease [49, 96], which may enhance conditions favorable for *Coccidioides* wet growth period and dry dispersion period. At the same time, anthropogenic climate change is expected, with medium-high confidence, to increase the duration, intensity, and frequency of temperature-driven drought in California [47, 97]. Accordingly, coccidioidomycosis incidence may continue to expand into historically wetter and cooler regions, such as coastal California counties. Future analysis should consider how these relationships resolved in this work will affect the spatiotemporal distribution of coccidioidomycosis under anticipated climate regimes in the decades to come. Increased drought is expected to dramatically change the agricultural landscape in California, as well as agricultural occupations and occupation-related mobility patterns. Future work should also consider how climate and land use interact to produce risk for *Coccidioides* transmission as well as how changes in future occupational exposures may affect incidence. We establish the individual and joint effects of temperature and precipitation, both key drivers of soil moisture, and thus factors affecting agricultural drought. Future work may apply similar approaches to understand the complex relationship between coccidioidomycosis and modelled soil moisture products [98]. Soil moisture estimates may offer a more direct measurement of the water available to soil-dwelling microbes, by accounting for irrigation and transport of snow melt. Future work may also resolve whether meteorological drought (lack of precipitation) or hydrological drought (reduced streamflow and groundwater levels) results in similar associations with coccidioidomycosis incidence [48].

The assignment of cases to census tracts of residence, which may lead to exposure misclassification, is a key limitation of the study, even as nearby census tracts, where a case may have been exposed (e.g., during work, recreation, or travel), may experience similar environmental conditions as in the census tract of residence. By using census tract-level environmental and outcome data, our study offers an improvement over prior work analyzing county-level data on minimizing statistical bias induced by aggregating heterogeneous environmental and outcome data across broad geographic regions, what is known as the modifiable areal unit problem (MAUP). However, our study is still subject to bias from aggregation of heterogeneously distributed spatial phenomenon, particularly in larger census tracts. What is more, when aggregating cases, patients were assigned to the month of their estimated date of onset, which was estimated either by the patient's own report, or, absent this, as the date of specimen collection. Therefore, the lag between a change in a climatic factor and its associated change in disease incidence also includes the incubation period for coccidioidomycosis, which varies between 7 and 21 days, and, for some patients, a lag between symptom onset and healthcare seeking, which is reported to be a median of 22 days [99], and a lag between healthcare seeking and testing. Some included cases may therefore have been exposed long before the September through November period, while some cases truly exposed

in September through November may be excluded. This exposure misclassification may bias the associations towards the null, even as we were able to detect strong associations.

Our focus on incident cases in the months of September through November, which enabled us to parse out the influence of specific timing of wet and dry periods, limited our ability to draw conclusions about how precipitation and temperature affect incidence at other times of the year. Still, September through November captures the peak transmission for coccidioidomycosis and incidence between September through November is highly correlated with the number of cases in a transmission year. Moreover, our ensemble models examined incidence throughout the entire year, and the findings align qualitatively with the regression models that were limited to incident cases in the months of September through November as both analyses found that incidence is suppressed during a transmission year with low precipitation, and a dry period before a wet period may amplify the usual transmission-enhancing effect of the wet period. Also, in modelling associations between climate variability and disease incidence, we do not control for factors that may lie on the causal pathway between temperature and incidence, such as near-surface winds or vegetation, even as they may play an important role in spore dispersal. Finally, our results pertain to California, where *C. immitis* dominates, and may not be reliably extrapolated to other endemic areas, such as Arizona, where *C. posadasii* prevails. While previous studies have largely focused on explaining the role of precipitation and temperature on the epidemiology of coccidioidomycosis in Arizona as compared to California, it remains to be seen how the effect of precipitation on coccidioidomycosis in Arizona is mediated by antecedent conditions, such as drought, or across regions with differing precipitation distributions.

While epidemiological studies are a powerful tool for relating climatic conditions to disease incidence and suggesting underling biological fungal responses driving patterns of disease, laboratory studies examining growth rate and spore production under experimentally modified moisture and temperature conditions are needed to validate the biological response of *Coccidioides* to various stimuli. Laboratory conditions may also seek to establish which soil microbes act as antagonists or symbionts for *Coccidioides*, and compare the range of conditions under which their survival is possible [88]. Large-scale field studies measuring the probability of *Coccidioides* detection under various climatic conditions can support these findings, while capturing and monitoring the complex microbial, plant, and animal diversity of the soil environment. Combinations of such studies are needed to validate or refute the “grow and blow” hypothesis, the small mammal reservoir hypothesis, or the “soil sterilization” hypothesis.

Our results offer the most in-depth treatment to date of the role of climatic factors on the transmission of coccidioidomycosis in California. Our ensemble modeling approach yielded a highly predictive model ($r^2 > 0.9$) that captured interaction between drought and non-drought conditions and nonlinear dynamics of temperature and precipitation through the inclusion of semi-parametric and machine learning algorithms, while simultaneously avoiding model overfitting through the use of leave-out-one-year cross validation and inclusion of simple, parametric models. The results provide evidence to motivate inclusion of drought monitoring and seasonal climate forecasts into coccidioidomycosis surveillance efforts and provide usable

timetables for coccidioidomycosis planning and prevention activities. For instance, prevention measures for coccidioidomycosis, including respirators for workers disturbing soil, wetting of soil before digging, and keeping car windows closed when driving through dusty areas, is needed to protect public health following droughts [100]. Increased awareness of the general public as well as health care providers of the heightened risk of coccidioidomycosis during wetter years, especially those following drought, may aid in faster diagnosis and treatment [101]. Nevertheless, prevention of coccidioidomycosis can be challenging and almost unavoidable in some circumstances (e.g., insufficient water for soil wetting, inability of wildland firefighters to wear N95s near active fires), increasing urgency to marshal resources towards ongoing vaccine development.

1.6 Conclusion

Drought may displace and amplify coccidioidomycosis incidence in subsequent years. Beyond the effect of droughts, wet winters, particularly following dry years, combined with hot summers, increase coccidioidomycosis incidence. These findings have implications for the future of coccidioidomycosis in California, where warming temperatures and increased drought may further shift the burden of coccidioidomycosis towards the wetter coastal and northern San Joaquin Valley counties.

Table 2. Summary of findings from generalized additive modeling of lagged precipitation and temperature and coccidioidomycosis incidence in the present study, and their relationship to established hypotheses and specific lifecycle components of *Coccidioides* spp.

Time period relative to estimated date of disease onset	Hypothesized lifecycle function	Hypothesis in question	Main finding on influence of meteorology	Variation in findings across space	Relationship to drought
Fall and summer (1-4 months prior to estimated date of disease onset)	Fragmentation of mycelia; dispersal and inhalation of spores via wind erosion	Blow of the 'Grow and Blow' hypothesis	High temperature is the driving factor increasing risk. Dry conditions also increase risk.	Most counties in the region examined are dry during fall and summer. Spatial variation in incidence rates is likely driven by spatial variation in temperatures, with some counties several degrees hotter than others, on average. Cooler counties may be heat-limited.	Higher temperatures during drought may increase cases, but the decline of precipitation in spring and winter seen during droughts seems to overcome this effect.
Spring and winter (5-10 months prior to estimated date of disease onset)	Growth of mycelia	Grow of the 'Grow and Blow' hypothesis	Precipitation is the driving factor increasing risk, with unimodal relationship observed. Cooler temperatures also increase risk.	Spatial variation in spring and winter precipitation is observed. Precipitation in dry counties are moisture-limited, and never exceed the maximum threshold beyond which additional precipitation suppresses risk. Precipitation in wet counties may surpass this threshold. We call these counties "unsaturated" and "oversaturated".	Low precipitation during drought may be driving the decline in cases seen among most regions during drought.
Two to three years prior to estimated date of disease onset	Growth and survival of spores and mycelia; regulation of upstream factors including: other microbial competitors, soil stabilizing vegetation, nutrients such as small mammals and vegetation	Soil sterilization	Antecedent climate modifies the effect of recent climate. The increased risk due to recent winter precipitation is most pronounced following consecutive years of drier than average conditions	Not examined	Incidence following drought is higher than expected.

CHAPTER 2. Role of rodents, burrows, and soil conditions on the presence of *Coccidioides immitis*, the causal fungal agent of coccidioidomycosis, in soils: experimental evidence from the Carrizo Plain

2.1 Abstract

Background

Researchers have demonstrated higher probability of detection for *Coccidioides*, the fungal causal agent for coccidioidomycosis, in burrows compared to surface soils. Burrows are subterranean areas where rodents may live and die. Burrows may modify soil conditions, like moisture and temperature, creating suitable microhabitats. On the other hand, rodents are known hosts for *Coccidioides* and the fungus can degrade keratin, which has raised a new hypothesis that rodents may be a zoonotic reservoir, releasing spores contained in lung granules upon death. The Carrizo Plain National Monument is located in a highly endemic region for coccidioidomycosis and home to a large population of burrowing mammals. Since 2007, experimental rodent exclosures have excluded rodents from certain areas, permitting for the first time disentanglement of the effect of rodents on *Coccidioides* from the effect of burrows on *Coccidioides*.

Methods

We collected 250 samples of soil across a factorial design that crossed burrows and surface soils with exclosures and non-exclosures. We collected information on soil moisture, temperature, and vegetation, and analyzed soil for presence of *C. immitis* DNA. We conducted a causal mediation analysis with g-computation to identify the total, direct, and indirect associations between rodents and *Coccidioides*, as mediated by burrows and soil conditions.

Results

Coccidioides was detected in 31% (31/100) of the burrow samples taken from outside rodent exclosures, and 23.4% (11/50) of the burrow samples collected from within the exclosures. Only six percent (3/50) of the surface soils samples where rodents were present had detectable *Coccidioides* DNA, and none of the surface soils samples taken inside the exclosures. Assuming that rodent presence is necessary for burrows under non-experimental conditions, we estimated that the odds of detecting *Coccidioides* when rodents are present is 15.3 times higher (95% CI: 8.9, 28.5) than when rodents are absent, and that that 74.1% (95% CI: 66.1, 82.3) of this association is mediated via burrows. Controlling for soil conditions, we estimated that the odds of detecting *Coccidioides* when rodents are present is 7.2 times higher (95% CI: 4.6, 12.5) than when rodents are absent, and that 54.3% (95% CI: 32.9, 57.6%) of this effect is mediated via burrows. Burrows and rodents are associated with lower soil moisture, and rodents are associated with lower vegetation. In generalized linear mixed models, lower soil moisture and higher vegetation were associated with greater odds of *Coccidioides.*

Conclusions

The notion that *Coccidioides* exists purely as a soil saprophyte is being challenged. Our study provides support for the endozoan small mammal hypothesis, while also pointing to burrows as suitable microhabitats, even in absence of rodents.

2.2 Introduction

Coccidioidomycosis is a respiratory disease caused by inhalation of infectious fungal spores of *Coccidioides immitis* or *Coccidioides posadasii*, which may be dispersed via dust or soil disruption [12]. It can infect mammalian hosts, including humans, dogs, cats, armadillos [102], bats [103], and burrowing rodents [19, 20]. *Coccidioides* has historically been recognized as a soil saprophyte, meaning that it derives its nutrients from the decaying organic matter, such as plants, in the soil [104]. However, the role that soil and vegetation play in providing nutrients for fungal growth remains unclear. One study found no correlation between vegetation type and soil pathogen presence [54], while research using remotely sensed information found that vegetation type was associated with *C. immitis* presence [105].

Researchers attempting to describe the ecological niche of *Coccidioides* by sampling for the pathogen in the soil have been limited by their ability to locate the fungus even in highly endemic areas, describing its distribution as "sporadic" [54]. Studies in California and Arizona show that *Coccidioides* spp. prefer alkaline, loamy, sandy soils that experience high temperature [54, 104, 106], which agrees with the classification of arid and hot regions as endemic coccidioidomycosis areas [14, 45]. Notably, the concentration of *Coccidioides* in rodent burrows has been found to be at least four times as high as in other surface or sub-surface soils Comparative genome analyses of *Coccidioides* and relatives that are, or are not, associated with animals demonstrate in *Coccidioides* and its family (Onygenaceae) many more copies of genes that encode enzymes that degrade animal protein and many fewer or no copies of genes that encode enzymes that deconstruct plant cell walls. These genomic changes suggest that the fungus has evolved to obtain nutrients from keratin-containing animal substrates rather than from plant matter [25]. If this is the case, keratin from rodents as well as their skin and fur in and around burrows – subterranean areas where rodents live and wall off their dead [21] – may provide nutrients to promote or sustain growth [24].

Coccidioides has been detected in living wild rodents since the 1940's [19, 20]. Emmons was the first to isolate the pathogen from rodents in 1940, and found a prevalence of 17% of the 29 kangaroo rats (*Dipodomys merriami*) and 15% of the 124 pocket mice (*Chaetodipus penicillatus*) surveyed [20]. Wild rodents infected with *Coccidioides* do not exhibit symptoms or disease unless exposed to concentrations of arthroconidia higher than those commonly occurring naturally [107]. Most recently, a study in Mexico published in 2014 analyzed 40 specimens across four rodent species for evidence of coccidioidal antibodies in serum, and discovered that 7.5% (two deer mice and one woodrat) indicated infection with *Coccidioides* [108]. These data, when coupled with evidence of keratin degradation, have led to the controversial hypothesis that *Coccidioides* spp. are not pure soil saprophytes but are also associated with dead animal hosts [24, 26, 109]. Rodents and other small mammals may serve as reservoir hosts of *Coccidioides* spp., whereby fungal spores live inside host lungs without causing disease, walled off by granules in the lung. When the hosts die, the granules dissolve, releasing the endospores into the environment, where they can exploit the decaying carcass as a substrate to grow hyphae and sporulate [26].

At the same time, there is active debate over whether the higher probability of *Coccidioides* detection in burrows is attributable to associations with mammalian hosts (e.g., fungus replicates in lungs of mammals, and uses mammals as a substrate upon death [26]) or due to beneficial conditions that burrow soils themselves exhibit (e.g., burrows may induce more suitable soil environments via changes to moisture and temperature [18, 24]). While *C. immitis* is generally found in hot and arid environments, its hyphal development is thought to be promoted by periods of relatively higher moisture, and cooler temperatures, similar to other filamentous fungi [57, 59, 110]. Rodent burrows experience fewer fluctuations in temperature than are experienced in ambient air or at the soil surface, and fluctuations usually lag those of ambient temperature [111, 112]. Many studies have observed that burrows decrease soil bulk density, which leads to increased infiltration rates and thus lower water retention [113-115]. Rodents also modify vegetation levels, which affects soil moisture as higher vegetation promotes moisture retention by improving soil aggregate stability, and reducing evaporation or runoff. However, it is unclear whether rodents reduce or enhance vegetation. On one hand, grass-eating rodents remove the vegetation around their burrows, leading to lower water retention [116]. By other accounts, higher biomass and plant diversity have been observed on burrows, likely due to greater nitrate availability [114, 117].

The objectives of this study are to elucidate how burrows, rodents, and other soil properties are associated with *C. immitis* presence in the soil. We examine the overall relationship between rodents, burrows, and *Coccidiodes* by comparing the probability of *C. immitis* detection between areas with no rodent activity and areas with known rodent activity. Using mediation analysis, we determine the degree to which associations between rodents and *C. immitis* are mediated by burrows. We leverage a 15-year experimental design in which rodents were removed from certain burrow-containing areas, and fenced exclosures prevented their return. This design permits, for the first time, disentanglement of the effect of burrows and rodents.

2.3 Methods

Study site

We collected soils from the Carrizo Plain National Monument, in San Luis Obispo County, California. The Carrizo Plain National Monument is the largest intact desert grassland within California [116]. It is home to at least 30 endangered and protected species, including several species of kangaroo rat that have previously been identified as possible hosts for *Coccidiodies* [19, 20]. The largest, the Giant kangaroo rat, or *Dipodomys ingens*, is considered an ecosystem engineer because they construct burrows that are utilized by other burrowing animals. Giant kangaroo rats build small colonies of burrows, known as precincts, which form a distinctive mound on landscapes (Figure 1) [118]. The Carrizo Plain National Monument, as well as adjacent areas [82], are highly endemic for *Coccidioides* and have been the source of numerous coccidioidomycosis outbreaks among researchers (*personal* correspondence, Gail Sondermeyer-Cooksey; *personal* correspondence, William Bean).

The research was conducted at ten experimental plots developed as part of the Carrizo Plain Ecology Project (CPEP), which was launched in 2007 to study the effect of Giant kangaroo rats

on vegetation, among other questions [116]. Each plot is 140 x 140 meters, and contains a 20 m x 20 m rodent exclosure in the middle of the plot (Figures 1 and 2). Exclosures were constructed from 1cm mesh hardware cloth and secured with rebar and T-posts. Hardware cloth extended 0.6 m below ground to prevent kangaroo rats from burrowing under, and 0.9 m above ground, with a 0.15 m overhang, to prevent kangaroo rats from climbing over [113]. The exclosure encompasses old precincts that have been absent of significant rodent activity for at most 15 years. Breaching of the exclosure by rodents – typically by the smaller San Joaquin antelope squirrel – occurs occasionally; each spring, rodents that have breached the exclosure are trapped and relocated outside the exclosure (*personal* communication: William Bean).

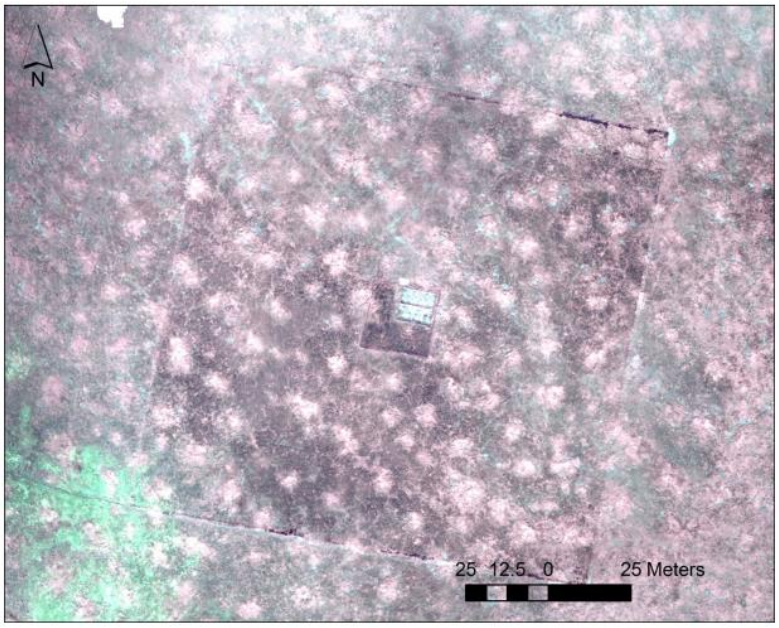


Figure 7. Satellite image of a representative experimental plot. Precincts are visible as lighter colored circles on the landscape, owing to the clipping of vegetation by rodents around their precinct. The square in the middle is the exclosure, and the larger scale delineates the boundaries of the experimental plot. Used, with permission, from prior CPEP reports [116].

Environmental sampling

In April 2021, we collected soil samples (n = 250) across a factorial design of burrows and surface soils crossed with exclosures & non-exclosures, for a total of four treatment combinations. We selected five clusters per plot: paired burrow and surface soil clusters within the exclosure; paired burrow and surface soil clusters outside the exclosure, and an unpaired burrow cluster (see Figure 2). We collected five soil samples per cluster, for a total of 25 samples per plot, and 250 samples across all ten plots. Surface soil was collected by digging a hole 10 cm deep with a trowel and sampling from the top 10 centimeters. Multiple burrow entrances were present for each precinct. Soils from precincts were taken from the mouths of five burrows, up to a depth of 24 centimeters.

Surface soils were sampled in a linear transect that radiated outward from the precinct, with the first sample 2 meters from the precinct, and the remaining four samples spaced 1 meter apart (Figure 2). Within the exclosure, burrows that seemed active or new due to potential breach were not sampled, nor were burrows that were within 2 meters of the exclosure fence. Outside the exclosure, preference was given to burrows within the selected precinct that had active markings, including tail marks, feet marks, fecal pellets, seed coats, or partially consumed seeds.

To select the two precincts outside the exclosure, we divided the plot into a 7 x 7 grid, excluding grid cells contiguous with the exclosure, for a total of 40 eligible cells (Figure 2). Using a simple random number generator, we selected two of the grid cells. If the grid cells were contiguous, we re-sampled. From each selected grid, we located the closest precinct that had signs of rodent activity, ensuring we were not moving into a contiguous grid cell of the other selected precinct. The precinct furthest from the exclosure was paired with surface soil samples. The precinct within the exclosure was selected as that previously identified by the CPEP study.

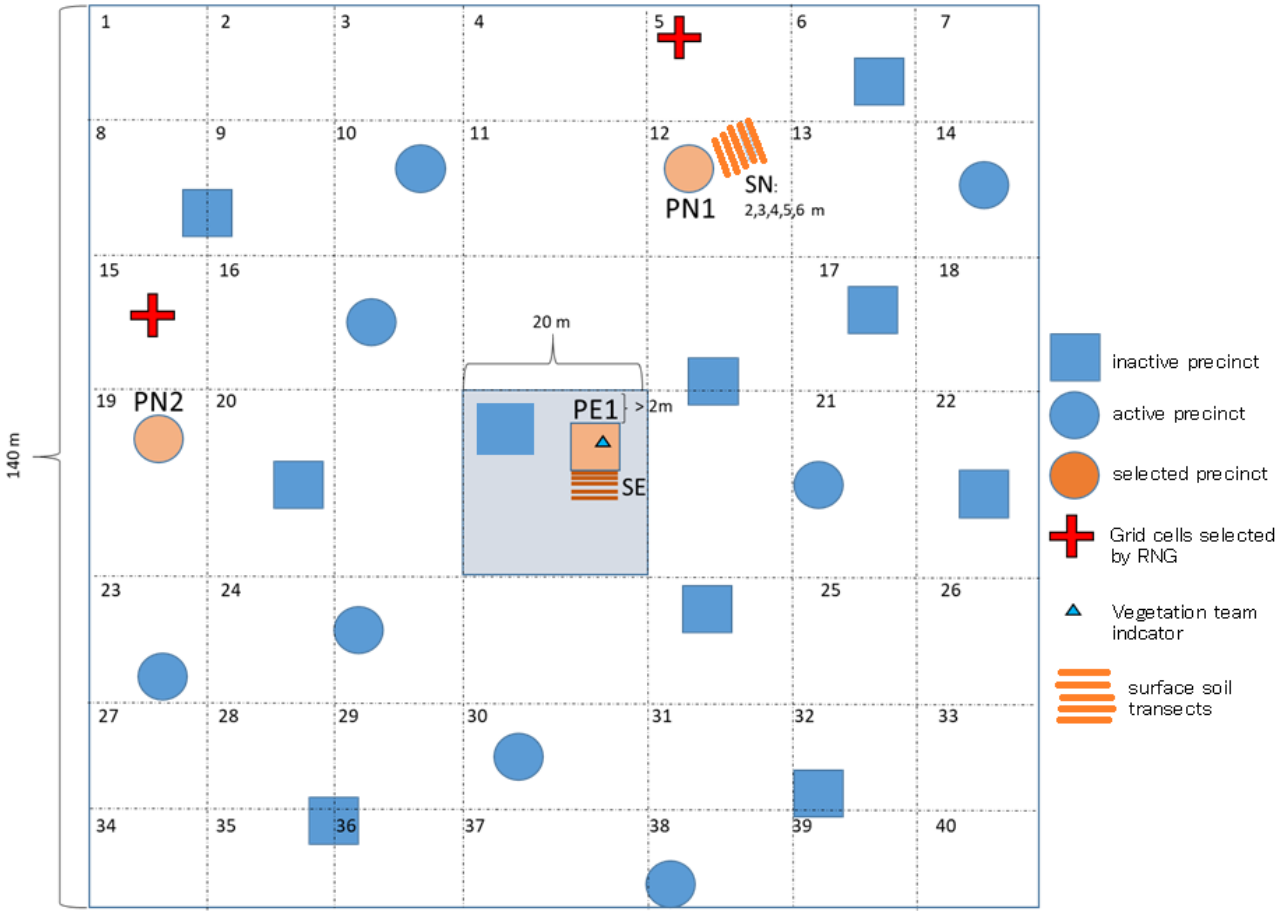


Figure 8. Diagram of an experimental plot. Each plot is 140m x 140m with a 20 x 20 m rodent exclosure in the middle. Two precincts from outside the exclosure were sampled and one precinct from within were sampled. One precinct in each the exclosure and the non-exclosure had a paired surface soil

sample, where samples were taken on a transect 2,3,4,5,6 m from the precinct. To select the precincts outside the exclosure, the plot was divided into grids, and two grid cells were selected using a random number generator. The nearest active precinct to the selected grid was taken. Cluster notation: P = precinct; S = surface; N = non-exclosure; E = exclosure; 1 = paired, 2 = unpaired.

In addition to soil, we collected information on soil moisture (measured as percent volumetric water content; Campbell Scientific® HS2 probe) soil temperature, and vegetation (type: barren/grass/shrub; level: none = 0/low = 1/medium = 2/high = 3). Data were collected using REDcap® mobile survey platform (UCSF). We also collected continuous measurement of temperature across the factorial design using temperature loggers placed within the burrows or buried 10 cm beneath the surface (HOBO® MX2202).

Laboratory analysis of soils

Soil was stored in 50-mL centrifuge tubes at room temperature until analysis. Following prior protocol, DNA was extracted using the Qiagen PowerSoil kit, and the CocciENV qPCR assay was used to determine presence of *C. immitis* in the sample [82]. Four replicates of each DNA aliquot were run, and a sample was considered positive if it was detected in at least 3 of the 4 replicates.

Statistical mediation analysis

We conducted a causal mediation analysis with g-computation to ascertain the association between rodent presence, burrows, and *Coccidioides* in soils, following the framework developed by Wang and Arah [119]. First, we posited a directed acyclic graph (DAG) to represent the data generating process. Our DAGs are shown in Figure 3. In one DAG, we include only rodents and burrows (Figure 3A), and in the other (Figure 3B), we include soil conditions, to demonstrate that some of the effect of rodents and burrows may be via soil conditions, including soil moisture, temperature, and vegetation. Rodents are considered the binary exposure of interest (X), burrows are the binary mediator (M), and *Coccidioides* presence in soil is the binary outcome (Y). Where used, soil conditions are considered the set of covariates ($\boldsymbol{Z}$). First, we modelled the mediator on the exposure and covariates, $E(M|X = x, \boldsymbol{Z} = \boldsymbol{z})$. While the experimental set up permits $M = 1$ under the scenario where $X = 0$ (i.e., burrows exist where rodents do not), this does not reflect real life conditions in the Carrizo Plain. Thus, we deterministically set values for this equation rather than modeled it using our collected data. Assuming the distribution of burrows are independent of covariates, that a rodent is required to form a burrow, and that rodents can be present in areas with both burrows and surface soils, we set $E(M|X = 0) = m_{x0} = 0$ (no rodent leads to no burrow) and $E(M|X = 1) = m_{x1} = 0.5$ (rodent leads to equal probability of burrow or surface soil).

Next, we modeled the outcome dependent on the exposure, mediator, and covariates, $E(Y|X = x, M = m, \boldsymbol{Z} = \boldsymbol{z})$, using a generalized linear mixed model (GLMM) to account for the clustered nature of the data. We regressed the binary indicator for *Coccidioides* presence

against fixed effect binary indicators for rodent presence and burrow, including nested random effect for plot number and cluster pair, and assuming a binomial distribution for the outcome. Because soil conditions, $\boldsymbol{Z}$, lie along the causal pathway, we conduct analyses that do and do not control for soil moisture, vegetation, and soil temperature. Effect estimates derived from analyses that do not control for such covariates capture the full effects of rodents and burrows, while effect estimates from analyses that do control for them capture the effects of rodents and burrows beyond abiotic changes to soil conditions (e.g., keratin in soil, enhanced nutrient profile, etc.)

Next, we simulated the potential mediators and outcomes, following the approach outlined by Wang and Arah [119]. Briefly, we drew 200 bootstrapped copies (resampling with replacement) of the original sample, clustered by plot. We simulated values for covariates, $\boldsymbol{Z}$, and rodent presence, X, that followed the same distribution as the observed data. We simulated each value of M (termed m_x, equal to 1, burrow vs. 0, surface soil) using the deterministic model for M described, conditional on $X = x$. Finally, we simulated Y using the fit GLMM, and various combinations of x (set to the observed distribution, set to 1, set to 0) and m_x. We regressed type of Y against simulated x to obtain different effects (see Table 2). To determine the percent of the association between rodents and *Coccidioides* that is mediated via burrows, we divided the total indirect effect by the total effect, prior to exponentiating. Since our models assumed logit link functions for a binary outcome, we exponentiated results to obtain odds ratios as the effects of interest.

Because it has been speculated that the association between burrows and *Coccidioides* might be attributable to more favorable soil conditions induced by burrows, we also examined the association between soil conditions, burrows, and rodents. We used similar modeling framework as described, using each soil condition in turn as the outcome (Figure 4). GLMMs for this analysis assumed a Gaussian distribution of the outcome and an identity link function, so results reflect differences.

2.4 Results

Descriptive results

We collected 250 samples of soil across a factorial design that included: rodents + burrows (n = 100); rodents + surface soils (n = 50); no rodents + burrows (n = 50); no rodents + surface soils (n = 50). *Coccidioides* was detected in 31% (31/100) of the burrow samples taken from outside rodent exclosures, and 23.4% (11/50) of the burrow samples collected from within the exclosures. Only six percent (3/50) of the surface soils samples where rodents were present had detectable *Coccidioides* DNA, and none of the surface soils samples taken from the exclosures. Of the three positive surface soils samples, one was taken closest to the precinct, and two were taken furthest away. The average soil moisture was 6.0%, the average soil temperature was 73^{o}C, and the vegetation was low grass (mean level = 1.1).

Contribution of rodents and burrows to Coccidioides *presence in soil*

Coefficients from the GLMM indicate a strong association between both burrows and rodents with odds of detecting *Coccidioides* in the soil (Table 1). In mediation analysis, the odds ratio (OR) representing the total effect (TE) of rodent presence on *Coccidioides* in soil was 15.3 (95% CI: 8.9, 28.5). In other words, the odds of detecting *Coccidioides* when rodents are present was 15.3 times higher than when rodents are absent (Figure 3A, Table 2). We estimated that 74.1% (95% CI: 66.1, 82.3) was mediated via burrows, meaning that 25.9% was not mediated via burrows. The direct, unmediated effects of rodents is represented by the total and pure direct effects (TDE; PDE). Here, we estimated that odds of detecting *Coccidioides* outside a burrow when rodents are present is 2.0 (95% CI: 1.0, 4.6) times the odds of detecting *Coccidioides* outside a burrow when rodents are absent, allowing for interaction between burrows and rodents.

Controlling for soil conditions reduced the estimated effects towards the null, but they were still significantly positive. Controlling for soil moisture, temperature, and vegetation, we estimated that the odds of detecting *Coccidioides* when rodents are present is 7.2 times higher (95% CI: 4.6, 12.5) than when rodents are absent (Figure 3B, Table 2). The lower OR implies that changes to soil conditions induced as a result of rodent and burrow presence accounts for some of the effect of rodents on *Coccidioides* presence. Removing the effect of soil conditions lowered the indirect effect of rodents via burrows (PIE: 3.2 (2.1, 5.6)), but slightly amplified the direct contribution of rodents (PDE: 2.8 (1.8, 4.9)). As a result, the percent of the total effect of rodent presence on *Coccidioides* that was mediated by burrows was lower. After removing the effect of soil conditions, we estimated that 54.3% (95% CI: 32.9, 57.6%) of the effect of rodents on *Coccidioides* presence was mediated via burrows, meaning 45.7% of the association is direct.

Table 3. Model coefficients from the generalized linear mixed model E(Y|X=x, M=m, Z=z) developed as part of the mediation models (Figure 3). Fixed effects are as shown in the table, and random effects are nested plot and paired cluster.

	Model A, OR (95% CI)	**Model B, OR (95% CI)**
Rodents present (*ref: exclosure*)	1.92 (1.06, 2.78)	3.58 (2.46, 4.71)
Burrows (*ref: surface soil*)	13.29 (12.07, 14.51)	7.25 (5.95, 8.55)
Soil moisture (water content %)	--	0.64 (0.43, 0.86)
Soil temperature (^{o}C)	--	0.99 (0.94, 1.05)
Vegetation level (ordinal, 0-3)	--	1.87 (1.29, 2.45)
Log likelihood	-98.14	-85.09
AIC	206.29	187.81
BIC	223.82	215.85

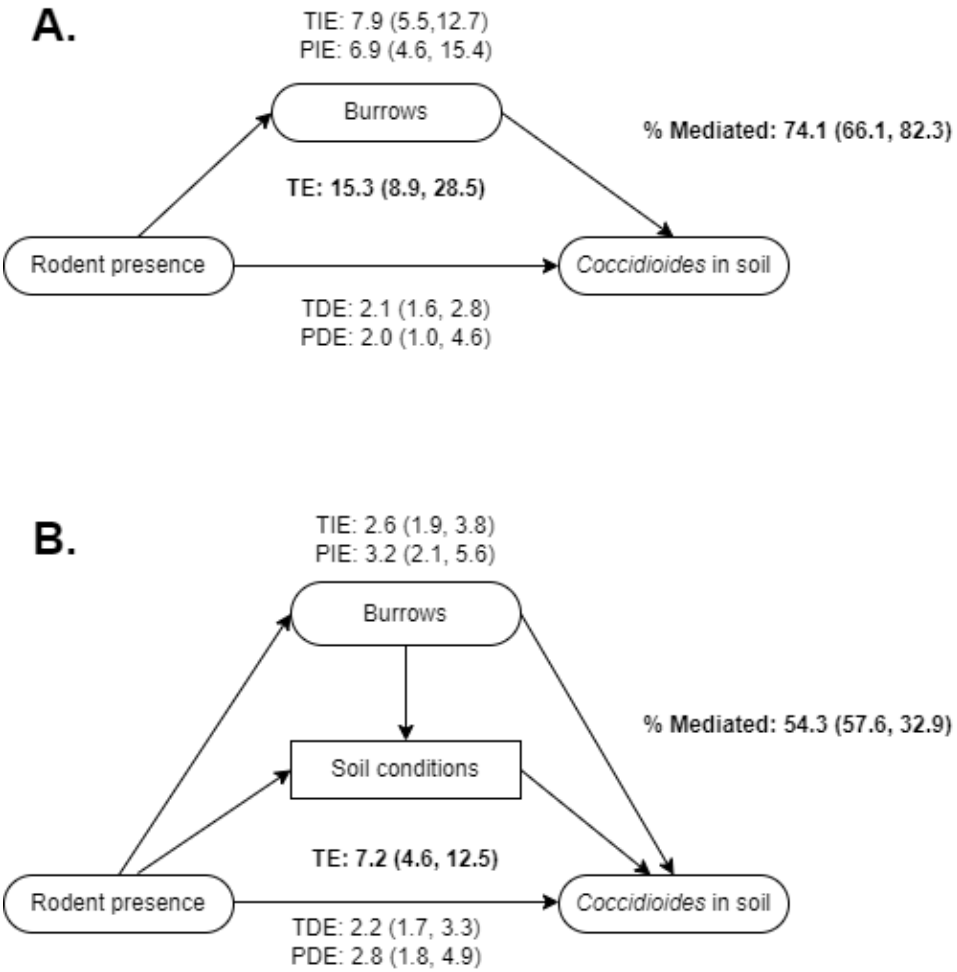


Figure 9. Directed acyclic graphs (DAGs) representing the hypothesized relationships between rodents, burrows, and *Coccidioides* presence in soils. Values reflect the odds ratios from causal mediation analysis not controlling for soil moisture, temperature, and vegetation (A), and controlling for these variables (B). TE = total effect; TDE = total direct effect; PDE = pure direct effect; TIE = total indirect effect via burrows; PIE = pure indirect effect via burrows.

Table 4. Odds ratios representing various effects from causal mediation analysis examining relationship between rodents, burrows, soil conditions, and *C. immitis* in soils.

Effect and counterfactual definition*	**Research question,** adapted from [[119]]	**Estimate,** *unadjusted for soil conditions* (Figure 2A) **OR (95% CI)**	**Estimate,** *adjusting for soil conditions* (Figure 2B) **OR (95% CI)**
Total effect (TE) $E[Y_{x1} - Y_{x0}]$	Overall, to what extent (relative) does rodent presence (X) increase probability of *Coccidioides* detection (Y)?	15.3 (8.9, 28.5)	7.2 (4.6, 12.5)
Total direct effect (TDE) $E[Y_{x1M_{x1}} - Y_{x0M_{x0}}]$	To what extent does rodent presence (X) cause *Coccidioides* presence (Y) other than via burrows (M) (e.g., surface soils), allowing burrows to boost up or tune down the effect?	2.1 (1.6, 2.8)	2.2 (1.7, 3.3)
Pure direct effect (PDE) $E[Y_{x1M_{x0}} - Y_{x0M_{x0}}]$	To what extent does rodent presence (X) cause *Coccidioides* presence (Y) other than via burrows (M)?	2.0 (1.0, 4.6)	2.8 (1.8, 4.9)
Total indirect effect (TIE) $E[Y_{x1M_{x1}} - Y_{x1M_{x0}}]$	To what extent does rodent presence (X) cause *Coccidioides* presence (Y) via burrows (M), including possible interactions between burrow and rodents?	7.9 (5.5, 12.7)	2.6 (1.9, 3.8)
Pure indirect effect (PIE) $E[Y_{x0M_{x1}} - Y_{x0M_{x0}}]$	To what extent does rodent presence (X) cause *Coccidioides* presence (Y) via burrows (M) only?	6.9 (4.6, 15.4)	3.2 (2.1, 5.6)
Percent mediated via burrow (%)	What proportion of the effect of rodents (X) on *Coccidioides* (Y) is mediated by burrows (M?)	74.1% (66.1%, 82.3%)	54.3 (32.9, 57.6)

*Y outcome (*Coccidioides*), X exposure (rodents present), M mediator (burrow), $x1$ and m_{x1} represent the index values while $x0$ and m_{x0} represent the reference, and $m_{x1} = P(m|x = 1)$. Counterfactual definition here is represented as difference for simplicity, but values show are odds ratios, not differences.

Role of soil conditions in Coccidioides *presence, and role of rodents and burrows in mediating soil conditions*

Rodents and burrows both had influences on soil moisture, vegetation, and temperatures. Compared to rodent exclosures and surface soils, soil moisture was lower in areas with rodents and in burrows, with 58.3% of the effect of rodents on soil moisture mediated via burrows (Figure 4A). The mean soil moisture in April was 6.0%. Overall, rodent presence was associated with a 1.5 lower percentage point in soil moisture. Both rodents and burrows also contributed to cooler soil temperatures, with 44.3% of the effect of rodents on soil temperature mediated by burrows (Figure 4B). Of note, all point estimates of temperatures were taken during the day, when conditions were hot. Analysis of continuous measurements of soil temperature revealed that at night, when conditions were cooler, burrows might have led to cooler warmer temperatures (Figure 5). Finally, rodents decreased vegetation by about one ordinal level (e.g., low to none, medium to low), with burrows having little mediating effect (Figure 4C).

We examined the contribution of soil conditions to *Coccidioides* presence by examining the odds ratios from the GLMM in the mediation analysis. Controlling for burrow and rodents, we found that the probability of *Coccidioides* detection was higher for lower soil moistures and higher vegetation levels (Table 1). For each percent increase in soil moisture, the odds of *Coccidioides* detection was reduced by 36% (95% CI: 14, 57%). For each ordinal level increase in vegetation (e.g., none to low, low to medium), the odds of *Coccidioides* detection increased by a factor of 1.87 (95% CI: 1.29, 2.44). Soil temperature was not associated with *Coccidioides* detection.

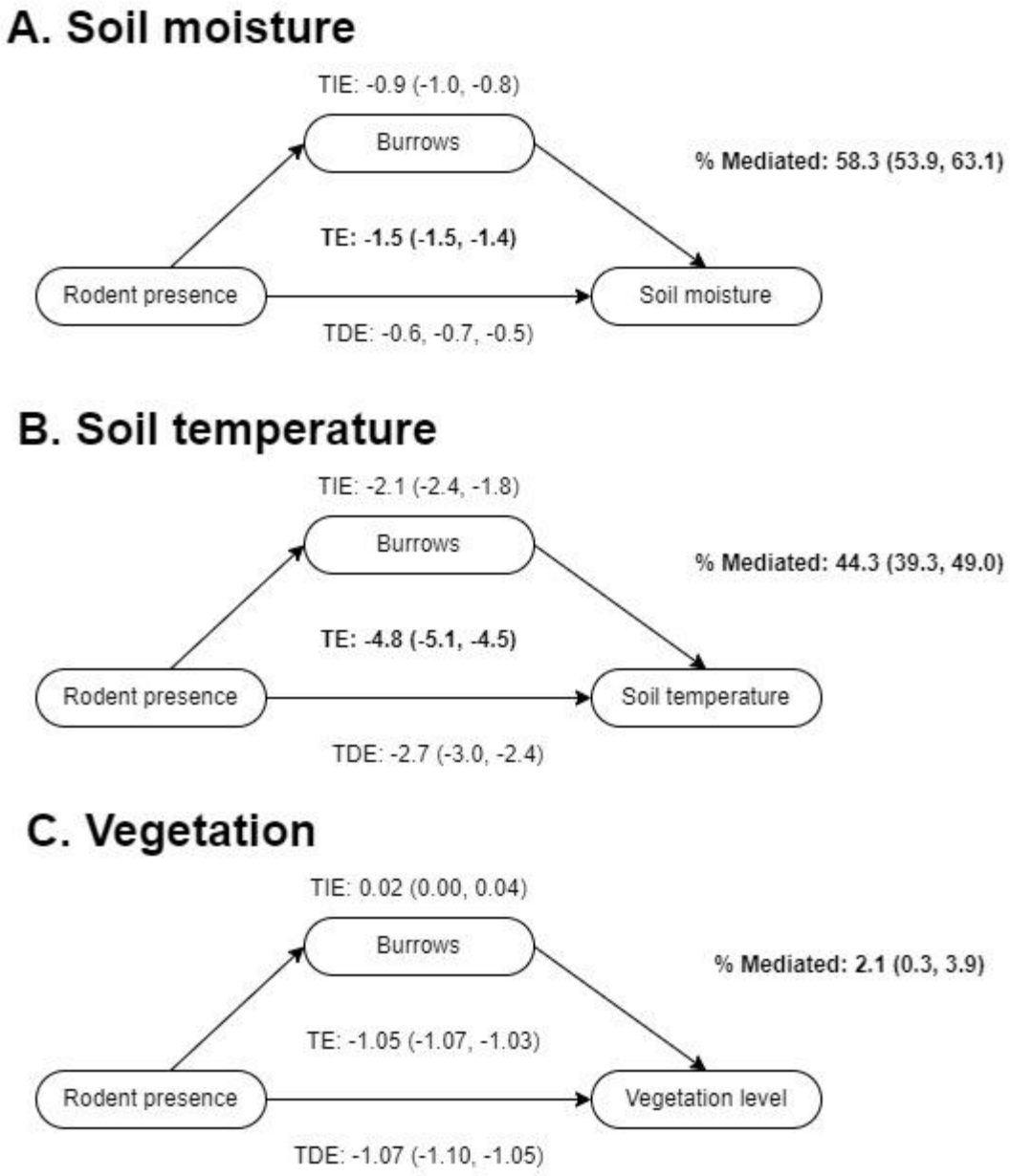


Figure 10. Directed acyclic graphs (DAGs) representing the hypothesized relationships between rodents, burrows, and soil conditions: moisture (% water content) (A), temperature (^{O}C) (B), vegetation level (ordinal, none=0, low = 1, medium = 2, high = 3) (C). Values reflect differences (not odds ratios as before). TE = total effect; TDE = total direct effect; PDE = pure direct effect; TIE = total indirect effect; PIE = pure indirect effect.

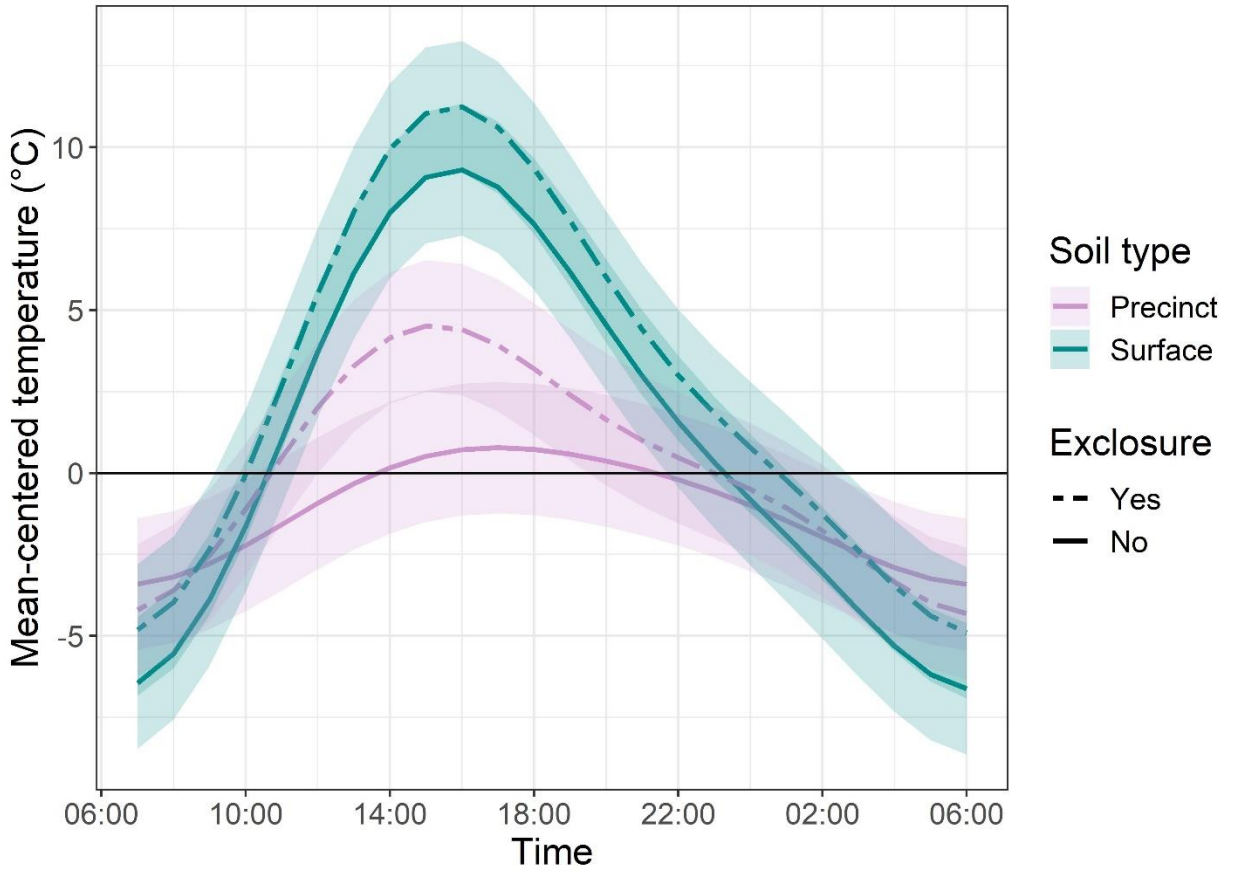


Figure 11. Comparison of continuous temperature measurement across burrows, surface soils, rodent exclosures, and rodent non-exclosures. Shaded color regions indicate 95% confidence intervals and solid lines represent the mean estimated mean-centered temperatures.

2.5 Discussion

We find evidence that both rodent presence and burrows are independently associated with increased probability of detection of *C. immitis* in soils. To our knowledge, this is the first study that has been able to disentangle, in part, the effect of rodents from the effect of burrows. Using a causal mediation approach, and assuming that in non-experimental conditions rodent presence is a requisite for burrow presence, we find that the odds of detecting *Coccidioides* when rodents are present is 15.3 times (7.2 times controlling for soil conditions) higher than when rodents are absent. We find that over 25% of the association between rodent presence and *Coccidioides* detection is not mediated through burrows. This lends support to the endozoan small mammal hypothesis, which postulates that spores from lung granules in infected rodents are released upon host death, releasing endospores that utilize the host body as a substrate for mycelial growth [26]. Controlling for soil conditions, we find that nearly 46% of the association between rodent presence and *Coccidioides* detection is not mediated via burrows. This result may suggest that burrows create important microhabitats that support the soil conditions conducive to *Coccidiodies*, while rodents themselves provide important chemical

(e.g., keratin, nitrogen) and microbiological factors (e.g., changes to antagonistic soil microbes) not controlled for in this analysis.

The strong indirect effects of burrows on *Coccidioides* presence additionally suggests that the burrow environment may serve as a suitable microhabitat for the fungus, regardless of rodent presence. When we examined soil conditions, we found that rodents mainly affect vegetation levels, leading to lower levels, while both rodents and their burrows lead to lower soil moisture levels. This finding is largely consistent with the literature, which attributes lower moisture to greater infiltration rates and porosity within burrows [113-115]. In our models, lower soil moisture was significantly associated with increased detection of *Coccidioides*, even after controlling for sample type (burrow vs. surface soil) and rodent presence. In addition, we found that soil temperatures within burrows experience greater homeostasis, at least compared to the top 10 cm of soil, in agreement with prior study [111, 112]. At the same time, *Coccidioides* spores have been shown to survive as long as 10 years in the soil [120], and in soils containing buried carcasses of mice experimentally infected with Coccidioides, the fungus was recovered as many as six years post burial [121]. Thus, it also remains possible that effects of historical rodent presence (including increased keratin in soil or old carcasses) are still contributing to this observed association.

The hypothesis that rodents serve as important reservoir hosts could further explain the observed relationships between coccidioidomycosis incidence and meteorological patterns [110]. For instance, if rodent hosts play a key role in the lifecycle of the fungus, rainfall could play an indirect role in supporting fungal growth by supporting vegetation growth that sustains large host populations, which in turn contribute to high levels of keratin in the soil [26]. Rodent host abundance declines markedly in years of drought and rebounds following the drought's end [91], much in the same way that human coccidioidomycosis incidence appears to rebound in years following a drought [59, 110, 122]. Drought provides a competitive resource advantage to small-bodied or rare host species, as larger or more prevalent species such as the Giant kangaroo rat experience steeper die-offs during drought [91].

This study has limitations. Although we were able to disentangle the influence of rodents and burrows to a degree, fully disentangling the influence of rodents, burrows, and soil conditions is a greater challenge. We examined soil moisture, temperature, and vegetation, yet other unmeasured abiotic factors, including humidity, soil texture, and porosity, may be affected by rodent presence and directly influence *Coccidioides* growth and survival [54, 83]. Beyond abiotic factors, burrowing rodents may have a profound influence on soil nutrient profiles, by modifying the local flora, bringing nutrients from subsurface levels to the surface, increasing litter decomposition, and creating defecation chambers that concentrate nitrogen [115, 117]. In prior study from the Carrizo Plain, Gurney and colleagues compared soil properties between mounds created by Giant kangaroo rats to surface soil between the burrow precincts and found 2.2 times higher concentration of nitrates [113]. In addition to nitrates, precinct mound soils in the Carrizo Plain were higher in organic matter, exchangeable calcium, bicarbonate and phosphorous, but lower in exchangeable magnesium and similar in exchangeable sodium, potassium, and cation exchange capacity [113]. We cannot untangle several possible, primary

sources of nutrition for Coccidioides, e.g., 1) recently dead or dying rodents, 2) rehydrated rodent carcasses or, 3) hair shed from rodents. Nor can we untangle the contribution of animal protein vs plant protein in the form of seed caches or other nutrients in the form of animal waste. Future laboratory experiments may seek to disentangle these sources.

Another limitation relates to our inability confirm whether burrows in the exclosure are truly inactive and burrows outside the exclosure are truly active, despite our visual assessments. Nevertheless, we are confident that, even with occasional breaching, rodent activity within the exclosures is meaningfully lower than outside exclosures. Active exclosures and inactive non-exclosures would bias the associations reported here rodents towards the null. Additionally, we assume that rodent presence is a requisite for burrow presence. This assumption may be flawed in that in the real world, there may exist sites with long abandoned burrows. However, in the context of our geographic region, where population of rodents is high and multiple species utilize the burrow systems built by the Giant kangaroo rat, we feel the assumption is tenable. Allowing the deterministic model for burrows conditional on rodents to burrow presence in the absence of rodent presence would move the total effect of rodents on *Coccidioides* closer to the null, while not affecting the association between burrows and *Coccidioides* presence.

This study suggests future work. For one, it has been speculated that the ability of *Coccidioides* to degrade keratin evolved as an evolutionary response to cope with extreme nutrient limitations within a desert environment [26]. If this is the case, we may expect to see that this association is strongest during periods of drought, when nutrients are most limited. We conducted this study during a period of extreme drought in California. Future work may seek to examine how climate modifies associations between rodents, burrows, and fungal presence in soils. Additionally, it remains unknown the risk that *Coccidioides* within burrows poses to individuals who do not directly dig in the soil. Prior laboratory experiment has buried carcasses of mice infected with *C. posadasii* at the bottom of soil columns, and reported that six days later, hyphae were detected at the surface of the soil that had been sterilized, but not at that which had been left unsterilized [123]. Intensive sampling around the mouth and on the top of burrows might clarify how *Coccidiodies* originating within burrows might become airborne.

The presence of rodents and/or their burrows has been noted at several historical outbreaks [124, 125]. In lieu of an available vaccine for coccidioidomycosis, prevention measures include education of outdoor workers for identification of symptoms, and provision of protections such as masks and soil wetting. Our findings motivate additional protection of workers in coccidioidomycosis endemic regions digging in soils that contain burrowing mammals.

2.6 Conclusion

Within an area highly endemic for coccidioidomycosis, we report direct effects of both rodents and burrows on the presence of *C. immitis* in soils. Our findings lend support for the both the endozoan small mammal hypothesis, as well as the idea that burrows create suitable microhabitats through altered soil characteristics, such as lowered moisture. Our findings also

motivate additional protection of workers in coccidioidomycosis endemic regions digging in soils that contain burrowing mammals.

CHAPTER 3. Model-based assessments of the effect of school closures on social mixing of children, and the effect of vaccination and other NPIs on SARS-CoV-2 Delta variant transmission dynamics within K-12 school populations

[2] Chapter 3 is included with permission from co-authors Kristin L. Andrejko, Qu Cheng, Philip A. Collender, Sophie Phillips, Anna Boser, Alexandra K. Heaney, Christopher M. Hoover, Sean L. Wu, Graham R. Northrup, Karen Click, Naomi S. Bardach, Joseph A. Lewnard, and Justin V. Remais.

3.1 Abstract

Background

School closures may reduce the size of social networks among children, potentially limiting infectious disease transmission. We estimated the impact of K-12 closures and reopening policies on children's social interactions and COVID-19 incidence in California's Bay Area. We examined school reopening policies for two upcoming semesters – one prior to vaccination availability, and one following availability of vaccine for individuals aged 12 years and older.

Methods

We collected data on social contacts among school-aged children in the California Bay Area and developed an individual-based transmission model to simulate transmission of SARS-CoV-2 in schools. We compared results for both Alpha and Delta variant circulations.

Results

Elementary and Hispanic children had more contacts during closures than high school and non-Hispanic children, respectively. We estimated that spring 2020 closures of elementary schools averted 2,167 cases in the Bay Area (95% CI: -985, 5,572), fewer than middle (5,884; 95% CI: 1,478, 11.550), high school (8,650; 95% CI: 3,054, 15,940) and workplace (15,813; 95% CI: 9,963, 22,617) closures. In the absence of a vaccine, and under moderate community transmission of the Alpha variant, we estimated that reopening for a four-month semester without any precautions will increase symptomatic illness among high school teachers (an additional 40.7% expected to experience symptomatic infection, 95% CI: 1.9, 61.1), middle school teachers (37.2%, 95% CI: 4.6, 58.1), and elementary school teachers (4.1%, 95% CI: -1.7, 12.0). Absent vaccine, we found that reopening policies for elementary schools that combine universal masking with classroom cohorts could result in few within-school transmissions, while high schools may require masking plus a staggered hybrid schedule. Stronger community interventions (e.g., remote work, social distancing) decreased the risk of within-school transmission across all measures studied, with the influence of community transmission minimized as the effectiveness of the within-school measures increased. At 70% vaccination coverage of students 12 and older, we found continued benefit (reductions in infections by >57%) of universal masking. In the absence of NPIs, increasing the vaccination coverage of community members from 50% to 70% or elementary teachers from 70% to 95% reduced the excess rate of infection among elementary school students by 24% and 37%, respectively.

Conclusions

Early school closures reduced contacts among children and averted infections, although the benefit may have been higher among non-Hispanic and higher income households. We found that schools are not inherently low risk, yet can be made so with high community vaccination coverages and masking. Vaccination of adults protects unvaccinated children.

3.2 Introduction

In response to the COVID-19 pandemic, long-term K-12 school closures were implemented across many settings to reduce risk of SARS-CoV-2 transmission among students, teachers, and family members. However, the long-term continuation of school closures poses a grave threat to healthy child development [29-31] and may exacerbate existing racial and socioeconomic gaps in school achievement [32] or nutrition [33]. The lack of data on children's social behavior during long-term closures has prevented robust assessment of school closure policies. Contact surveys among children have found weakened contact networks during short-term school closures [126], weekends, and holidays [127], but the impact of long-term COVID-19 related school closures on children's contact networks remains unclear. Much of our understanding about social contact patterns during the COVID-19 pandemic has been limited to adult behaviors [128-130] with only one study quantifying social contacts amongst children [131].

COVID-19 outbreaks within schools that held in-person instruction without physical distancing modifications [132] highlight the need to rigorously determine—and enact—effective risk reduction measures. A U.S. modelling study estimates that reductions in within-school mixing of children via classroom cohorts or hybrid schedules may limit risk of school-attributable infection by 4 to 7-fold, respectively [133]. Modification of individual behaviors, such as wearing face masks [133-135], quarantine of contacts of sick individuals [136], and increased testing [17], is also expected to reduce school-based transmission. K-12 schools in North Carolina reported only 32 school-acquired infections among over 90,000 students that attended in-person schooling with precautions involving universal masking, daily symptom monitoring, and a 2-day per week hybrid schedule [137]. Nevertheless, these studies may be limited by non-detection of asymptomatic transmission. The REACT study, in the UK, assessed time trend data of both asymptomatic and symptomatic infection, finding that children ages 13-17 years had a similar infection prevalence as working age adults, and only slightly higher than children ages 5-12 years [138]. Increases in prevalence were observed in children, and other age groups, after the reopening of schools in September, 2020 [138]; however, national reopening guidelines recommended that masks not be used in any classroom [139].

Differences in school size and social mixing patterns across age groups, as well as possible differences in susceptibility and transmissibility by age and the variants circulating [140], may contribute to heterogeneity in transmission risk across schooling levels. Meta-analysis found that children below 10 years of age had 48% lower odds of secondary infection of SARS-CoV-2 compared to adults, whereas there was no significant difference between adolescents and adults [141]. Secondary attack rates derived from contact tracing data of child index cases are conflicting, and it remains unclear whether children and adults are similarly infectious [142-146]. Empirically, differences in transmission between elementary and high school aged children are observed. In England, a study of over 9 million adults found that living with a child aged 12-18 years, but not a child 0-11 years, was associated with slightly increased risk of SARS-CoV-2 infection [147]. Serological testing prior to closures in France revealed limited evidence of secondary transmission within primary schools [148], but a high seroprevalence of 38% among high school students and 43% among high school teachers after reopening [148]. Accordingly, it is imperative that the impact of school closures be evaluated separately for

elementary, middle, and high schools. At the same time, teachers and staff may experience higher risk of infection as compared to students. In the UK, monitoring of over 19,000 schools between June 1 – July 17 revealed 210 cases across 55 outbreaks [149]. Staff made up 73% of cases, and 26 outbreaks were driven by staff to staff transmission [149]. Therefore, it is also critical to assess impacts in different school community groups—teachers, students, and family members.

In March of 2020, the California Bay Area was among the first in the nation to close schools, moving the 2020 spring semester to remote instruction [28]. The objectives of this study were to: 1) estimate social contact patterns among school-aged children during Bay Area (California) COVID-19 related school closures; 2) estimate the cumulative incidence of COVID-19 throughout the 2020 spring semester under counterfactual scenarios had schools or workplaces remained open, or social distancing policies not been enacted; and 3) estimate the effect of various school reopening strategies in Bay Area schools by grade level and across a new school semesters. For objective 3, estimated impact of reopening strategies for two semesters with different variants in circulation and different interventions available are reported. During the spring 2021 semester, the dominant variant in circulation was the Alpha variant, and vaccinations were not available for any age group. The fall 2021 semester featured rising rates of the more transmissible Delta variant across the U.S. [150], but vaccines with high effectiveness against infection [151-154] with SARS-CoV-2 were available for individuals aged 12 years and older.

3.3 Methods

We conducted a survey to ascertain the contact rates of children and their adult family members during spring school closures. We used these contact rates within an individual-based transmission model to examine the impact of spring school closures and reopening strategies.

Survey methodology

We implemented a social contact survey of school-aged children in nine Bay Area counties (Alameda, Contra Costa, Marin, Napa, San Francisco, San Mateo, Santa Clara, Solano, Sonoma) during county-wide shelter-in-place orders. Survey respondents reported the number and location of non-household contacts made within six age categories (0-4, 5-12, 13-17, 18-39, 40-64, and 65+ years) throughout the day prior. A contact was defined as an interaction within six feet lasting over five seconds.

Eligible households contained at least one school-aged child (preK – grade 12). A first sample was obtained using a web-based contact diary distributed in English via social networks (Nextdoor, Berkeley Parents Network) between May 4 and June 1, 2020. A second sample was procured between May 18 and June 1, 2020 via an online panel provider (Qualtrics) to be representative of Bay Area race/ethnicity and income. In both samples, surveys asked one adult respondent per household to respond on their behalf and for all children in their household. The survey also recorded household demographic information, including adult occupation status. A copy of the survey tool is included in the Supporting Information.

Survey analysis

To adjust for potential selection bias, we calculated post-stratification weights reflecting joint distributions of race/ethnicity and income of the counties' combined population using the 2018 one-year American Community Survey Public Use Microdata Sample (PUMS) from the nine counties. To account for potential bias due to occasional non-response on location questions, we applied a second set of weights equal to the inverse of the probability of response, conditional on race and income (fixed effect) and household ID (random effect). Weighted and unweighted survey data yielded similar results (Supporting Information; Figure S2).

Contact matrices generated using weighted and unweighted survey data were stratified by income, race, and location of contact. To determine whether an individual's total reported contacts varied by key covariates, we fit a multivariable linear regression model accounting for a household random effect and fixed effects for age, race, household income, number of household members, single parent household, weekday of reported contact, school type, and a binary indicator of whether more adults within the household worked at home during shelter-in-place compared to before shelter-in-place.

We conducted all statistical analyses using R (version 3.2.2; R Foundation for Statistical Computing; Vienna, Austria), and fit random effects models using the *lme4* package [155].

Ethics statement

Ethical approval was obtained from the Office for Protection of Human Subjects at the University of California, Berkeley (Protocol Number: 2020-04-13180). Prior to taking the anonymous survey, parents were provided a description and asked to provide written informed consent.

Transmission model

Using survey-derived estimates of contact patterns, we developed a transmission model to estimate the number of cases, hospitalizations, and deaths that would have occurred under various counterfactual intervention scenarios (e.g., if schools had remained open), and used this model to simulate the impact of various school reopening strategies.

First, we generated 1,000 synthetic populations representative of the demographic composition of Oakland, California, following previous methods (Supporting Information) [126]. Each individual was assigned an age, household, and occupation status (student, teacher, school staff, other employment, not employed) upon which membership in a class or workplace was based. Each individual represented 25 individuals in the real population. All possible pairings of individuals were partitioned into one of six types of interactions, according to a hierarchy of highest shared membership: household > classroom or workplace > grade > school > community [156]. Community interaction represented the number of contacts expected between individuals from age groups i and j scaled by the total number of individuals in age group j, such that the total number of contacts per agent stayed constant were the

simulated population to be scaled up. We separated schools into elementary (grades K–5), middle (grades 6-8) and high (grades 9-12) schools.

We then developed a discrete-time, age-structured, individual-based stochastic model to simulate SARS-CoV-2 transmission dynamics in the synthetic population (Figure 1A). At each time increment, representative of one day, each individual is associated with an epidemiological state: susceptible (S), exposed (E), asymptomatic (A), symptomatic with non-severe illness (C), symptomatic with severe illness (H1, D1) resulting in eventual hospitalization before recovery (H2) or hospitalization before death (D2), recovery (R), or death (M). For the post-vaccination scenario, we added a compartment representing vaccinated individuals (V). A full description of the transmission model methodology is provided in the Supporting Information.

Based on their type of interaction (e.g. household, class, community), the daily contact rate between individuals *i* and *j* on day *t*, $K_{ij,t}$, was estimated for pairs of individuals following previous study [156]. Contact rates were scaled by a time-dependent factor between 0 (complete closure) and 1 (no intervention) representing a social distancing intervention to reduce contact between individual pairs. Pairs with a school or workplace interaction were reassigned as community interactions under closures. Because symptomatic individuals mix less with the community [157], we incorporated isolation of symptomatic individuals and quarantine of their household members. Following prior work, we simulated a 100% reduction in daily school or work contacts and a 75% reduction in community contacts for a proportion of symptomatic individuals, and an additional proportion of their household members [158]. This means that a proportion of students and staff would stay home from school if they themselves were symptomatic, while a smaller percentage would stay home from school if one of their household members was symptomatic. We assumed that individuals were in the infectious class for up to three days prior to observing symptoms [159], during which time they did not reduce their daily contacts.

To parameterize the model, we calculated the mean transmission rate of the pathogen, $\bar{\beta}$, using the next-generation matrix method [160]. Briefly, assuming an initial R_0 of 2.5 [161, 162], we solved for $\bar{\beta}$ as the ratio between R_0 and the product of the infection duration and the weighted mean number of daily contacts per individual during the pre-intervention period (Supporting Information, equation 2). To represent age-varying susceptibility [163], we then calculated an age-stratified β_i, that incorporated varying relative susceptibility by age while permitting the population mean to be $\bar{\beta}$ (Supporting Information, equations 3-4). Due to uncertainty in the relative susceptibility of children to SARS-CoV-2 infection compared with adults [141], we modeled scenarios where children under 10 years were half as susceptible as older children and adults, children under 20 years were half as susceptible as adults, and all individuals were equally as susceptible (see the Supporting Information for tabular review of studies on age-dependent susceptibility). Using these methods, we calculated the secondary attack rate among household members to be between 9.6% and 11.1%, in agreement with prior studies [143, 164-166]. For the semester with predominant Delta variant circulation, we calculated R_0

as 4.6, based on an average of R_0 for the Alpha (R_0 = 2.5) and Delta (R_0 = 5) variant weighted by the proportion of circulating variants in summer 2021 [167, 168].

Transmission was implemented probabilistically for contacts between susceptible (S) or vaccinated (V) and infectious individuals in the asymptomatic (A) or symptomatic and non-hospitalized states (C, H1, D1). Movement of individual *i* on day *t* from a susceptible to exposed class is determined by a Bernoulli random draw with probability of success given by the force of infection, $\lambda_{i,t}$:

$$\lambda_{i,t} = \alpha\beta_i \sum_{j=1}^{N} K_{ij,t} A_{j,t} + \beta_i \sum_{j=1}^{N} K_{ij,t}(C_{j,t} + H1_{j,t} + D1_{j,t}) \quad (1)$$

where *N* is the number of individuals in the synthetic population (*N*=16,000), and α is the ratio of the transmissibility of asymptomatic individuals to symptomatic individuals. Using estimates from studies evaluating risk of symptoms by age [163], we assumed 21% of infected individuals <20 years and 69% of infected individuals 20 years and older experienced symptoms [163]. Following previous work [163], we assumed α to be less than one, as asymptomatic individuals may be less likely to transmit infectious droplets by sneezing or coughing [169]. We explored differences in age-dependent transmissibility by modelling scenarios that varied α.

Vaccines were modelled by adjusting agents' susceptibility to infection, probability of developing symptoms after being infected, and probability of developing severe disease after being infected. Prior to simulating transmission over the school semester, a proportion of susceptible individuals aged 12 and older were moved to the vaccinated compartment, according to a Bernoulli random draw with probability of success equal to the proportion vaccination coverage among the eligible population, and the disease progression tracks that the vaccinated individuals would follow post infection were updated (Supplemental information, equations 6-7). For most simulations, vaccine effectiveness was 77% against any infection [170], 85% against symptomatic infection [171], and 93% against severe infection [172]. To account for lower effectiveness due to waning immunity or new variants, we also explore scenarios with lower effectiveness.

Whether an individual remained asymptomatic, or was hospitalized, or died was determined via Bernoulli random draws from age-stratified conditional probabilities (Figure 1B, Table S5). The duration of time spent in each disease stage were sampled from Weibull distributions (Table S5). Simulations were initiated on January 17, two weeks before the first known case [173], assuming a fully susceptible population seeded with a random number (range: 5-10) of exposed individuals. We averaged results over 1,000 independent realizations, using one random draw from the synthetic population, and estimated confidence intervals as the 2.5th and 97.5th percentile of all realizations.

Estimating impact of spring 2020 school closures

A shelter-in-place order was announced for Bay Area counties on March 16 [28], following which 28% of work continued in-person [174], and schools were closed. Between January 17

and March 16, transmission was simulated as described above, deriving community contact rates during typical conditions (POLYMOD) [175].

We then simulated transmission March 17 - June 1, the remainder of the spring semester in the 2019-2020 academic year (Figure 1C), first under real-world conditions: no school contacts, 28% workforce participation [174], and community contacts derived from our social contact survey. Modelled output matched well with available data on hospitalizations, deaths, and seroprevalence (Figure S5). We then simulated transmission under counterfactual scenarios where: 1) schools remained open; 2) workplaces remained open; and 3) non-essential community contacts continued.

Community contact matrices were derived for each intervention based on survey and POLYMOD data to account for differences in location-specific contacts (e.g., transportation contacts increase for in-person work, daycare contacts decrease when school is in session) (Figure S2 and Table S4). For all counterfactual scenarios, except those permitting non-essential community contacts, we assumed 50% of household members of symptomatic cases reduced their community contacts by 75% and their work or school contacts by 100% [158]. We estimated the number of cases, hospitalizations, and deaths averted by the intervention as the difference between these outcomes for the counterfactual scenarios minus the modelled real-world scenario.

Estimating impact of reopening strategies prior to vaccine availability, Alpha variant circulation

We simulated the effect of school reopening strategies over a subsequent four-month semester; Figure 1C). We established initial conditions for these simulations by initiating model runs spanning a school closure period, and then modeled the effect of reopening strategies under two susceptibility assumptions (children <20 half vs. equally as susceptible as adults), and two transmission contexts (high and moderate community transmission). The high transmission context is characterized by 75% of workplaces remaining open and non-essential community contacts double what we observed in our survey; the moderate transmission context is characterized by 50% of workplaces remaining open and non-essential community contacts equal to that observed in our survey after Memorial Day (May 25). In our simulations, the school closure period aligned with summer break, and the reopening period with the 2020 fall semester; however, Bay Area school districts remained closed throughout the duration of the fall 2020 semester. We thus model various transmission scenarios in the school closure period so as to enable model simulations for the new semester to be generalizable to either a fall or a spring semester reopening after a variable closure period.

We simulated six school reopening strategies (Figure 1C; see the Supporting Information for details): 1) schools open without precautions; 2) classroom groups are enforced, reducing other grade and school contacts by (a) 50% (weak cohort), or (b) 75% (strong cohort); 3) hybrid with class sizes halved, and each half attends two staggered days each week; 4) hybrid with class sizes maintained, and half the school attends two staggered days each week according to grade groups; 5) all students and faculty wear masks; 6) faculty and/or students are tested with 85% sensitivity on a (a) weekly or (b) monthly basis [176], with positive cases isolated and their class

quarantined for 14 days (periodic test-trace-isolate, TTI). We examined the six interventions by themselves and in combination (e.g., cohorts, masks, and TTI). The average class size was 20 students.

Masks were assumed to reduce both outward and inward transmission by η_i [177], where η_i represents the efficacy of the mask for individual *i*. Meta-analyses that included cotton masks worn by general population found a reduction of infection risk of about 50% to the adult wearer [178]. Mask efficacy is lower among children than adults, and lower in younger children (about 15%) compared to older children, possibly related to inferior fit, or compliance with continuous use [179, 180]. We therefore assumed age-dependent mask efficacy (15% for elementary students, 25% for middle school students, 35% for high school students, 50% for teachers/staff). We estimate excess infections (symptomatic only and all infections), hospitalizations, and deaths attributable to school-based transmission as the cumulative incidence of infections, hospitalizations, and deaths under each school reopening scenario minus the cumulative incidence under a school closure scenario. We then identified which set of interventions are needed to reduce excess risk of symptomatic illness for teachers (the sub-population determined to be at highest risk) such that less than one additional percent becomes infected.

Estimating impact of reopening strategies post vaccine introduction, Delta variant circulation

We examined the effect masking, testing, and masking plus cohorts across three levels of community vaccination coverage (50%, 60%, 70%), assuming that vaccination coverage within school children 12 years and older and teachers matches that in the community. While all of the nine Bay Area counties have achieved vaccination coverages of at least 60% as of summer, 2021, and some over 80% [181], we include the lower 50% to make the findings more generalizable to areas outside the Bay Area who may otherwise have similar demographics. Results with vaccination coverages above 70% are included in the supplement.

Next, we considered within-school vaccination coverage in the absence of within-school NPIs (masking, testing, cohorting). We assumed a community vaccination coverage among the eligible population of 70%, which represented a conservative level of vaccination coverage among a Bay Area county [181]. We then examined COVID-19 outcomes if students 12 and older and teachers/staff had higher vaccination coverages (ranging from 70% to 95% coverage).

Finally, we estimated the additional cases averted in each population by masking the entire student and teacher population, compared to masking only the unvaccinated student and teacher population, in the absence of additional interventions. We held community and within-school vaccination coverage of the eligible (12+) population at 70%, and varied vaccine effectiveness from low (41% any infection, 45% symptomatic infection, 49% severe infection) to medium (59% any infection, 65% symptomatic infection, 71% severe infection) to high.

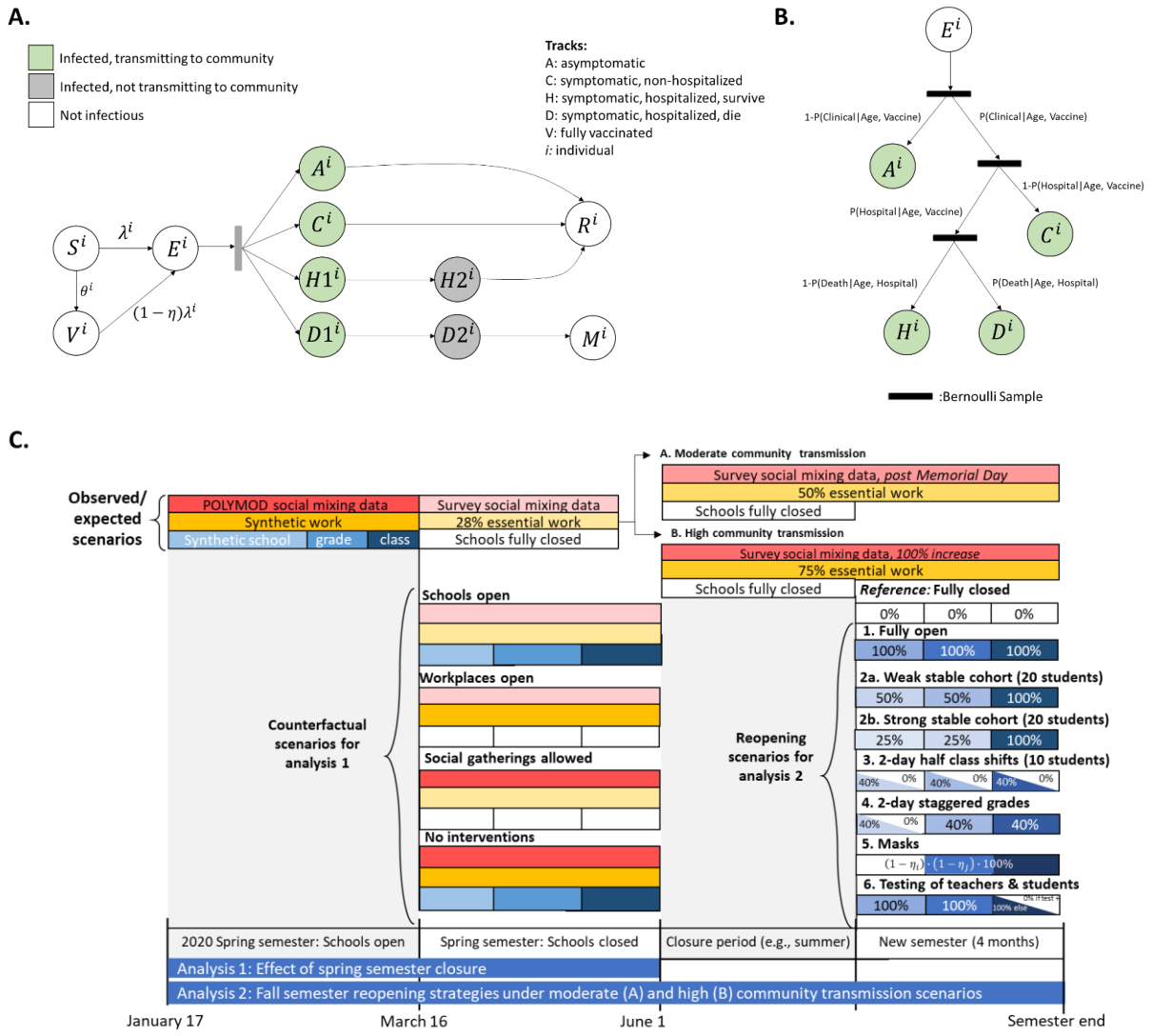


Figure 1. Model schematic A) Schematic of the agent-based susceptible-exposed-infected-recovered (SEIR) model. S = susceptible; E = exposed, A = asymptomatic; C = symptomatic, will recover; H1 = symptomatic and will recover, not yet hospitalized; H2 = hospitalized and will recover; D1 = symptomatic, not yet hospitalized; D2 = hospitalized and will die; R = recovered; M = dead; V = vaccinated; λ = force of infection defining movement from S to E. Superscript *i* refers to individual. After an agent enters the exposed class, they enter along their predetermined track, with waiting times between stage progression drawn from a Weibull distribution. B) Schematic of the conditional probabilities by which agents are assigned a predetermined track. C) Schematic of interventions simulated in the SEIR model. The first analysis examines transmission between January 17 and June 1, and tests the effect of several counterfactual scenarios that took place between the enactment of Shelter in Place (March 16) and the original end of the spring semester (June 1). The second analysis examines transmission over a subsequent four-month semester, and tests the effect of several simulated reopening strategies for the semester, expected to occur under a high and moderate community transmission scenario. Boxes represent categories of social contacts, including community (red), work (yellow), school (light blue), grade (medium blue) and classroom (dark blue). Percentages in the boxes represent the percentage of the contact rate experienced under a given intervention or counterfactual scenario (e.g., 0% represents a full closure). TTI = test-trace-isolate.

3.4 Results

Contact patterns

612 households provided contact histories on behalf of 819 school-aged children in the Bay Area (Table S1). The majority of non-household contacts occurred between individuals in the same age category, and while performing essential activities (such as grocery shopping, laundering clothing, or receiving health care), at work, home, or during an outdoor leisure activity (Figure 2A; Figure 2C). Children aged 5-12 years had twice as many non-household contacts (1.58 contacts per child per day) as teenagers aged 13-17 years (0.78 contacts per teenager per day) (Figure 2B).

In multivariable models adjusting for demographic and household characteristics, households identifying as Hispanic or Latinx had 2.32 (95% CI: 0.08-4.50) more contacts on average compared to non-Hispanic or Latinx households (Table 1). Households that did not indicate an increase in the number of adults working from home during shelter-in-place compared to before shelter-in-place had 1.85 (95% CI: 0.16-3.52) more contacts than households with more adults working at home during shelter-in-place.

Table 1. Differences in social contacts by demographic variables. Coefficients from multivariable linear mixed model adjusted for race (reference: White alone), self-reported household income (reference: < $150,000), whether household identified as Hispanic (reference: not Hispanic), whether household was a single parent household (reference: multi-parent household), whether date of reported contacts were weekend or weekday (reference: weekday), whether child attended a public or non-public school (including private, charter, homeschool, or other), age of individual in years, whether the date of reported contacts occurred over memorial day weekend (May 24 – May 26, reference: not the holiday weekend), and the change in number of adults working at home during shelter in place (reference: more adults working at home during SIP). SIP = shelter-in-place.

		Average adjusted difference in daily contact rate (95% CI)
Race (*ref: White alone)*		
	Asian alone	-0.77 (-2.4, 0.89)
	Black or African American alone	-1.33 (-3.93, 1.35)
	Other race alone	-2.94 (-6.46, 0.69)
	Two or more races	-1.43 (-4.66, 1.72)
Hispanic Household		2.32 (0.08, 4.5)
Household Income > $150K		-0.35 (-1.8, 1.12)
No. Individuals in Household		0.25 (-0.59, 1.05)
Single Parent Household		-0.32 (-3.73, 3.13)
Weekend		1.63 (-0.45, 3.69)
Public School		-0.2 (-1.79, 1.41)
Age		0.0 (-0.16, 0.16)
Memorial Day Weekend		1.28 (-1.03, 3.62)
Less or same no. adults working from home during SIP		1.85 (0.16, 3.52)

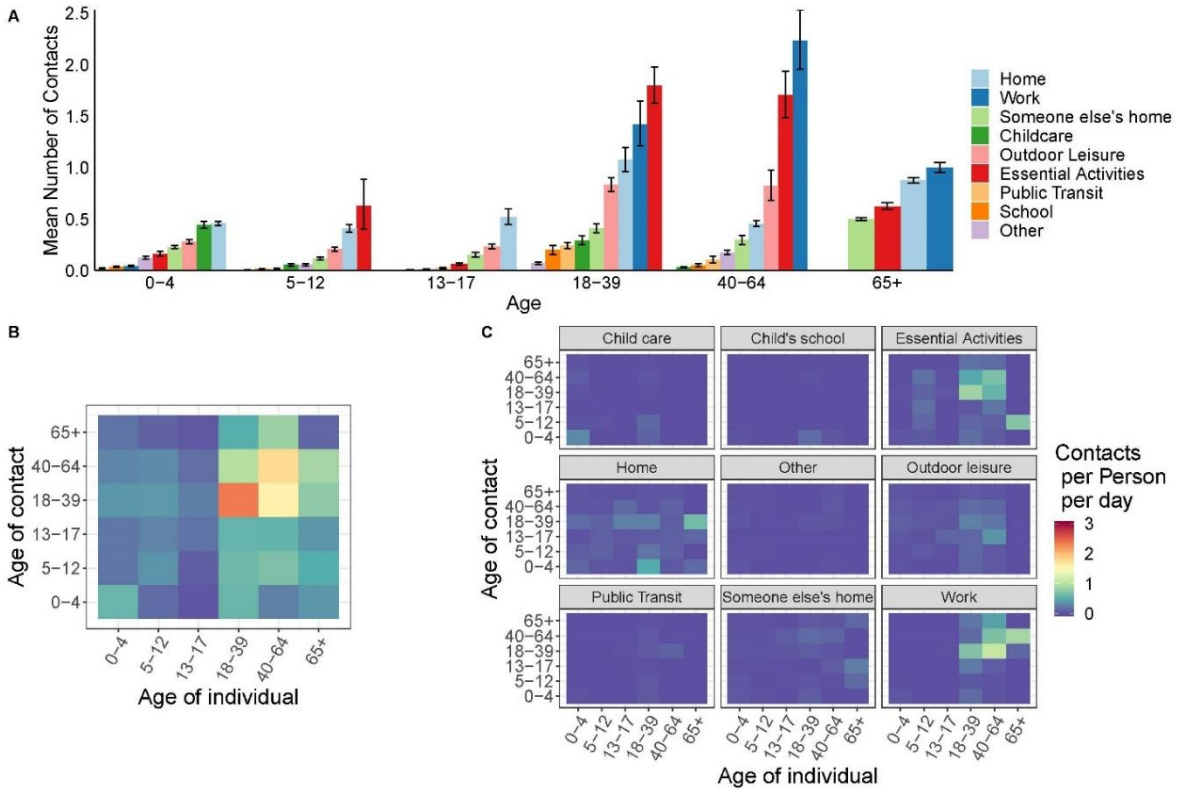


Figure 2. Social contact patterns between children and adult family members of Bay Area households, May 4—June 1, 2020. A) Average daily contacts per age group at nine pre-specified locations. B) Average daily contacts per person by age category of the survey respondent and reported contact, unweighted. C) Average daily contacts per person at each of the nine locations. Panels B and C share a legend.

Impact of spring 2020 school-closure policies

As of June 1, the nine Bay Area counties had reported 14,202 cases of COVID-19 [182]. Assuming a ratio (α) of the transmissibility of asymptomatic individuals to symptomatic individuals of 0.5, and susceptibility of children under 10 years set to half that of older children and adults, we estimated that there would have been 1.98 (95% CI: 0.44, 2.6) times more cases of COVID-19 throughout the nine Bay Area counties between March 16 – June 1 than observed had all K-12 schools remained open (Figure 3), corresponding to 13,842 (95% CI: 6,290, 23.040) excess confirmed cases. We estimated 3.16 (95% CI: 1.79, 4.89) times more cases would have occurred among families of students grades K-12 than observed. Examining cases averted by school level closures, we estimated that if elementary schools alone had remained open, the Bay Area would have recorded 2,167 additional cases (95% CI: -985, 5,572), while if only middle schools had remained open, an additional 5,884 cases (95% CI: 1,478, 11,550) would have been observed, and if high schools alone had remained open, an additional 8,650 cases would have been observed (95% CI: 3,054, 15,940). An additional 6,370 (95% CI: 1,853, 12,122) cases would have been recorded if middle schools and elementary schools had remained open. This means that when one level of schooling is closed, each additional closure has a smaller marginal benefit. This is in part driven by households with multiple school-aged children, who share the

same household contacts to whom an infection acquired within school could spread (Supporting Information; Figure S6a).

By comparison, had all workplaces remained open, we estimated that, as of June 1, there would have been 15,813 additional confirmed cases (95% CI: 9,963, 22,617), reflecting 2.11 (95% CI: 1.70, 2.59) times more cases than observed. If non-essential outings and social gatherings had been permitted, we estimated that there would have been an additional 7,030 (95% CI: 3,118, 11,676) confirmed cases, reflecting 1.50 (95% CI: 1.22, 1.82) times more cases than observed. All three interventions together helped avert an estimated 49,023 confirmed cases. The excess cases associated with opening both workplaces and schools was additive (Figure S6b). The effects of limiting social gatherings depended upon whether there were concurrent workplace or school closures; the number of excess cases associated with allowing social gatherings and in-person work, or allowing social gatherings and in-person school, was higher than the excess cases associated with either individually. This suggests that, by itself, social distancing is the least effective intervention; yet it becomes an important control measure when workplaces or schools are open. Reopening a school or workplace raises an individual's exposure to infection, which then increases the risk of a social gathering of individuals from multiple schools or workplaces, while also permitting infections to jump workplaces or schools (Supporting Information).

We find that both school and workplace closures in the spring of 2020 were necessary to achieve a sustained $R<1$. We estimated that the highest COVID-19 hospitalization occupancy that would have been observed on any one day during shelter-in-place if schools were open was 10.6 (95% CI: 6.0, 16.0) per 10,000 population, representing an excess of 4.42 individuals per 10,000 from the modelled real-world hospitalization occupancy. As the Bay Area has, on average, 12.3 beds available per 10,000 (22 beds per 10,000 capacity at 56% non-occupancy rate) [183], school closures permitted over a third of available beds to remain available, but were not necessary to keep Bay Area healthcare systems under capacity. As of June 1, 2020, the Bay Area had 3,997 confirmed deaths from COVID-19 [182]. We estimate that school closures averted 0.63 deaths (95% CI: -1.25, 3.75) per 10,000 population, corresponding to 663 averted deaths across the Bay Area, fewer than workplace closures (estimated 828 deaths averted) and more than restrictions on social gatherings (estimated 503 deaths averted).

At low levels of susceptibility (i.e., 25% that of adults) among children, the impact of school closures was small, and the ratio of transmissibility of asymptomatic individuals to symptomatic individuals (α) had little influence on the impact of spring school closure policies (Figure 4A). As children increase in susceptibility relative to adults, the influence of α becomes more pronounced (Figure 4A).

We found a significant positive relationship between the number of cases averted by school closures and the proportion of households in the population with children under 18 years (Figure 4B). For each 1% increase in the proportion of total households that have children under 18, we estimate an additional 5.8% increase over observed incidence had schools remained open throughout the spring semester.

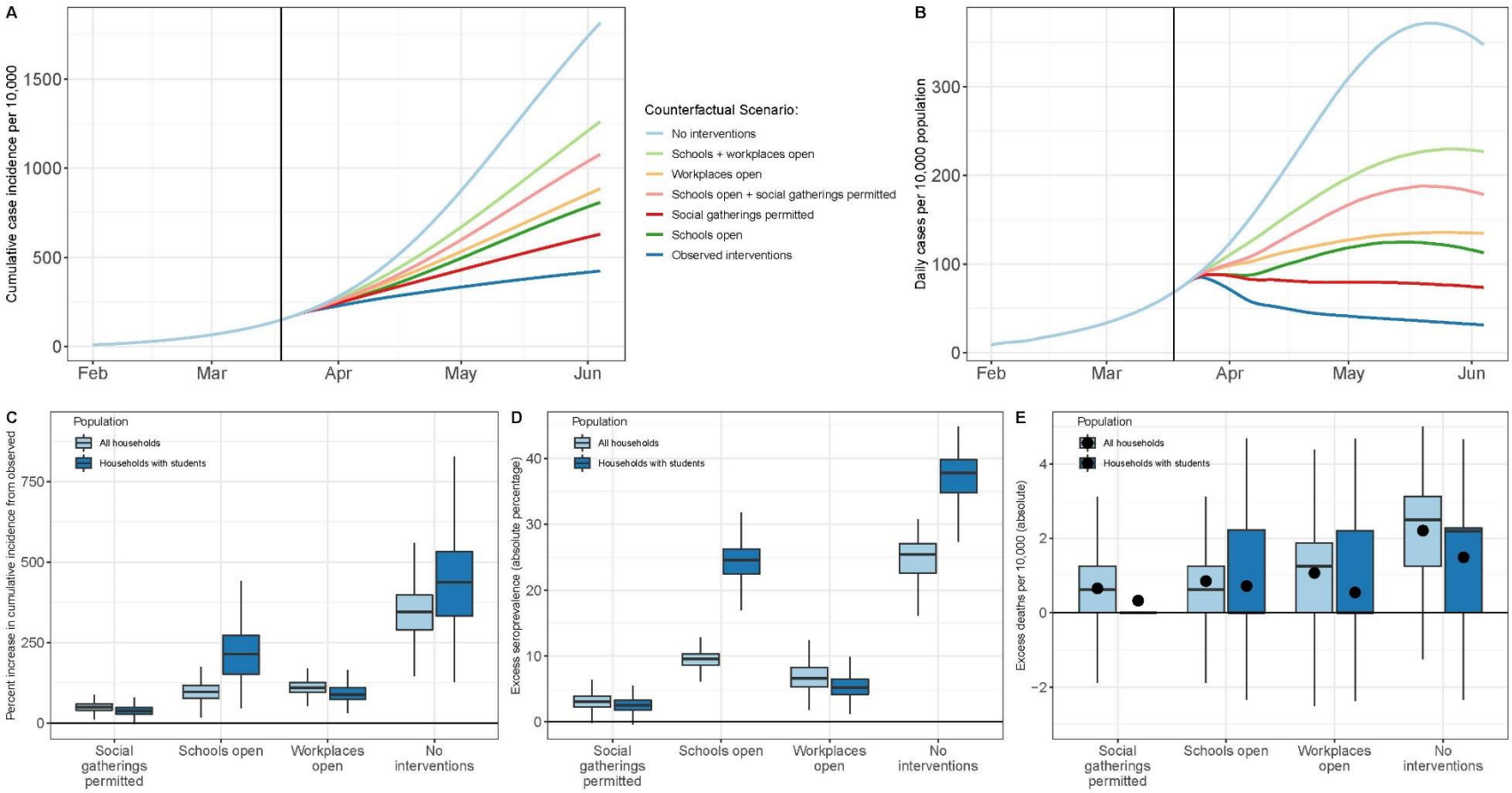


Figure 3. Effect of spring semester interventions. We simulated transmission between February 17 and June 1 assuming children <10 years are half as susceptible to infection compared with older children and adults. Between March 16 (enactment of shelter-in-place orders) and June 1 (the end of the spring school semester), we assessed potential outcomes under various counterfactual scenarios: 1) schools had remained open for the remainder of the school semester; 2) workplaces had remained open; 3) social gatherings were permitted; 4) no interventions were enacted. A) Modelled cumulative incidence according to the counterfactual scenario examined. Modelled predictions are not adjusted for under-reporting, which is expected to be substantial. B) Daily incidence per 10,000 per counterfactual scenario examined. C) The percent increase in cumulative incidence from observed incidence between February 17 and June 1, stratified by counterfactual scenario and population sub-group. D) The absolute difference in the percent of population seropositive for each counterfactual scenario compared to the modelled, observed seroprevalence between February 17 and June 1, stratified by population sub-group. E) The percent increase in deaths per 10,000 from observed

between February 17 and June 1, stratified by counterfactual scenario and population sub-group. The distribution of estimated death rate across 1,000 realizations was skewed, so black dots representing the mean number of excess deaths per 10,000 are added.

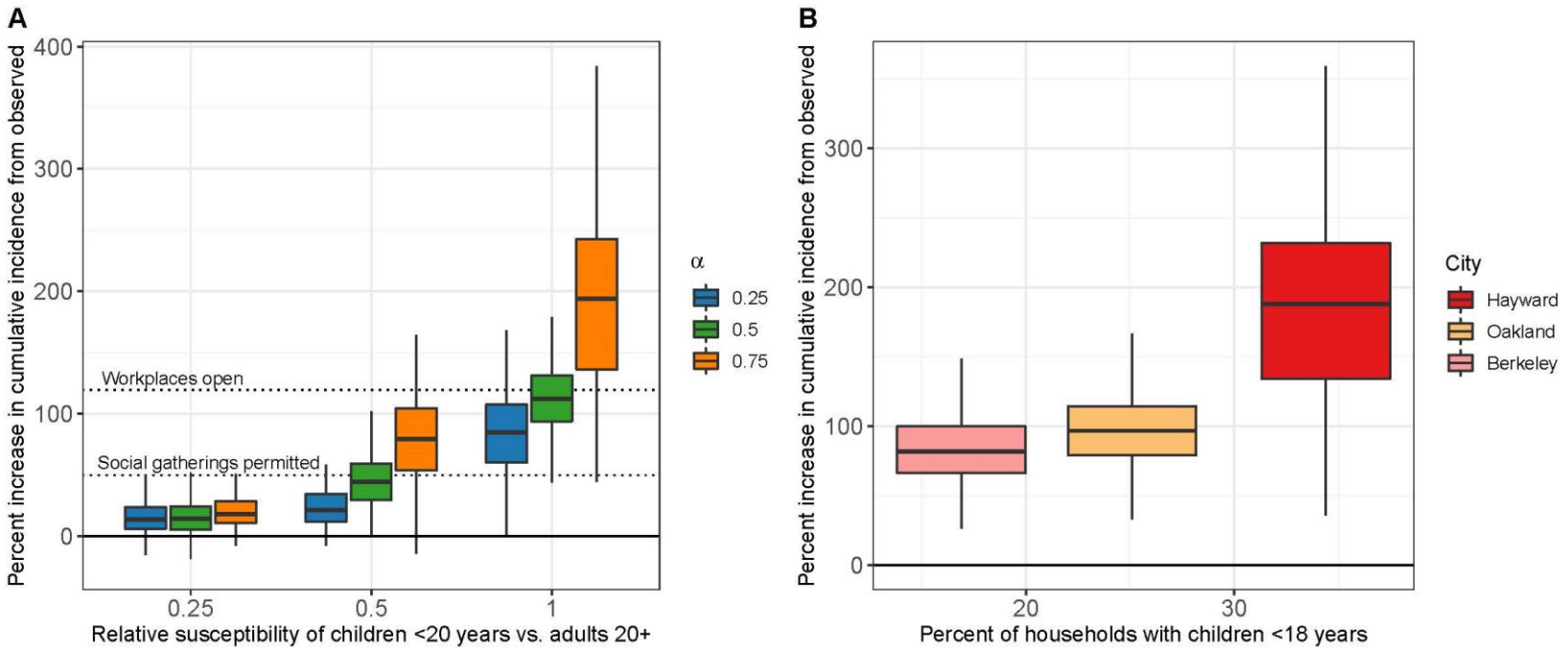


Figure 4. Influence of key epidemiological parameters on the effectiveness of school closures. The percent increase in cumulative incidence from observed incidence over the period February 17 - June 1 had schools remained open between March 17 and June 1. A) Results are reported for modelling scenarios that varied the ratio of the susceptibility of individuals under 20 years to adults 20 or older, and the ratio of the force of infection for asymptomatic infections to symptomatic infections (α). Dashed lines indicate the percent increase in incidence from observed that would have been expected if workplaces had remained open, and if social gatherings were permitted. B) Results are reported for synthetic populations with varying levels of the proportion of households with children under 18 years of age, reflecting three major Bay Area cities (Berkeley, Oakland, Hayward, assuming children under 10 are half as susceptible as older children and adults.

Simulated impact of reopening strategies: less transmission variant, no vaccine availability

The estimated risk of symptomatic infection associated with reopening for a four-month semester—across moderate to high transmission contexts—is highest for teachers and other school staff, followed by students and other household members of students and teachers/staff (Figure 5). Owing to larger average school sizes, we found high schools were at higher risk, followed by middle schools, then elementary schools. Staggered 2-day school weeks with halved class sizes provided the largest reduction in risk among all interventions considered, followed by strong stable cohorts of class groups, then wearing face masks. In the absence of other interventions, periodic (tests administered weekly or monthly) test-trace-isolate (TTI) strategies have low effectiveness, but when combined with strict social distancing measures, a modest reduction in community cases was possible as infectious individuals and their contacts identified in the school environment were quarantined (i.e., have their community contacts reduced by 75% for 14 days). Excess seroprevalence, hospitalizations, and deaths associated with school reopening, as they varied with respect to differing assumptions about child susceptibility and community controls, are detailed in Tables S6-S9.

We examined the effect of school reopening when modest community controls (e.g., 50% in-person work and continued social distancing) were in place, leading to moderate community transmission. Assuming individuals <20 years are half as susceptible as adults, with no precautions taken within school settings, we estimated that an additional 21.0% (95% CI: 0, 46.0%) of high school teachers, 13.4% (95% CI: -2.2, 38.6%) of middle school teachers, and 4.1% (95% CI: -1.7, 12.0%) of elementary school teachers would experience symptomatic illness over the four-month reopening period, compared to expectations if schools were closed (Figure 5). We estimated that the daily hospitalization occupancy rate would increase by an average of 0.53 (95% CI: -0.58, 1.73) hospitalizations per 10,000 individuals (roughly 4.2% of Bay Area available bed capacity), of which 0.13 (95% CI: -0.29, 0.58) and 0.33 (95% CI: -0.58, 1.30) hospitalizations per 10,000 would be among household members of students and other community members, respectively (Figure 6B). We estimated an excess total death rate of 0.56 (95% CI: -1.88, 3.13) per 10,000 over the four-month period, corresponding to 434 (95% CI: -1,451, 2,418) deaths across the Bay Area, of which 287 would be among community members without students in their household, 114 among household members of students, 31 among teachers, and one among students.

We also examined the effect of reopening when lessened community controls (e.g., 75% in-person work and limited social distancing) were in place, leading to high community transmission. With no precautions taken within school settings, we estimated that an additional 33.3% (95% CI: 11.1, 53.6%) of high school teachers, 24.4% (95% CI: 4.3, 44.4%) of middle school teachers, and 9.1% (95% CI: 0.9, 20.0%) of elementary school teachers would experience symptomatic illness (Figure 5). We estimated that the daily hospitalization occupancy rate would increase by an average of 1.65 (95% CI: -0.17, 3.38) hospitalizations per 10,000 individuals, of which 0.37 (95% CI: -0.22, 1.01) and 1.17 (95% CI: -0.36, 2.70) per 10,000 would be among household members of students or teachers and other community members, respectively (Figure 6B). We estimated an excess total death rate of 1.73 (95% CI: -2.50, 6.25)

per 10,000, corresponding to 1,341 (95% CI: -1,934, 4,837) deaths across the Bay Area, of which 1,026 would be among community members, 254 among household members, 60 among teachers, and one among students.

At moderate community transmission, we estimated that reducing excess risk of symptomatic illness for teachers to less than 1% would require either strict adherence to staggered school weeks (either as half classes or grades), or a combination of stable cohorts (weak or strong), wearing face masks, and monthly test-trace-isolate (see Table 2, which also details interventions necessary in high transmission contexts). Strong stable cohorts, 2-day staggered grades, or strong stable cohorts combined with wearing masks and periodic test-trace-isolate protocols are associated with reductions in deaths of 85%, 95%, and 95%, respectively.

We found that reducing community transmission via enhanced community controls significantly reduced the excess risk to teachers across all grades, from 18.4% (95% CI: 7.7, 27.9%) to 10.3% (95% CI: 0.4, 20.7%) in the no precaution scenario, with the influence of community transmission levels minimized as school-based interventions became stronger. Under minimal within-school interventions, the level of community transmission strongly determined whether the effect of school reopenings would be associated with increased incidence among the general community (non-students, teachers or family members). In high transmission settings where schools open without precautions, we estimated that the majority (59%) of the excess cases would be among community members, whereas in moderate transmission settings, fewer than half (45%) of the excess cases would be among community members (Figure 6A).

Regardless of the relative susceptibility of children to adults, across both moderate and high community transmission settings, a strict adherence to a combination of within-school distancing interventions (e.g., combining staggered half classes or staggered grades with stable cohorts; combining stable cohorts with wearing face masks and monthly test-trace-isolate protocols) was required to reduce the excess risk of symptomatic illness for high school teachers and all other school staff to less than 1% (Table 2). The benefit of having a strong (75%) versus a weak (50%) reduction in non-classroom (non-cohort) contacts is most notable when children are highly susceptible. For instance, in a high transmission context, reducing non-classroom contacts by 50% and 75% lowers the excess risk to all teachers from 32.1% to 15.3% and 5.3%, respectively. If children are half as susceptible, the excess risk to all teachers is lowered from 18.4% to 5.2% and 3.4%, respectively (Figure 5).

Table 2. School-based interventions to reduce risk. This table colors the reopening strategies examined by whether or not they are sufficient to reduce the additional proportion of teachers and other school staff experiencing symptomatic illness across a four-month semester to <1% of teachers. Strategies colored in green are strategies which reduce the excess number of teachers with symptomatic illness to <1%. Strategies colored in gray are strategies which do not reduce the excess number of teachers with symptomatic illness to <1%. Results are stratified by high school and elementary school teachers.

		Elementary school		High school	
	Community transmission:	Moderate	High	Moderate	High
Children half as susceptible	Stable cohorts (weak)				
	Masks				
	Stable cohorts (strong)				
	2-day staggered grades				
	2-day half class shifts				
	Stable cohorts*, masks + monthly TTI				
	2-day staggered grades + stable cohorts*				
	2-day half classes + stable cohorts *				
	All interventions**				
Children equally as susceptible	Stable cohorts (weak)				
	Masks				
	Stable cohorts (strong)				
	2-day staggered grades				
	2-day half class shifts				
	Stable cohorts*, masks + monthly TTI				
	2-day staggered grades + stable cohorts*				
	2-day half classes + stable cohorts *				
	All interventions**				

* Weak or strong
** All interventions include: masks, staggered grades, stable cohorts, and monthly TTI
TTI = test-trace-isolate

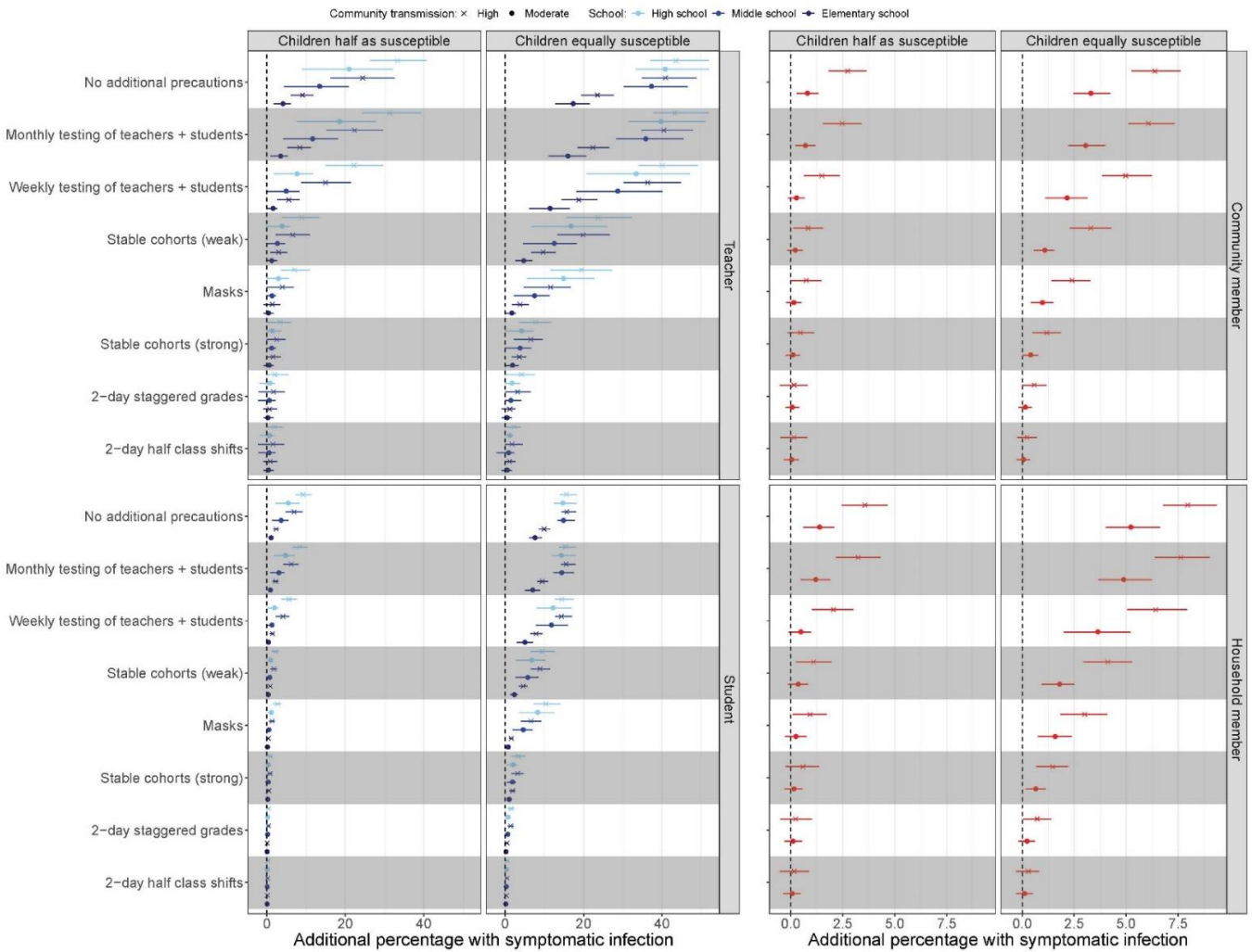


Figure 5. Excess risk by subgroup associated with school reopening strategies over a subsequent four-month semester. Panel A shows the additional proportion (mean and IQR) of each subgroup expected to experience clinical infection over the course of a four-month semester compared to if schools were closed under each reopening scenario and the four transmission contexts: children half and equally as susceptible as adults crossed with moderate and high community transmission. Colors indicate the transmission across levels of schooling (elementary, middle, and high) while the shape of the mean point indicates the level of community transmission (circle = moderate, cross = high). “Teachers” include teachers and all other school staff.

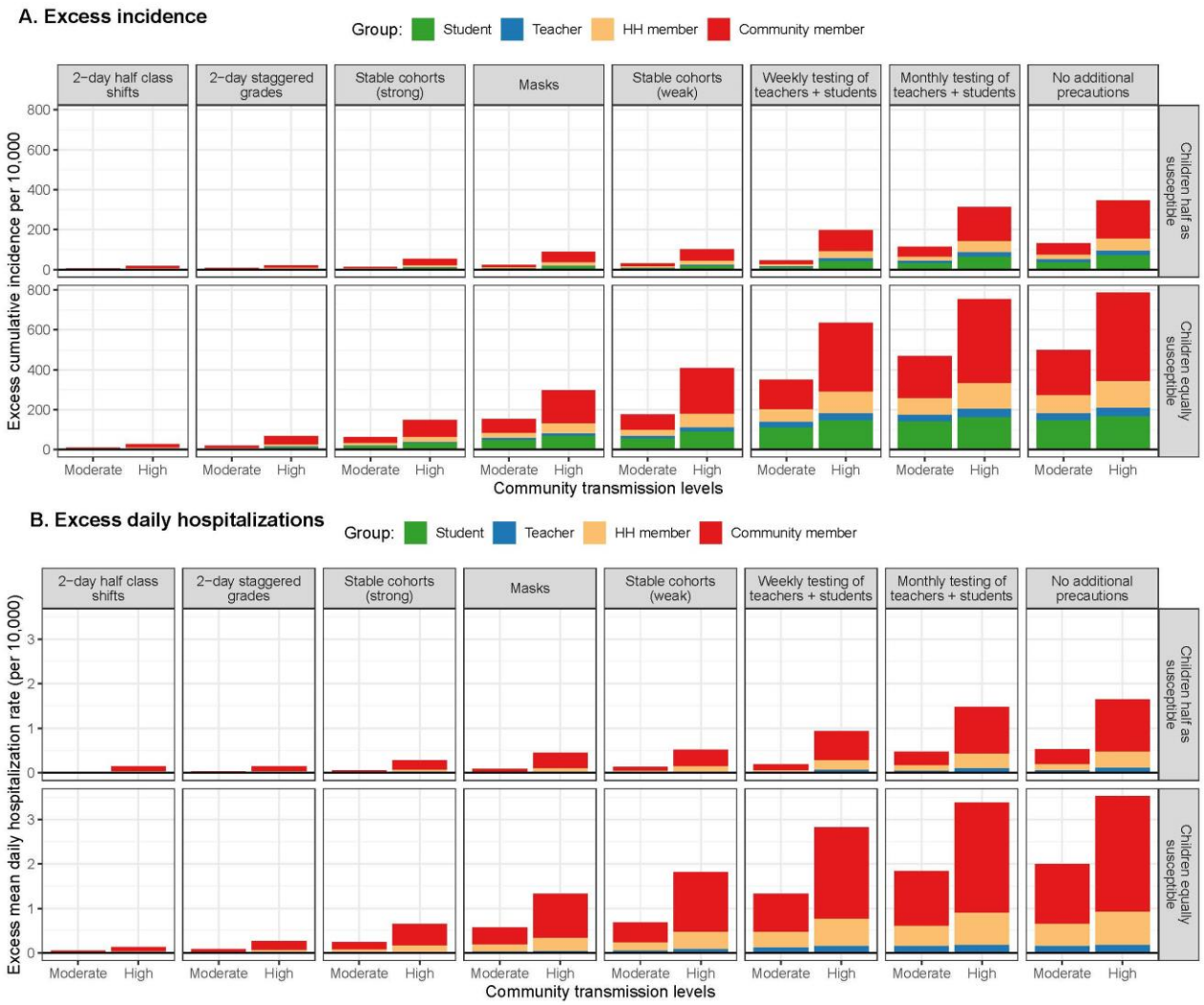


Figure 6. Population level excess incidence and hospitalizations association with reopening strategies over a four-month semester. Excess cumulative incidence per 10,000 (A) and excess daily hospitalization, on average, per 10,000 (B) that would be expected over a four-month semester for each reopening strategy compared to if schools were closed. Bars are stratified by the moderate and high community transmission scenario and colored according to the subgroup contributing cases. Bars are stratified by the moderate and high community transmission scenario and colored according to the subgroup contributing cases. “Teachers” include teachers and all other school staff.

Simulated impact of reopening strategies: more transmission variant, vaccine availability for ages 12+

Effect of within-school precautions under various community vaccination coverages (children under 10 years half as susceptible to infection)

We estimated higher rates of excess illness among elementary and middle school students as compared to high school students across all combinations of NPIs tested (Table 3; Table S14; Fig. 7). Excess illness was also higher among elementary and middle school teachers, as compared to high school teachers, but differences between schooling levels were smaller among teachers as compared to students (Table S14; Fig. 7). Increasing community and school vaccination coverage reduced excess illness attributable to school transmission among all populations, but particularly among the vaccine-eligible population (i.e., teachers and high school students) (Fig. 7), both in the absence and presence of additional NPIs.

Upon achieving a 70% community vaccination coverage or higher (the coverage observed in May 2021 in most Bay Area counties)[181] and without additional NPIs, we estimated the average excess incidence rate as between 8-10 symptomatic cases per 100 students across all age groups (Fig. 7). Expressed as excess cases per school attributable to school transmission, this amounts to an estimated 55 excess cases per high school, 41 excess cases per middle school, and 37 excess symptomatic cases per elementary school across a 128-day semester (Table 3). Tables S10-11 display results for 50% and 60% vaccine coverage. Full results for symptomatic and asymptomatic infection are included in Tables S14-15, and Fig. S6 displays results for vaccine coverages of 80% and 100%.

Table 3. The number of excess student cases attributable to school transmission expected across a four-month (128-day) semester, for 70% community vaccination coverage, which is seen in most Bay Area counties [181]. The mask row is highlighted to demonstrate the current minimum required scenario for schools within the Bay Area.

	Excess student cases attributable to within-school transmission within:			
	380-person **elementary schools** (*half susceptibility*)	380-person **elementary schools** (*equal susceptibility*)	420-person **middle schools**	620-person **high schools**
No precautions	35 cases per school	51 cases per school	41 cases per school	56 cases per school
Universal masking	13 cases per school	36 cases per school	18 cases per school	12 cases per school
Masks + testing	7 cases per school	28 cases per school	10 cases per school	6 cases per school
Masks + cohorts	2 cases per school	9 cases per school	3 cases per school	2 cases per school

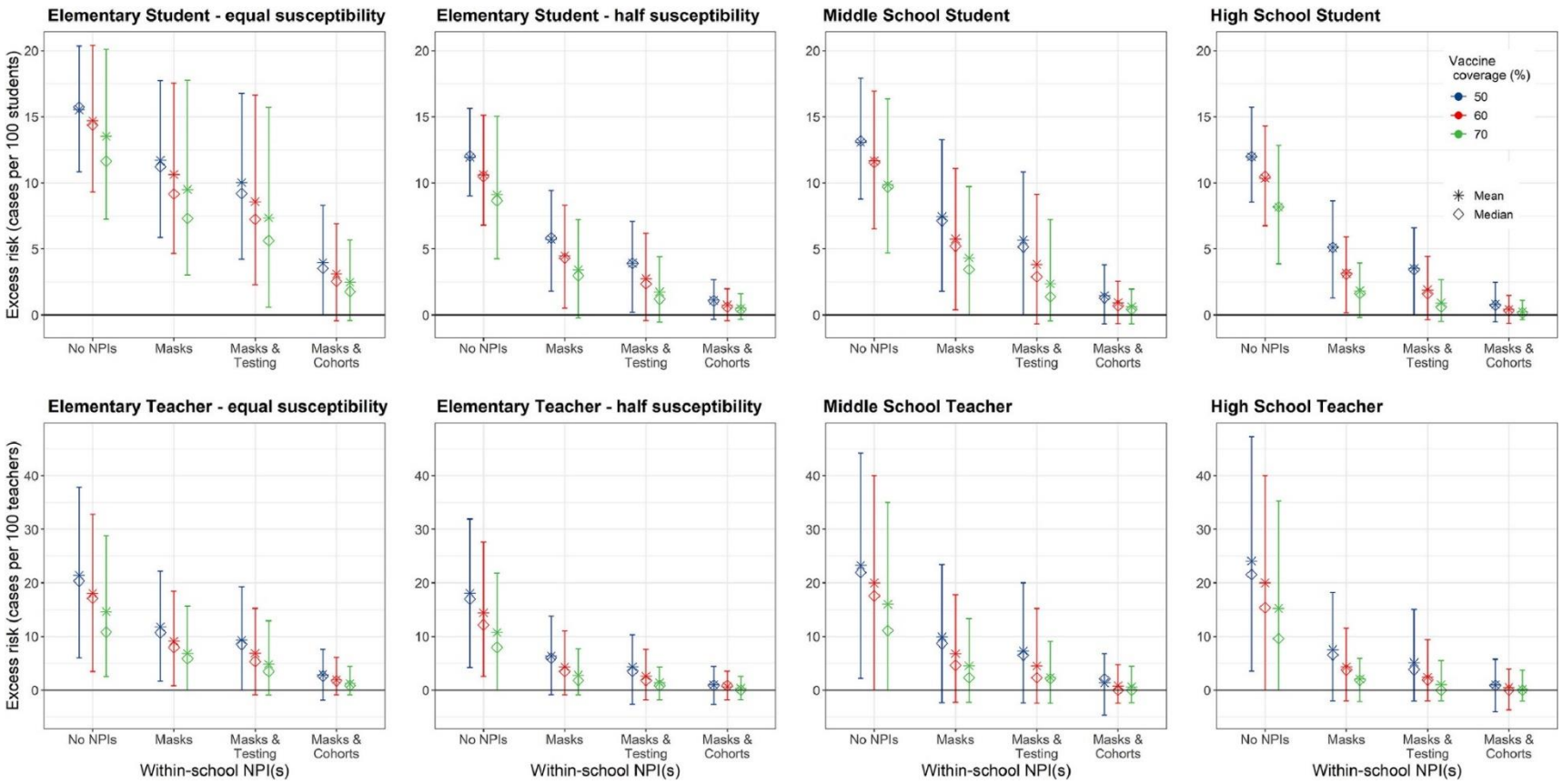


Figure 7. Effect of non-pharmaceutical interventions. We examined the effect of three non-pharmaceutical interventions across three levels of community vaccination coverage (50%, 60%, 70%), assuming that vaccination coverage within school children 12+ and teachers matches that in the community and the vaccine effectiveness is 77% against infection, 85% against symptomatic infection, and 93% against severe infection. Masks indicate universal masks regardless of vaccination status. We calculated the mean (stars) and median (diamonds) of excess cases per 100 persons attributable to school transmission among population subgroups across 1,000 model realizations. Vertical lines reflect the 89thpercentile high probability density interval (HPDI).

Under the most likely reopening scenario for Bay Area schools – dominant circulation of the Delta variant, vaccination coverages of at least 70% and universal masks (Table 3) – we estimated an excess of 13 symptomatic cases per elementary school, 18 cases per middle school, and 12 cases per high school attributable to school transmission over a 128-day semester. This equates to school-attributable illness in an additional 3.4% of elementary school students, 4.3% of middle school students, and 1.8% of high school students owing to school transmission. We estimated that an additional 2.8% of elementary school teachers, 4.5% of middle school teachers, and 2.1% of high school teachers would experience symptomatic infection attributable to school transmission across a semester. Of these symptomatic infections among teachers, 71% were estimated to occur among unvaccinated teachers (Fig. 8A), while 84% of severe infections among teachers were estimated to occur among unvaccinated teachers. Nearly 90% of all infections among children were estimated to occur among unvaccinated children (Fig. 8A). The fraction of cases occurring among the unvaccinated population increased with lower vaccination coverages (Fig. 8B) and decreased with lower vaccine effectiveness (Fig. 8C-D).

While children <12 years remain ineligible for vaccination, increasing vaccination among the community and teachers lowered risk of asymptomatic and symptomatic illness among young children. As simulated community vaccination coverage of the eligible population increased from 50% to 60% to 70%, we estimated that the expected percent of elementary school children with a school-attributable symptomatic illness fell from 11.9% to 10.6% to 9.1%, representing a 23.5% decline in school-attributable transmission. This suggests that adult-to-child transmission represents an important source of school-attributable illnesses (Fig. 7). Under the current reopening plan, the excess rate of symptomatic infection and severe infection among household members of students was estimated to be 1.76 and 1.14 times that of other community members, respectively, suggesting that having a school child in the home would increase the risk of symptomatic infection to household members by 76% and the risk of severe infection by 14% (Fig. 8A; Table S16).

Within-school NPIs were most effective at reducing excess symptomatic cases within elementary and middle schools regardless of levels of community vaccination coverage, and within high schools with lower community vaccination coverages (Fig. 7). For instance, where community vaccine coverage was 50% and no additional NPIs were taken, we estimated an excess incidence of 11.9 cases (89% HPDI: 9.0, 15.6) per 100 students in elementary schools, 13.1 (89% HPDI: 8.8, 17.9) per 100 students in middle schools and 12.0 per 100 students in high schools (89% HPDI: 8.6, 15.7). Adding masks but holding vaccine coverage constant, we estimated an excess incidence of 5.7 cases (89% HPDI: 1.8, 5.9) per 100 elementary students, 7.5 cases (89% HPDI: 1.8, 13.3) per 100 middle school students, and 5.1 (89% HPDI: 1.3, 8.4) cases per 100 high school students.

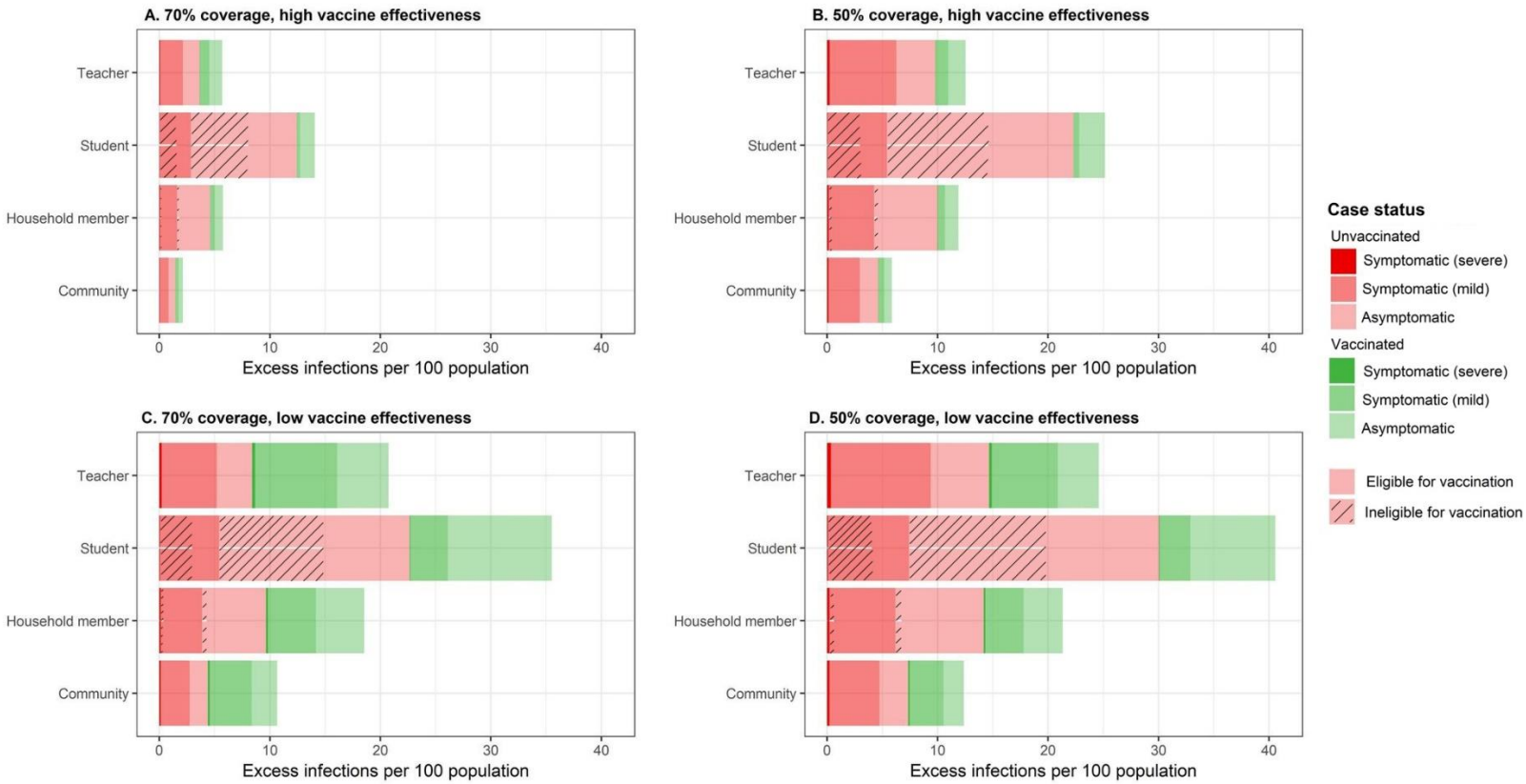


Figure 8. Share of the excess risk by vaccination status and disease outcome, across various vaccine coverages and effectiveness and populations. Red colors reflect the share of excess infections among unvaccinated persons, while green represents excess infection among vaccinated persons. Dark hues represent severe disease (i.e., needing hospitalization), medium hues represent symptomatic but not severe infection, and light hues represent asymptomatic infection. Hashes on the student and the household members represent infected among unvaccinated individuals who are ineligible for vaccination. The high vaccine effectiveness scenario (panels A, B) models vaccines that are 77% effective against any infection, 85% effective against symptomatic infection, and 93% effective against severe infection. The low vaccine effectiveness scenario (panels C, D) models vaccines that are 41% effective against any infection, 45% effective against symptomatic infection, and 49% effective against severe infection.

Effect of increasing vaccination among the school population in the absence of other interventions

We examined under what vaccination coverages, if any, it might be possible to have a return to schooling without any additional NPIs (Fig. 9). Increasing vaccination coverage of the eligible school population from 70% to 95% reduced mean estimates of excess cases among elementary students, suggesting that increasing vaccination coverage among elementary school teachers can reduce the force of infection among their students. For instance, increasing the vaccination coverage of the eligible school population (here, teachers) from 70% to 95% reduced the estimated excess rate of infection from 9.1 (89% HPDI: 4.3, 15.0) to 5.7 (89% HPDI: 0.2, 12.5) symptomatic cases per 100 elementary students across the four-month semester, representing a reduction of 37%. At the same time, increasing vaccination of teachers/staff from 70% to 95% reduced the estimated excess rate of infection among elementary teachers from 10.8 (89% HPDI: 0, 21.8) to 2.9 (89% HPDI: 0, 7.0) symptomatic cases per 100 teachers across the four-month semester, representing a reduction of 73%.

While increasing within-school vaccine coverage indirectly reduced infections among elementary and middle school students, the effect of increasing within-school vaccination coverage was most pronounced among high school students and teachers of all grade levels. Compared to other schooling levels, high school teachers and students achieved the lowest rates of infection attributable to school transmission using vaccination only without NPIs (Table 4). At 70% coverage of the eligible school population, we estimated an excess of 8.2 (89% HPDI: 3.9, 12.8) symptomatic cases per 100 high school students and 15.2 (89% HPDI: 0, 35.3) per 100 teachers across the 128-day semester, and at 95% coverage an excess of 2.7 (89% HPDI: 0, 5.4) cases per 100 students and 2.4 (89% HPDI: 0, 7.7) per 100 teachers across the 128-day semester (Fig. 9).

Interventions required to reduce incidence attributable within schools below certain risk tolerances

We examined whether layering NPIs or increasing within-school vaccination could reduce incidence attributable to school transmission below specific risk tolerances (Table 4). We estimated that universal masking and 70% community and within-school vaccination coverage or higher could reduce the number of excess cases attributable to school transmission to <50 per 1,000 students and teachers across all grade levels. In high school students, increasing the vaccine coverage among the vaccine-eligible school population above 80% could also reduce excess transmission to <50 per 1,000 students and teachers in the absence of NPIs. However, achieving lower risk levels among elementary school students—e.g., <10 cases per 1,000 students or teachers—required additional NPIs, such as testing or cohorts, and was not achievable through the NPIs investigated here if children under 10 years are equally as susceptible as adults. On a per school basis, reducing the excess cases attributable to school transmission to fewer than two cases per school across the full semester (i.e., <50% probability of a case per school per month) required both masks and cohorts. Tables S12-13 display the minimum NPIs required to achieve the various risk tolerances assuming 50% and 60% vaccine coverage in the eligible community, respectively.

Table 4. The minimum non-pharmaceutical intervention(s), or minimum within-school vaccination coverage of the eligible population, needed to reduce the risk of symptomatic infection to beneath a given risk level (e.g., 50 cases per 1,000 population), assuming that 70% of the vaccine-eligible community has received a vaccine at 85% effectiveness. 'Not observed' indicates that no combination of interventions *examined in this study* reduced excess risk beneath the indicated threshold. Masks refers to universal masking regardless of vaccination status.

		Population-wide risk tolerance — symptomatic cases per 1,000 population			**School-based risk tolerance — < 2 cases per school***
		<50	**<25**	**<10**	
Students	Elementary school – *half susceptibility*	Masks	Masks + testing	Masks + cohorts	Masks + cohorts
	Elementary school – *equal susceptibility*	Masks + cohorts	Masks + cohorts	Not observed**	Not observed**
	Middle school	Masks or 95% coverage	Masks + testing	Masks + cohorts	Not observed**
	High school	Masks or 80% coverage	Masks	Masks + testing	Masks + cohorts
Teachers/staff	Elementary school – *half susceptibility*	Masks or 80% coverage	Masks + testing	Masks + cohorts	
	Elementary school – *equal susceptibility*	Masks + testing	Masks + cohorts	Not observed**	
	Middle school	Masks or 90% coverage	Masks + testing	Masks + cohorts	
	High school	Masks or 80% coverage	Masks or 95% coverage	Masks + testing	

*Assuming a 380-person elementary school, 420-person middle school, and 680-person high school

**Not observed under the interventions examined here

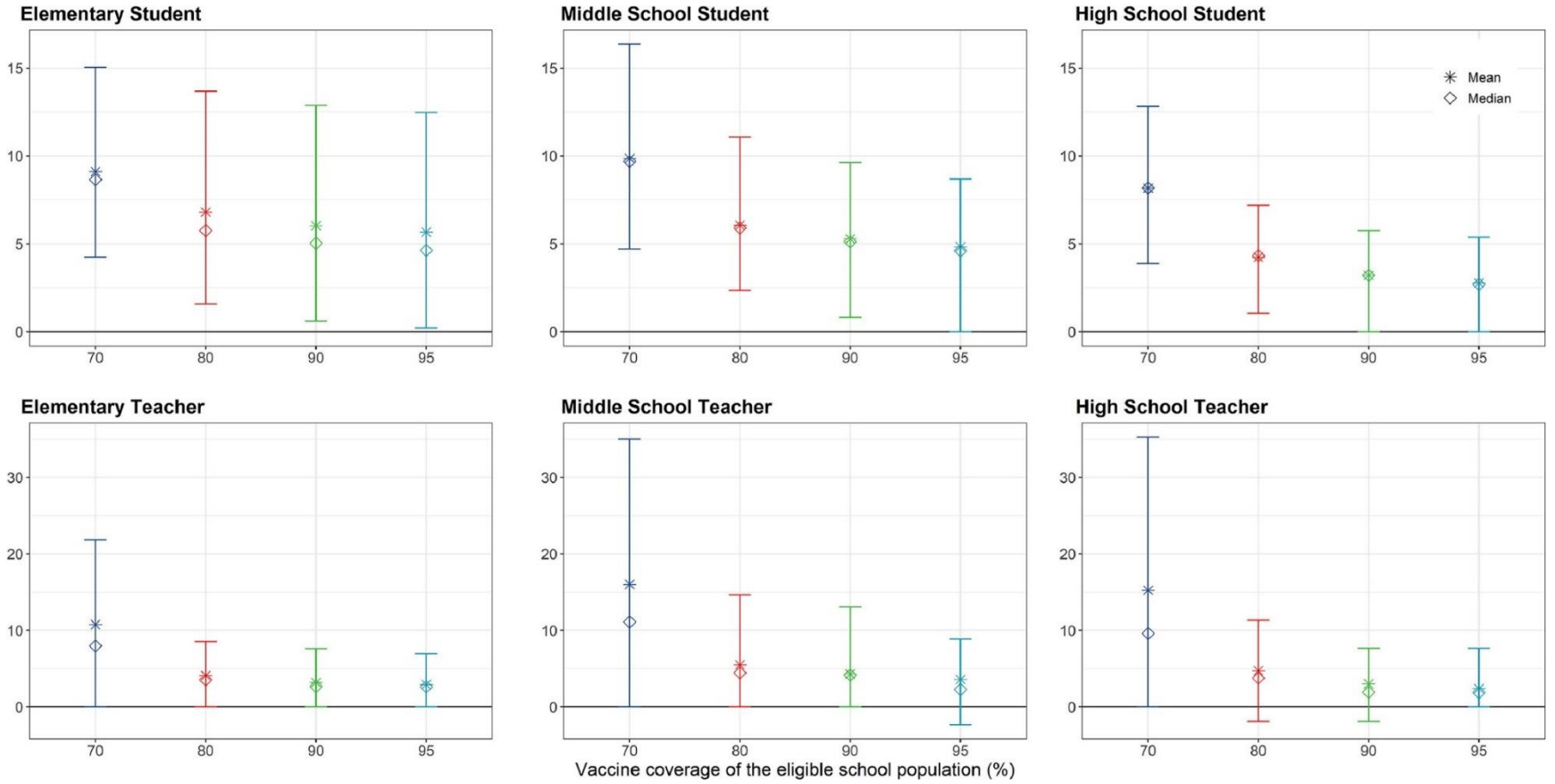


Figure 9. Effect of increasing within-school vaccination coverage. We examined the effect of increasing vaccination coverage among school populations, in the absence of additional non-pharmaceutical interventions, and holding community and within-school vaccination coverage of the eligible (12+) population at 70%. We calculated the mean (stars) and median(diamonds) of excess risk per 100 persons attributable to school transmission among population subgroups across 1,000 model realizations. Vertical lines reflect the 89thpercentile high probability density interval (HPDI).

Effect of masking all individuals in a school compared to masking only unvaccinated individuals

We compared the differences in school-attributable transmission under scenarios where only unvaccinated individuals wore masks compared to if all individuals masked, across different levels of vaccine effectiveness (VE), assuming 70% of the eligible population is fully vaccinated (Fig. 10). Since all elementary students are unvaccinated, such a rule would change behaviors only among the vaccinated teachers, about 5% of the overall school population. In contrast, such a rule would affect the entirety of the vaccinated high school population, both students and teachers, about 70% of the overall school population. The difference between masking the entire student and teacher population as compared to only the unvaccinated school population is thus most apparent in middle and high school populations, and at lower VEs. For instance, given 45% VE, masking all middle and high school students and teachers would avert symptomatic infection for 4.0% of middle school students, 5.4% of high school students, 1.2% of middle school teachers, and 14.6% of high school teachers compared to masking only unvaccinated students and teachers. At 85% VE, masking all students and teachers would avert symptomatic infection for 1.4% of middle school students, 1.7% of high school students, 3.0% of middle school teachers, and 3.4% of high school teachers compared to masking only unvaccinated students and teachers.

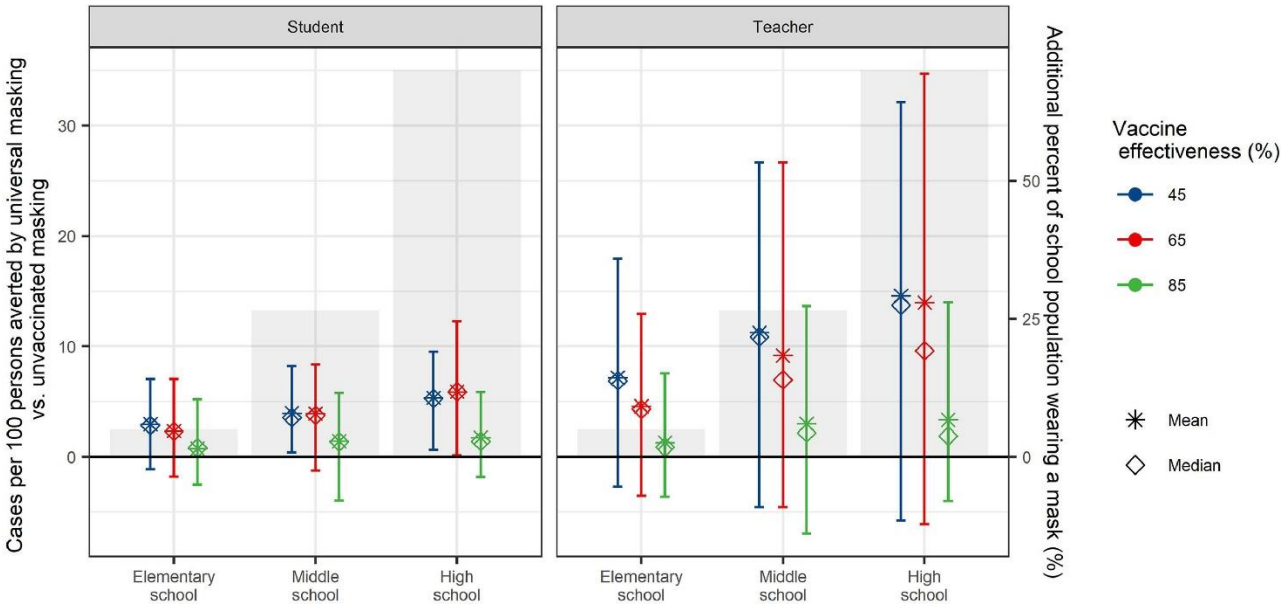


Figure 10. Effect of universal masking compared to masking only of unvaccinated individuals. We estimated the additional cases averted in each population by masking the entire student and teacher population, compared to masking only the unvaccinated student and teacher population, in the absence of additional interventions. We held community and within-school vaccination coverage of the eligible (12+) population at 70%, and varied vaccine efficacy (VE). We calculated the mean (stars) and median(diamonds) of excess risk per 100 persons attributable to school transmission among population subgroups across 1,000 model realizations. Vertical lines reflect the 89thpercentile high probability density interval (HPDI). Shaded bars and right axis reflect the vaccinated percent of the school population, for whom a universal masking rule as compared to a masking rule among the unvaccinated would apply.

3.5 Discussion

Gaps in our understanding of contact patterns among U.S. schoolchildren have limited previous efforts to estimate the effect of school closures on COVID-19 transmission in a community of demographically heterogeneous households. We found evidence of a higher average community contact rate among lower income and Hispanic children during shelter-in-place orders, consistent with literature demonstrating limited ability of low-income communities to shelter-in-place [184], which contributes to the disproportionately high incidence and mortality rates among low-income or Hispanic communities [185]. Differences in total contacts between Hispanic and non-Hispanic respondents were driven by working-aged adults (18-65 years) and young children (0-12 years). As Hispanic individuals make up a disproportionate number of essential workers in the Bay Area [174], these findings may reflect both contacts at work and childcare. Indeed, while our survey found higher contact rates in elementary students as compared to high school students, social mixing data during non-epidemic periods report higher community contact rates among high school students [175]. Elementary students may have more limited ability to shelter-in-place than high school students due to accompanying family members during essential activities and requiring daycare.

In the March 17 - June 1 spring 2020 semester period, we estimated that school closures averted 13,842 confirmed cases, and 663 deaths in the Bay Area. Under the lowest risk scenario examined, we found that reopening for a four-month semester without any precautions would increase risk for students (an additional 3.0% of students across all grades levels infected over the four-month reopening period), family members of students (an additional 1.4% infected), and especially teachers/staff (an additional 10.3% across all grade levels). Our results are consistent with other models that project large increases in transmission due to in-person schooling conducted with no safety measures, with substantial reductions in school-attributable transmission possible when within-school and community intervention measures are in place [133, 134, 136, 186, 187]. Our results are also consistent with empirical evidence showing high transmission among a summer camp where children interacted in large cohorts [188], high seroprevalence among teachers and students from a high school setting with limited safety measures [148], moderate transmission among teachers from schools with rare face mask use and some social distancing [149, 189], and low transmission in schools that adopted a cohort or hybrid system, masks, or TTI protocols [137, 190].

Some reopening strategies can result in few in-school transmissions among students and teachers alike, according to our findings. Most notably, our model found that, in absence of a vaccine, reducing in-school mixing via classroom cohorts or hybrid scheduling is an effective means of reducing risk of school-attributable illness across all levels of education, especially when combined with universal masking. These findings concur with observations of schools that reopened with universal masking, social distancing, and a hybrid or cohort approach and avoided large outbreaks [135, 137, 190]. While we find that high community transmission increases the risk of within-school transmission across all measures studied, the influence of community transmission is minimized as the effectiveness of the within-school measures increases. Our findings therefore support the most recent CDC guidance, which states that

community transmission rates are important to monitor when planning for the reopening of schools, but the essential elements for reopening are implementation of within-school measures—masks, physical distancing, handwashing, and contact tracing—with priority given to masks and distancing [191]. We found that if the prioritized essential elements of masks and physical distancing via a cohort or hybrid system are not met, outbreaks are plausible. Under such scenarios with minimal within-school interventions, community interventions (e.g., workplace closures and reductions in social gatherings) play a larger role in moderating within-school transmission. This is consistent with outbreaks documented in childcare settings that lack of safety precautions [132, 188, 192], and reports from the UK that risk of outbreaks in schools without mask requirements increased with community transmission levels [139, 149]. We found that teachers and staff would bear a disproportionate burden of infection if an outbreak occurred, in agreement with available data on school transmission [149, 189]. It is thus essential to ensure that specific precautions are available to support this population, including safe spaces for lunch breaks, virtual faculty meetings, and financial and logistical support if quarantine is needed.

We find that, in absence of a vaccine, reducing risk of school-attributable illness to below 1% in each population sub-group is most feasible in elementary schools (using, for instance, masks and stable cohorts). The idea that elementary schools pose lower transmission risk than high schools is widely supported [29], both from modelling studies [133, 187, 193], as well as empirically [147, 148, 192]. For one, high school environments have larger student, teacher and staff populations. Even if younger children are as susceptible as older children, we estimate that reopening high schools without precautions yields an estimated 3-5 times more risk of symptomatic infection to teachers/staff when compared with reopening of elementary schools, depending on the level of community transmission. If susceptibility increases with age, as some evidence suggests [141, 163, 194], we estimated that high school teachers may experience as much as 5-10 times greater risk of symptomatic infection when compared with elementary school teachers, depending on the level of community transmission. These findings agree with empirical data from Sweden, which found that risk to teachers increased with student age [189].

Vaccination is recognized by the CDC and CDPH as the leading public health strategy for reducing within-school transmission [195, 196], and our results highlight that increased vaccination coverage—both among the general community and among the eligible school population—plays an essential role in limiting symptomatic illness attributable to school transmission. However, even in the presence of a vaccine, in the midst of Delta variant circulation, our findings support the use of universal masks as precaution within schools, particularly elementary and middle schools, but also high schools that have within-school vaccine coverage <90%. Masks are supported as one of the simplest, yet effective, mitigation strategies [195-197]. Masking is of particular importance for elementary and many middle school students who remain ineligible for vaccination; inadequate mask use has been implicated in school-based transmission in the United States and elsewhere [197-200].

Increased vaccine coverage of community members and teachers helped reduce illness among children not yet age-eligible for vaccination. We estimated that increasing vaccination coverage of the general population reduced the excess risk of transmission by 24% among elementary students. Similarly, we estimated that increasing vaccination coverage among teachers from 70% to 95% reduced the excess risk of school transmission by 37% among elementary students This suggests that teacher-to-student transmission is an important route of transmission that can be eliminated by increased vaccination. This finding agrees with conclusions from recent epidemiological investigations of school-associated outbreaks. In Georgia, investigation of 31 cases across six elementary school populations found that two outbreaks were initiated by teacher-to-teacher transmission, followed by teacher-to-student transmission, accounting for nearly 50% of the cases [198]. Similarly, outbreaks at three child care facilities in Utah were linked to adult index cases [146], and a large, prospective study in England found that staff-to-staff transmission and staff-to-student transmission were responsible for initiating 50% and 23%, respectively, of 30 confirmed school outbreaks [149]. At the same time, child-to-adult transmission is documented as well [146, 198, 201]. While reduced frequency and severity of symptoms in children may correspond to lower infectivity, viral load in children has been found to be equal to that in adults after controlling for symptoms [202].

Limitations

The age-structured contact rates from the Bay Area are similar to those captured from households with children from other major cities, including New York, Atlanta, Phoenix, and Boston [128]. However, extrapolation of contacts rates requires caution because the Bay Area differs from the broader United States in several dimensions: higher household income, higher educational attainment, larger workforce, smaller household sizes, smaller proportion of African Americans, and higher compliance with social distancing [203]. During the spring semester, the Bay Area had a higher proportion of essential workers than the national average [174], which could translate into a larger impact of workplace closures non-Bay Area cities. As we demonstrated, the impact of school closures varies by the proportion of households that have school aged children, as well as the average school and class size of local public schools. Accordingly, the risk associated with school-based transmission will be higher in cities with a greater proportion of school aged children, as well as larger school or classroom sizes. Nevertheless, many findings pertaining to school reopening are generalizable—such as teachers experiencing the greatest risks; high schools being at higher risk than elementary schools; high community transmission increasing risk in the absence of safety measures put in place; and the relative ranking of interventions. After all, key epidemiologic parameters (e.g., susceptibility of children, asymptomatic transmission, mask effectiveness) apply across locations, and several population-level parameters (e.g., household size) apply to other urban areas.

Selection bias in our survey is possible because it was administered in English, and respondents were less likely to be essential workers. Discrepancies observed in the number of contacts by work location (outside vs. inside the home) and ethnicity (Hispanic vs. non-Hispanic) are thus expected to be biased towards the null. Our sample does not capture contact patterns among and between adults who do not have children, particularly missing those of young adults (18-

29) or older adults (65+). However, our results are similar to estimates captured in another Bay Area contact survey that targeted households with and without children [128].

Community contacts under modelled school closure scenarios account for increases in daycare contacts only at the rates observed in our community survey, when fewer adults were permitted to work in-person. Therefore, modelled school closures or staggered weeks while reopening for a subsequent four-month semester may not adequately account for increases in community contacts from daycare settings. Similarly, the attributional effect of school reopening does not account for increases in workplace transmission that may occur if working parents return to in-person work once their child's school resumes in-person instruction

All of our modelled estimates depend, in part, on imperfectly understood epidemiologic parameters, such as the relative susceptibility of children [141, 194] and transmissibility of asymptomatic individuals [141, 169, 194]. We compare modeling results across various assumptions of each but contact tracing studies that seek to capture the relative susceptibility and infectiousness of symptomatically and asymptomatically infected children across ages are urgently needed.

While our model accounts for isolation of symptomatic individuals and quarantine of household members, modeled community interventions do not necessarily include the full effects of population level contact tracing. However, based on modelled estimates of the effect of contact tracing used by the Bay Area over this period, we do not expect our conclusions about school closures would change substantially if accounting for this [166, 204]. While we found large reductions in risk with mask use and physical distancing, modeled within-school interventions did not include infection control measures, such as improved ventilation, increased handwashing, desk spacing, or reduced sharing of supplies, which may further reduce transmission. Based on conversations about feasibility with school districts, we chose to model a periodic test-trace-isolate (TTI) intervention, in which testing was conducted on a monthly or weekly basis, rather than reactively based on symptom presentation. Other studies have demonstrated that reactive TTI can prevent a second transmission wave caused by school reopening [136].

3.6 Conclusion

Given the myriad individual and societal consequences of school closures, policymakers must urgently dedicate resources to support the package of interventions necessary to mitigate risk in schools. Vaccination remains the most effective and sustainable means of risk reduction and efforts should focus on increasing vaccination coverage among the eligible community members and school population. Among populations not yet eligible for vaccination and communities with lower vaccination coverage, prevention measures, such as masking, may be needed to reduce the risk of school outbreaks. Schools may consider layering testing or cohorting as additional safety measures, particularly as the Delta variant takes hold.

Acknowledgments

We thank Dr. Jon Krosnik for advice regarding survey questionnaire development. We thank Ms. Sarah Vidmar, Ms. Anna Johnson, Ms. Shelley Facente, and Dr. Stephanie Holm for assistance in distributing the survey.

Data Accessibility

Data and code for this analysis can be found at: https://github.com/jrhead/COVIDandSchools

CHAPTER 4. Early evidence of inactivated enterovirus 71 vaccine impact against hand, foot, and mouth disease in a major center of ongoing transmission in China, 2011-2018: a longitudinal surveillance study[3]

[3] Chapter 4 is included with permission from co-authors Philip A. Collender, Joseph A. Lewnard, Nicholas K. Skaff, Ling Li, Qu Cheng, Julia M. Baker, Charles Li, Dehao Chen, Alison Ohringer, Song Liang, Changhong Yang, Alan Hubbard, Benjamin Lopman, Justin V. Remais

4.1 Abstract

Background

Enterovirus 71 (EV71) is a major causative agent of hand, foot, and mouth disease (HFMD), associated with severe manifestations of the disease. Pediatric immunization with inactivated EV71 vaccine was initiated in 2016 in the Asia-Pacific Region, including China. We analyzed time series of HFMD cases attributable to EV71, coxsackievirus A16 (CA16), and other enteroviruses in Chengdu, a major transmission center in China, to assess early impacts of immunization.

Methods

Reported HFMD cases were obtained from China's notifiable disease surveillance system. We compared observed post-vaccination incidence rates during 2017-18 with counterfactual predictions made from a negative binomial regression and a random forest model fitted to pre-vaccine years (2011-15). We fit a change point model to the full time series to evaluate whether the trend of EV71 HFMD changed following vaccination.

Results

Between 2011-18, 279,352 HFMD cases were reported in the study region. The average incidence rate of EV71 HFMD in 2017-2018 was 60% (95% prediction interval (PI): 41%–72%) lower than predicted in the absence of immunization, corresponding to an estimated 6,911 (95% PI: 3,246, 11,542) EV71 cases averted over two years. There were 52% (95% PI: 0.42, 0.60) fewer severe HFMD cases than predicted. However, the incidence rate of non-CA16 and non-EV71 HFMD was elevated in 2018. We identified a significant decline in the trend of EV71 HFMD four months into the post-vaccine period.

Conclusions

We provide the first real-world evidence that programmatic vaccination against EV71 is effective against childhood HFMD and present an approach to detect early vaccine impact or unintended consequences from surveillance data.

4.2 Introduction

Enterovirus 71 (EV71) is a major causative agent of hand, foot, and mouth disease (HFMD), associated with severe manifestations of the illness. HFMD causes substantial burden in the Asia-Pacific region, including China, which reported over 20 million cases between 2008-2018 [37]. Between 2004-2013, HFMD had the highest incidence of any infectious disease in China, and was the leading cause of death for children under five years old amongst all 39 notifiable infectious diseases [43, 205]. The most prevalent enterovirus serotypes causing HFMD in China are EV71 and coxsackievirus A16 (CA16), with EV71 implicated in 70% of severe cases and 92% of deaths [38-41]. In China, EV71 cost an estimated $180-$330 million USD in 2016 [206], and caused over 3,000 deaths between 2008-2017 [41].

In December 2015, a monovalent, inactivated whole-virus vaccine against EV71 developed by the Institute of Medical Biology, Chinese Academy of Medical Sciences (CAMS) was licensed in China [207, 208]. Another was licensed in January 2016, by Sinovac Biotech Co. (Beijing) [209], and a third in late 2016 by Vigoo Biological Co. (Beijing) [210]. In phase 3 clinical trials among healthy children 6-35 months, vaccine efficacy exceeded 90% against EV71-associated HFMD after one dose and 98.8% after two doses [208, 209, 211]. In 2016, the vaccine was gradually made available for a fee at healthcare centers in China, targeting children 6-59 months. For best efficacy, two doses are encouraged, one month apart, starting at 6 months of age.

Despite promising clinical trials, the real-world impact of China's vaccination program remains unknown. In 2017-2018, after vaccine introduction, Beijing and Chengdu reported outbreaks of all-cause HFMD [212, 213]. Infection with EV71 or CA16 is thought to confer lifelong serotype-specific immunity [38], as well as transient (~7 weeks) cross-serotype protection [42]. Given that the vaccine is not expected to confer cross immunity [42, 43], increases in HFMD incidence due to non-vaccine serotypes post vaccine introduction may indicate serotype replacement [35, 36].

Here, we estimate the effect of EV71 vaccine introduction on reported incidence of HFMD due to EV71 and non-EV71 serotypes in Chengdu, an urban prefecture in Sichuan Province with incidence that is consistently among the highest in southwest China [214]. The introduction of the EV71 vaccine in 2016, combined with high-quality surveillance data, make Chengdu a vital setting in which to study early evidence of the impact of EV71 vaccination, and its potential benefits in major Asian population centers with similarly high EV71 burdens.

4.3 Methods

Epidemiologic and covariate data

The study region of Chengdu Prefecture in southwest China contains over 15 million people in an area of 14,380 km^2, making it the 5th most populous Chinese metropolitan region [215]. HFMD cases in China are highly seasonal, with transmission beginning in urban centers and spreading to the urban-rural interface [216]. As Chengdu is surrounded by rural prefectures, the transmission dynamics of HFMD there have implications for Sichuan Province [214].

HFMD cases in Chengdu reported between Jan 1, 2011 and Dec 31, 2018 were obtained from China's National Infectious Disease Reporting System (NIDRS)—a passive electronic surveillance system covering almost all Chinese healthcare facilities [217]. Upon diagnosis, all HFMD cases are required to be reported to NIDRS within 24 hours.

HFMD cases were defined as a patient presenting with papular or vesicular rash on hands, feet, mouth, or buttocks [218]. The distinctness of clinical features of HFMD and differences in the epidemiology, seasonality, and progression of the differential diagnoses limits concerns of clinical misdiagnosis [219, 220]. To identify the proportion of serotypes in circulation, serotypes were determined by reverse-transcriptase polymerase chain reaction (RT-PCR), real-time RT-PCR, or virus isolation for 9,879 (3.5%) of patients who met the HFMD case definition [218]. Serotype determination was more common during certain years, at certain facilities, and for severe cases.

Year-end, age-stratified population data were obtained for 2011-2017 [215], and projected for 2018 based on prior years' population using splines. Data on the number of doses given at healthcare facilities in Chengdu were collected by the Sichuan Center for Disease Control and Prevention immunization information system, which tracks daily vaccinations provided at each inoculation point throughout the province, and their manufacturer. The majority (~80%) of vaccines administered in Chengdu were produced by CAMS, with the remaining by Sinovac (Figure S1).

Statistical Analysis

Construction of time series adjusted for probability of serotype testing

Because a non-random subset of cases underwent serotype determination, we constructed adjusted time series for HFMD cases caused by EV71, CA16, and other (non-EV71 and non-CA16) enteroviruses via inverse probability weighting. First, we regressed a binary indicator for receiving laboratory testing against month of diagnosis, case severity (binary, defined as suffering cardiopulmonary or neurological complications [38, 221]), age group (0-4 years, 5-9 years, 10+ years), sex, and a linear trend on year stratified by severity (Supplemental Text 1). We weighted each tested case by the inverse of its model–predicted probability of being tested to reflect the total underlying number of cases with the same etiology. Predicted probabilities closely matched the observed data (Figure ST1.1). We compiled the adjusted monthly time series of EV71, CA16, and other HFMD cases by summing these weights, representing the most probable serotype-specific case counts.

Establishing counterfactual expectations and estimating vaccine impact

We took parametric and non-parametric approaches to estimate the incidence of HFMD due to EV71, CA16, and other etiologies as well as severe HFMD incidence expected in the absence of vaccination. In both approaches, we fit four separate predictive models to the adjusted time series of serotype-specific and severe cases from the pre-vaccine period (2011-2015), and used

the fitted models to predict incidence rates in the post-vaccine period (2017-2018). Following others' work [222-224], we considered 2016 a transitional period and excluded it from analysis.

Candidate predictive generalized linear models (GLMs) assumed a negative binomial distribution (Supplemental Text 2) and included an offset for yearly population, as well as combinations of terms for age-specific fixed-effects; seasonality (monthly fixed effects or 6 and/or 12 month harmonic terms); population immunity (all-cause or serotype-specific incidence in the prior year); biennial periodicity (24 month harmonic term); and yearly trend (linear or cubic spline). We applied blocked cross validation within the pre-vaccine period, selecting the model for each outcome with the lowest out-of-sample mean squared error across holdout years (Supplemental Text 3) [225, 226]. We compared model-predicted, expected rates in the post-vaccine period to observed rates and calculated incidence rate ratios (IRR) as the sum of the observed cases over the sum of the expected cases. We calculated 95% prediction intervals (PIs) via 1,000,000 Monte Carlo simulations using parameter estimates from the fitted models. We propagated uncertainty year by year for models that included an autoregressive term capturing last year's incidence, using the distribution of simulated incidence in the prior year as inputs to predictions in the current year.

To reduce concerns surrounding GLM misspecification, we performed a sensitivity analysis using random forest regression ("ranger" package in R, v.3.5.1), a non-parametric machine learning algorithm, to predict monthly serotype-specific HFMD rates, using the same candidate covariates we did for the GLMs [80] . Predictions from each tree were averaged to generate an ensemble estimate. We adapted bootstrap sampling routine to randomly exclude a year of pre-vaccine data during training of each tree to ensure generalizability to unobserved years [227]. The 95% prediction intervals were calculated from the .025 and 0.975 quantiles of the estimated conditional distribution of the response variable [228].

As an alternative approach to estimating vaccine impact, we fit a change point model to the full time series to determine whether vaccine introduction preceded a significant change in the epidemiologic trend. Unlike interrupted time series analysis, change point methods do not pre-assume the point at which changes in incidence occur but rather assesses whether changes in incidence trends follow, and are thus likely attributable to, vaccine implementation [229, 230]. To identify the most likely change point, we fit negative binomial GLMs for EV71 HFMD incidence, incorporating harmonic terms for seasonal and biennial periodicities, as well as pre- and post-change point yearly linear trends (Supplemental Text 2). We assessed each month in the full time series as a candidate change point, and selected the model with the highest log-likelihood. We generated counterfactual predictions of post-change point incidence rates by assuming a continuation of pre-change point trends. Prediction intervals were estimated as before.

4.4 Results

Descriptive

Between January 1, 2011 and December 31, 2018, 279,352 HFMD cases were reported in Chengdu Prefecture, resulting in 19 deaths (0.2%) and 2,430 severe cases (0.9%) (Table 1, Table S1 for yearly breakdown). The median age was 5 years (range 0-84) and 58% of cases were male. Of 9,879 (3.5%) tested cases, 2,503 (20.8%) were positive for EV71 and 1,859 (18.8%) were positive for CA16. Among cases with known etiology, EV71 was associated with 48% of severe cases and 85% of deaths. Across Chengdu, immunizations began in June 2016, with the rate of vaccination increasing monthly until stabilizing in July 2017 (Figure S1). Over this period, 880,673 doses were administered, corresponding to a coverage of 54.3% of the eligible 811,700 children if all vaccinated children received two doses.

The adjusted time series of monthly incidence of EV71, CA16 and other enteroviruses exhibits seasonal peaks following six, twelve, and 24-month periodocities (Figure 1). Prior to 2016, EV71 and CA16 alternated years as the most prevalent known serotype, with the proportion of EV71 cases declining post vaccine introduction and the proportion of CA16 cases remaining stable (Figure 1). Annual oscillations are less clear for the proportion of non-CA16 and non-EV71 cases. The peak incidence rate of EV71, observed in 2016, was followed by a substantial decline, and 2018 recorded both the fewest annual cases of EV71 in the study period and the highest annual cases of non-CA16 and non-EV71 HFMD.

Table 5. Distribution of HFMD incidence by sex, age, severity, and serotype in Chengdu, China, 2011-2018

		Cases with known serotypes (N = 9,879)		
	All cases	**EV71**	**CA16**	**Other enteroviruses**
Total, N (%)	279,352 (100)	2,053 (20.8)	1,859 (18.8)	5,967 (60.4)
Male, N (%)	162,757 (58.3)	1,231 (20.3)	1,134 (18.8)	3,670 (60.8)
Age, Median (Range)	5 (0-84)	7 (0 - 36)	6 (0 - 35)	5 (0 - 38)
Severe, N (%)	2,430 (0.88)	639 (46.6)	115 (8.4)	616 (45.0)
Deaths, N (%)	19 (<0.01)	12 (85.7)	0 (0)	2 (14.3)

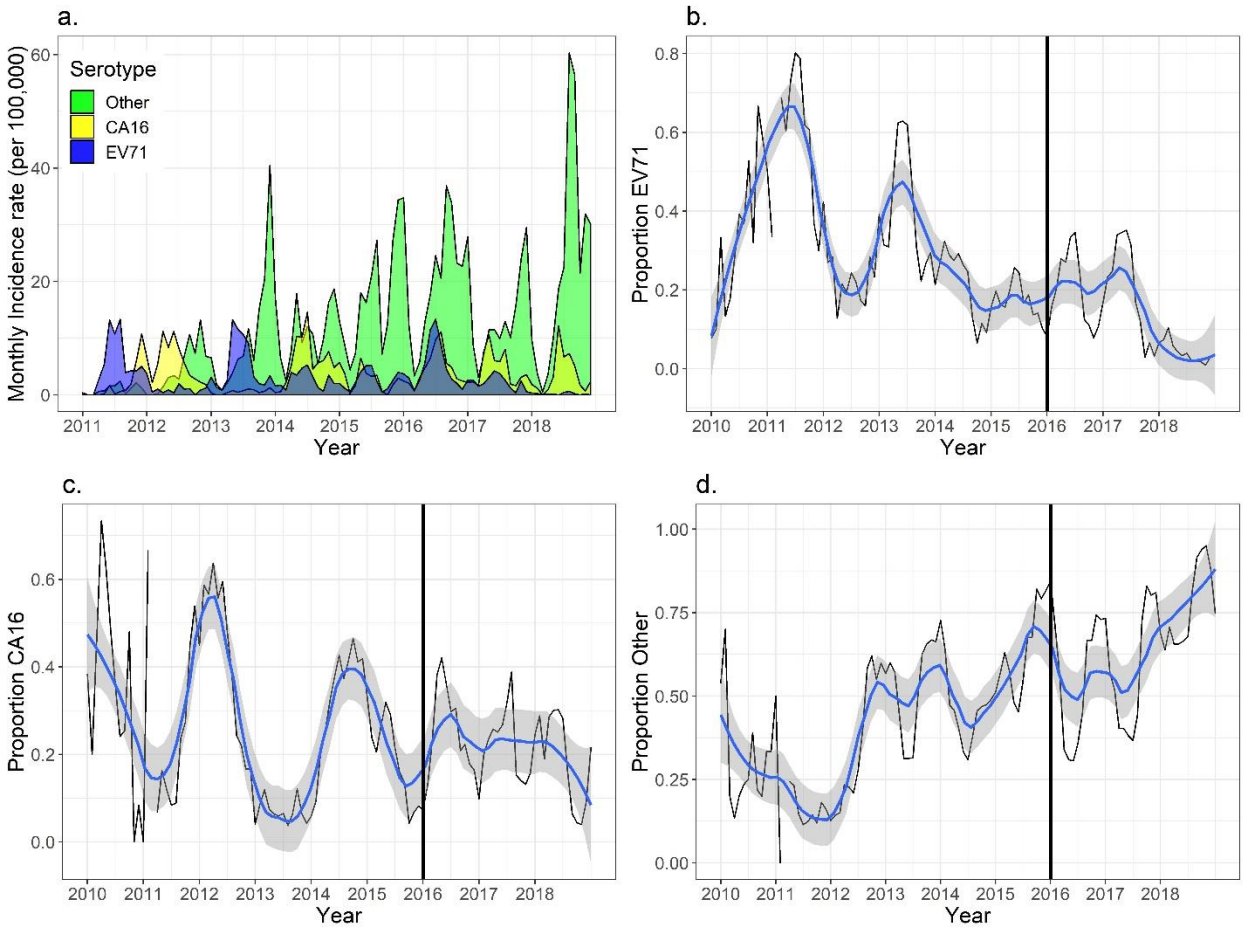


Figure 1. a) Monthly incidence rate per 100,000 of adjusted serotypes for HFMD in Chengdu, China, 2011-18. Adjustment involved calculating a weighted sum of cases for each serotype determination, where weights were equal to the inverse probability of having been tested. b) The proportion of EV71 cases that make up all HMFD cases; c) The proportion of CA16 cases that make up all HMFD cases; d) The proportion of non-CA16 and non-EV71 cases that make up all HMFD cases.

Estimated vaccine impacts on EV71 HFMD and severe HFMD

During the two years following EV71 immunization in Chengdu (2017-2018), EV71 incidence rates were lower than expected based on predictions from GLM and random forest models fit to pre-vaccine data, as well as the best change point model. The IRR of observed EV71 HFMD over the expected incidence based on the best predictive GLM was 0.40 (95% PI: 0.28, 0.59) (Table 2; Figure 2). A cumulative total of 6,911 (95% PI: 3,246, 11,542) fewer EV71 cases were reported than expected over 2017-2018, with 5,664 (95% PI: 2,186, 11,059) averted cases in 2018. The random forest model yielded similar results, estimating a total of 8,314 (95% PI: -1,942, 25,427; IRR: 0.36, 95% PI: 0.15, 1.73) fewer EV71 cases reported than expected over 2017-2018, with 4,622 (95% PI: 182, 12,266) averted cases in 2018 (Table S2; Figure S2).

Similarly, the IRR for the observed over the expected incidence of severe HFMD cases was 0.48 (95% PI: 0.40, 0.58; Table 2; Figure 2) according to the best predictive GLM, and 0.46 (95% PI: 0.22, 1.80) according to the random forest model. A cumulative total of 286 (95% PI: 188, 396) fewer severe cases were reported than expected.

The change point model that achieved the highest log-likelihood identified a change in EV71 trend in April 2017, roughly 10 months after vaccine introduction (Figure 3). The resultant model estimated an IRR of 0.18 (95% PI: 0.08, 0.42), corresponding to 14,653 (95% PI: 4,267, 33,874) EV71 cases averted, when comparing observed incidence rates for May 2017-December 2018 with the counterfactual predictions.

Vaccine impacts on CA16 and other HFMD etiologies

Comparisons of observed CA16 and other HFMD incidence rates with forecasts from their respective best predictive GLMs did not provide strong statistical evidence for changes in CA16 transmission when examined over the two-year post-vaccine period (Table 2). For CA16, the estimated post-vaccine IRR was 0.81 (95% PI: 0.25, 4.06); however, for non-EV71 and non-CA16 HFMD the IRR was 2.01 (95% PI: 0.96, 4.82). We observed 25,574 (95% PI: 3,938, 36,727) more cases of HFMD caused by non-CA16 and non-EV71 enteroviruses in 2018 than predicted by our GLM (IRR: 2.59, 95% PI: 1.10, 8.45). Point estimates from random forest models were consistent in direction of effect with GLM results, but did not indicate a statistically significant decline or increase of cases for either the 2017-2018 period or for 2018 in isolation (Table S2).

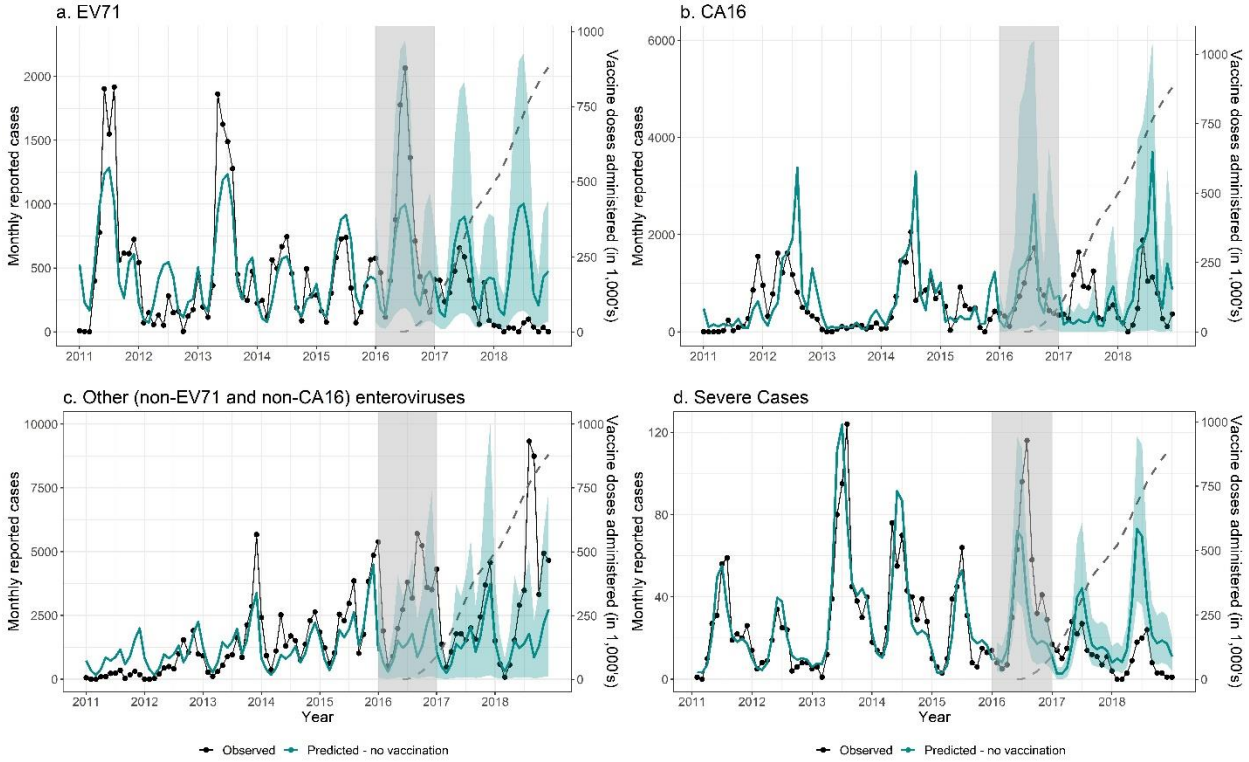


Figure 2. Results of analyses examining trends from 2011-2018 for a) EV71-specific HFMD, b) CA16-specific HFMD, c) HFMD caused by another enterovirus, and d) severe HFMD, and associated comparisons of observed events (black dotted lines and dots) to predicted events (teal solid line) before and after vaccination. Expected number of cases and 80% prediction intervals (teal shading) are based on predictions from GLMs fitted to historic data. The gray box indicates the year vaccination was first began and is removed from calculation of averted cases. The gray dotted line indicates the cumulative number of doses of the vaccine (in 1,000's) given in Chengdu (right axis).

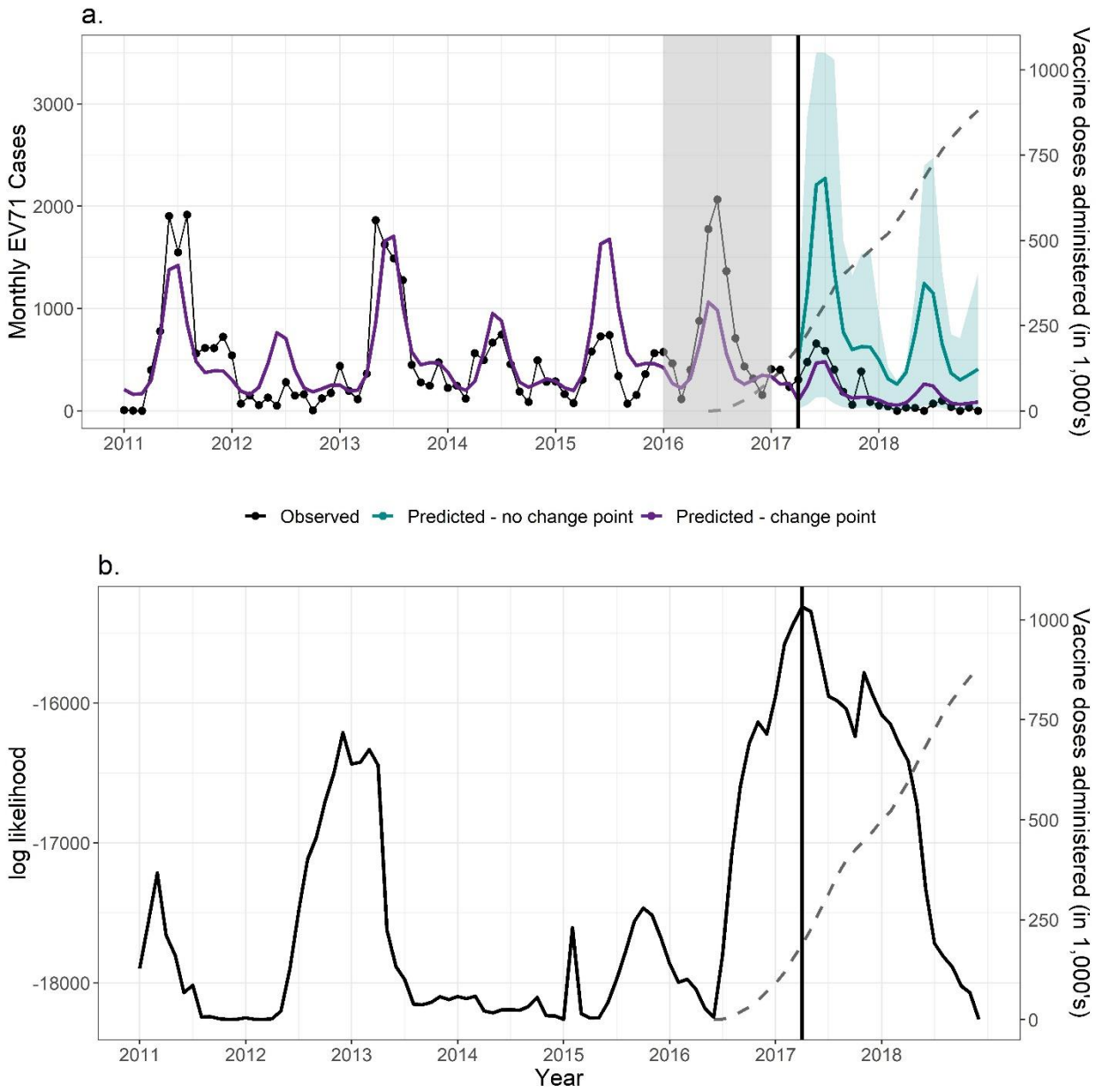


Figure 3. Results of the change point analysis testing a change in trend for individual months in the full time series, 2011-2018. The gray dotted line in both panels indicates the total number of doses of the vaccine (in 1,000's) administered in Chengdu (right axis). a) Observed cases (black) against predicted cases (purple), setting the change point at the most likely month (estimated as April, 2017). Counterfactual predictions (teal) indicate continuation of the trend if the effect of the change in slope were removed. b) The log-likelihood of the model when a change in the slope was included at the month indicated (maximum likelihood indicated by vertical line, April, 2017).

Table 6. Observed EV71, CA16, and other etiology HFMD case counts, expected cases under GLMs, and incidence rate ratios for the post-vaccine period 2017-2018 in Chengdu, China

	Cases in 2017		
	Observed	**Averted (95% PI)****	**IRR (95% PI)**
EV71*	4,198	1,247 (-1,851, 60,080)	0.77 (0.07, 1.79)
CA16*	8,465	-4,262 (-7,816, 5,433)	2.01 (0.61, 13.04)
Other *	26,647	-8,725 (-21,529, 19,624)	1.49 (0.58, 5.21)
Severe cases	175	36 (-18, 100)	0.83 (0.64, 1.11)
Total	39,310	-11,740 (-27,519, 20,247)	1.43 (0.66, 3.33)
	Cases in 2018		
	Observed	**Averted (95% PI)**	**IRR (95% PI)**
EV71*	400	5,664 (2,186, 11,059)	0.07 (0.03, 0.15)
CA16*	6,702	7,706 (-4,766, 47,579)	0.47 (0.12, 3.46)
Other *	41,658	-25,574 (-36,727, -3,938)	2.59 (1.10, 8.45)
Severe cases	90	250 (170, 342)	0.26 (0.21, 0.35)
Total	48,757	-11,804 (-33,404, 33,675)	1.32 (0.59, 3.18)
	Combined Cases, 2017 - 2018		
	Observed	**Averted (95% PI)**	**IRR (95% PI)**
EV71*	4,598	6,911 (3,246, 11,542)	0.40 (0.28, 0.59)
CA16*	15,167	3,444 (-11,439, 45,394)	0.81 (0.25, 4.06)
Other *	68,304	-34,299 (-54,132, 3,188)	2.01 (0.96, 4.82)
Severe cases	265	286 (188, 396)	0.48 (0.40, 0.58)
Total	88,069	-23,944 (-55,494, 35, 214)	1.37 (0.71, 2.70)

IRR = Incidence Rate Ratio; PI = Prediction Interval

*Adjusted by inverse probability of testing. See Supplemental Text 1.

**Prediction intervals calculated via 1,000,000 Monte Carlo simulation draws using fitted model parameters

4.5 Discussion

We present the first estimates of the real-world impact of vaccination of children 6-59 months against EV71 in Chengdu Prefecture, a major center of EV71 transmission. We observed a substantial decrease in reported EV71-specific and severe HFMD cases in Chengdu in 2017-2018, the years following vaccine introduction. Infection with EV71 was confirmed in 85% of HFMD-associated deaths, which corresponds with a previous report of high (92%) proportion of deaths attributable to EV71 [41]. We found no statistical evidence of post-vaccine reductions in non-EV71 HFMD etiologies, suggesting that the observed changes in EV71 transmission are likely attributable to the vaccine rather than external factors.

We employed two distinct methods to estimate counterfactual EV71 HFMD incidence in the absence of vaccination. The first, predictive model selection via cross-validation with forecasting to the post-vaccine period, makes no assumption on the functional form needed to capture changes in incidence after vaccine implementation, but uses *a priori* delineation of the intervention period. The second, change point modeling, utilizes the full time series to identify the most likely timing of major changes in incidence, but assumes a more restrictive functional form to fit the post-change point data (i.e., log-linear trend over time). Change point analysis can be more susceptible to influence of phenomena that are poorly captured by the pre-intervention model (e.g., the disruption of 24-month periodicity of EV71 HFMD between 2013-2016). In this study, both methods identified significant decreases in EV71 incidence in the post-vaccine period. Estimates of case reduction due to vaccination from the change point model in 2017 were substantially higher than forecasting estimates, potentially resulting from the influence of high case counts for EV71 in 2016 on pre-change point parameter estimates.

Individually-randomized clinical trials of the vaccines licensed by Sinovac and CAMS demonstrated a vaccine efficacy against EV71 HFMD of over 98% after two doses [208, 209, 211]. Assuming an efficacy of 50% for a single dose, and 90-98% for two doses, Wang et al. predicted that full immunization would avert 7.57 cases of EV71 HFMD per vaccinated infant per year across all of China [231]. While difficult to draw a direct comparison given uncertainties regarding the number of individuals immunized, the proportion receiving one or two doses, and their age and geographic distribution, Wang's predictions are consistent with our estimates: 3,246-11,542 cases averted over 2 years, after administering 880,673 doses. If all children received two doses, this corresponds to 3.69-13.11 cases averted per thousand vaccinated children per year by our estimate.

Given that the vaccine, unlike EV71 infection, is not expected to confer short-term protection against non-EV71 serotypes [42, 43], increases in non-vaccine serotypes post-vaccination may indicate serotype replacement [35, 36]. While our GLM forecasts show that incidence of HFMD etiologies other than EV71 and CA16 increased in 2018, it is difficult to discern if this is attributable to serotype replacement or other factors, as 2018 was a year of anomalously high HFMD incidence [212, 213]. Mathematical simulations predict that serotype replacement, if it occurred, would be minimal [42, 232]. Our model variances are large for all serotypes, so continuous monitoring of non-vaccine serotypes is warranted.

This study has limitations. Although a small proportion of cases underwent serotype determination, inverse probability weighting permitted adjustment of the case counts for each serotype to account for testing propensities. While the post-vaccine period is of limited duration, the length of our pre-vaccine period is consistent with similar studies and captured inter- and intra-annual variation in incidence. Leave-out-one-year cross validation in model selection expresses well the uncertainty in extrapolating estimates to unforeseen years [233].

The best predictive model for EV71 captured two-year periodicity in the training period using an autoregressive term rather than 24-month harmonic terms. Consequently, mean predictions for both 2017 and 2018 indicated moderate incidence, rather than continued alternation between low and high incidence years; variance in predicted prior years' incidence was absorbed into prediction uncertainty. While this may result in mean predictions inconsistent with the pre-vaccination pattern, the additional prediction variance should lead to conservative lower bound estimates of cases averted, and is not expected to create bias when examining the full post-vaccine period. The 2016 rise in EV71 incidence could have caused, in part, the observed decrease in 2017 if the susceptible population were depleted. However, if vaccination had not been effective, we would have expected a rebound in 2018 consistent with the observed biennial patterns. Also, inclusion of prior years' EV71 incidence as a covariate in the model accounted for population immunity in our counterfactual predictions.

As only 19 HFMD deaths were reported between 2011-2018, we did not evaluate the impact of vaccination on mortality. Between one and six deaths were reported annually from 2011-2016, while only one death was reported in 2017 and none in 2018 (Table S1). Limited reporting among children under five years at the beginning of the time series prevented stratification of the vaccine impact by age group, as age-specific models for pre-vaccine EV71 incidence were unstable. The adjusted incidence of EV71 declines in 2017-2018 across all ages, potentially suggesting indirect and direct effects of vaccination (Figure S3). As the median age of EV71 infection is seven years, there may also be delayed direct effect in children aged five to seven.

4.6 Conclusion

We provide early evidence of vaccine impact on HFMD incidence in an urban area with high EV71 transmission. After initiation of pediatric immunization with inactivated EV71 vaccine, a dramatic decline in EV71-associated HFMD incidence was observed in Chengdu, accompanied by a similar relative decline in severe HFMD. It is unknown whether the 2018 increase in cases caused by other serotypes resulted from vaccine-driven serotype replacement, continuing emergence of non-EV71 HFMD, or external factors. Continued monitoring of non-EV71 serotypes will thus be crucial. For now, the strong evidence from clinical trials, simulations, and the substantial population-level decreases in EV71 HFMD incidence documented here support use of EV71 vaccine to control severe childhood HFMD.

CONCLUSION

Despite decreasing trends in infectious disease transmission over the past decades in the US and globally [234], the SARS-CoV-2 pandemic lays bare the threat that emerging infectious diseases can have on population health. Amidst ongoing environmental, social, and demographic changes, we may expect to continue to see rises in the number of EIDs [2, 8]. This dissertation demonstrates the important role of environmental variability, individual behavior, and public health policy in the transmission dynamics of EIDs.

In Chapters 1 and 2, I demonstrate the importance of environmental conditions in the transmission of coccidioidomycosis and the distribution of *Coccidioides* in the environment. Chapter 1 finds that a multiannual sequence of drought-like conditions, followed by a wet winter and a hot, dry fall amplifies incidence of coccidioidomycosis. What is more, increased temperatures across the southwestern US, expected as part of global climate change, is expected to increase the frequency and severity of drought. Chapter 1 demonstrates that drought overall increases incidence, by amplifying cases in the years following the drought. Chapter 2 examines the role of environmental conditions on the environmental distribution for *Coccidioides immitis*. I demonstrate that burrows are associated with increased presence of *C. immitis*, suggesting that burrows provide suitable local conditions, such as reduced soil moisture. At the same time, I demonstrate for the first time, a strong effect of rodents on *C. immitis* presence, beyond the influence of their burrows. Many EIDs have important interactions with wildlife, and shifting climate patterns can greatly affect how zoonotic hosts interact with environmental reservoirs.

In Chapters 3 and 4, I demonstrate the role of individual contact patterns and public health policy in mitigating circulation of SARS-CoV-2 and enteroviruses that cause hand, foot, and mouth disease (HFMD). Chapter 3 finds that reductions in child social contact patterns as a result of school closures may have averted substantial number of cases of COVID-19 during the first emergence of SARS-CoV-2 to the Bay Area in the spring of 2020. Moreover, Chapter 3 demonstrates how individual and collective mask use, decisions to get vaccinated, and school policies for cohorts, mask and vaccine mandates, and testing could help reduce the spread of SARS-CoV-2 within the K-12 environment. Finally, Chapter 4 finds that pediatric immunization against EV71 averted 60% of EV71-HFMD cases, without causing significant replacement of other, non-EV71 enteroviruses that cause HFMD.

EIDs can be associated with uncertain transmission dynamics, unknown risk factors for severe disease, and lack of evidence for effective health interventions, all of which present challenges for public health policies for prevention and control [8]. The earlier that researchers can detect emergence of a pathogen and describe the shifting transmission dynamics, the more prepared we will be to generate solutions. In the decades to come, the field of Public Health will have to contend with the challenge of rapid detection and characterization of EIDs. This dissertation demonstrates application of many tools that can be used by public health researchers seeking

to detect and understand the mechanisms for disease emergence – including disease and environmental surveillance, causal inference, and models for disease prediction.

First, disease surveillance has long been recognized as a critical tool for early detection of pathogen emergence or re-emergence [235], especially when coupled with genomics or laboratory data [236]. Chapters 1 and 4 of this dissertation demonstrate the value of surveillance data for detecting pathogen emergence and ascertaining underlying factors behind the emergence.

Second, the majority of EIDs have zoonotic and/or environmental reservoirs [2]. Pathogens may emerge first in such reservoirs, before being detectable in the human population. Environmental surveillance can be a useful tool for detecting diseases in wildlife hosts, domestic animal hosts, or environmental media such as water or soils, in advance of spillover events from animals to humans or environmentally–mediated transmission events [237]. Environmental sampling has also been used to understand the environmental reservoir of emerging pathogens [238]. Chapter 2 describes the initial phases of a research agenda for environmental surveillance of *Coccidioides* within a highly endemic area. While Chapter 2 describes data only from April of 2021, ongoing soil sampling is continuing, which will permit us to understand fluctuations and trends in pathogen concentration in the soils over time. The environmental surveillance in Chapter 2 makes important headway towards understanding whether the most important reservoir for *Coccidioides* is soils, rodents, or both. While *Coccidioides* is not transmitted directly from host-to-host, I have demonstrated that rodents are important contributors to *Coccidioides* presence within the soil.

Third, while disease and environmental surveillance are critical for gathering information on disease trends across space and time, formal hypothesis testing is needed to understand whether fluctuations in disease patterns are indicative of meaningful shifts, and to investigate the underlying factors behind the shifts. Causal inference and counterfactual theory can be important tools for answering these questions. Chapter 1 demonstrates application of ensemble prediction algorithm with g-computation to generate a counterfactual estimate of the number of cases of coccidioidomycosis that would have occurred absent two major droughts in California. Inclusion of parametric and non-parametric algorithms within a weighted averaged model permitted flexible model fitting while improving out-of-sample prediction error by ensuring against over-fitting. Chapter 3 demonstrates application of causal mediation analysis with g-computation to obtain a mechanistic understanding of the role of rodents, burrows, and soil conditions in *Coccidioides* environmental distribution. Finally, Chapter 4 demonstrates application of multiple time series approaches – including both parametric and machine learned interrupted time series regression and a change point analysis – to general counterfactual estimates of disease that would have occurred absent vaccine introduction.

Finally, robust models for disease prediction and forecasting can help public health researchers make informed policy decisions, in absence of prior evidence. During the SARS-CoV-2 pandemic, mathematical modeling was utilized frequently as a tool for anticipating the effect of various

public health measures. Chapter 3 demonstrates application of agent-based models, coupled with primary collected data on soil contact patterns, for making decisions regarding public health prevention measures for K-12 populations.

All in all, public health must be prepared to encounter emerging and re-emerging pathogens in the decades to come. This dissertation explores mechanisms behind the emergence and control of EIDs and demonstrates methodological tools that can help researchers understand mitigate transmission of EIDs. The transmission dynamics of many EIDs – including coccidioidomycosis and SARS-CoV-2 – remains poorly understood. Myriad opportunity exists to continue to understand mechanisms underlying transmission and shifting epidemiology o these and other EIDs, along with the most effective prevention measures.

References

1. National Institute of Allergy and Infectious Diseases. NIAID Emerging Infectious Diseases/Pathogens Washington, DC2018 [Available from: https://www.niaid.nih.gov/research/emerging-infectious-diseases-pathogens.
2. Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, et al. Global trends in emerging infectious diseases. Nature. 2008;451:990.
3. Morens DM, Folkers GK, Fauci AS. The challenge of emerging and re-emerging infectious diseases. Nature. 2004;430(6996):242-9.
4. Chen LH, Wilson ME. The Role of the Traveler in Emerging Infections and Magnitude of Travel. Medical Clinics of North America. 2008;92(6):1409-32.
5. Nucci M, Marr KA. Emerging fungal diseases. Clin Infect Dis. 2005;41(4):521-6.
6. Baker RE, Mahmud AS, Miller IF, Rajeev M, Rasambainarivo F, Rice BL, et al. Infectious disease in an era of global change. Nature Reviews Microbiology. 2021.
7. Jones BA, Grace D, Kock R, Alonso S, Rushton J, Said MY, et al. Zoonosis emergence linked to agricultural intensification and environmental change. Proceedings of the National Academy of Sciences. 2013;110(21):8399-404.
8. Oaks Jr SC, Shope RE, Lederberg J. Emerging infections: microbial threats to health in the United States. 1992.
9. Binder S, Levitt AM, Sacks JJ, Hughes JM. Emerging Infectious Diseases: Public Health Issues for the 21st Century. Science. 1999;284(5418):1311-3.
10. Martin G, Yanez-Arenas C, Chen C, Plowright RK, Webb RJ, Skerratt LF. Climate change could increase the geographic extent of Hendra virus spillover risk. EcoHealth. 2018;15(3):509-25.
11. Skaff NK, Cheng Q, Clemesha RE, Collender PA, Gershunov A, Head JR, et al. Thermal thresholds heighten sensitivity of West Nile virus transmission to changing temperatures in coastal California. Proceedings of the Royal Society B. 2020;287(1932):20201065.
12. Galgiani JN, Ampel NM, Blair JE, Catanzaro A, Johnson RH, Stevens DA, et al. Coccidioidomycosis. Clin Infect Dis. 2005;41(9):1217-23.
13. Tabnak F LL, Sondermeyer G, Vugia D. . Epidemiologic Summary of Coccidioidomycosis in California, 2009-2012. Richmond, CA: California Department of Public Health; 2014.
14. Brown J, Benedict K, Park BJ, Thompson GR, 3rd. Coccidioidomycosis: epidemiology. Clinical epidemiology. 2013;5:185-97.
15. Cooksey GLS, Nguyen A, Vugia D, Jain S. Regional Analysis of Coccidioidomycosis Incidence—California, 2000–2018. Morbidity and Mortality Weekly Report. 2020;69(48):1817.
16. Gorris ME, Treseder KK, Zender CS, Randerson JT. Expansion of Coccidioidomycosis Endemic Regions in the United States in Response to Climate Change. GeoHealth. 2019;3(10):308-27.
17. Robins J. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. Mathematical Modelling. 1986;7(9):1393-512.
18. Lacy GH, Swatek FE. Soil ecology of Coccidioides immitis at Amerindian middens in California. Applied microbiology. 1974;27(2):379-88.

19. Emmons CW. Coccidioidomycosis in Wild Rodents. A Method of Determining the Extent of Endemic Areas. Public Health Reports (1896-1970). 1943;58(1):1-5.
20. Emmons CW, Ashburn LL. The Isolation of Haplosporangium parvum n. sp. and Coccidioides immitis from Wild Rodents. Their Relationship to Coccidioidomycosis. Public Health Reports (1896-1970). 1942;57(46):1715-27.
21. Smith R. Natural history of the prairie dog in Kansas. Museum Univ Kansas, Misc Pubs. 1967;69:1-36.
22. Barker BM, Tabor JA, Shubitz LF, Perrill R, Orbach MJ. Detection and phylogenetic analysis of Coccidioides posadasii in Arizona soil samples. Fungal Ecology. 2012;5(2):163-76.
23. Ernest SKM, Brown JH, Parmenter RR. Rodents, Plants, and Precipitation: Spatial and Temporal Dynamics of Consumers and Resources. Oikos. 2000;88(3):470-82.
24. Del Rocío Reyes-Montes M, Pérez-Huitrón MA, Ocaña-Monroy JL, Frías-De-León MG, Martínez-Herrera E, Arenas R, et al. The habitat of Coccidioides spp. and the role of animals as reservoirs and disseminators in nature. BMC infectious diseases. 2016;16(1):550-.
25. Sharpton TJ, Stajich JE, Rounsley SD, Gardner MJ, Wortman JR, Jordar VS, et al. Comparative genomic analyses of the human fungal pathogens Coccidioides and their relatives. Genome research. 2009;19(10):1722-31.
26. Taylor JW, Barker BM. The endozoan, small-mammal reservoir hypothesis and the life cycle of Coccidioides species. Medical Mycology. 2019;57(Supplement_1):S16-S20.
27. Anderson RM, Heesterbeek H, Klinkenberg D, Hollingsworth TD. How will country-based mitigation measures influence the course of the COVID-19 epidemic? Lancet. 2020;395(10228):931-4.
28. San Francisco Chronicle Staff. Timeline: How the Bay Area has combated the coronavirus. 2020 2020 [Available from: https://projects.sfchronicle.com/2020/coronavirus-timeline/
29. Levinson M, Cevik M, Lipsitch M. Reopening Primary Schools during the Pandemic. New England Journal of Medicine. 2020.
30. Yoshikawa H, Wuermli AJ, Britto PR, Dreyer B, Leckman JF, Lye SJ, et al. Effects of the Global Coronavirus Disease-2019 Pandemic on Early Childhood Development: Short- and Long-Term Risks and Mitigating Program and Policy Actions. The Journal of pediatrics. 2020;223:188-93.
31. Armitage R, Nellums LB. Considering inequalities in the school closure response to COVID-19. The Lancet Global Health. 2020;8(5):e644.
32. Van Lancker W, Parolin Z. COVID-19, school closures, and child poverty: a social crisis in the making. The Lancet Public Health. 2020;5(5):e243-e4.
33. Dunn CG, Kenney E, Fleischhacker SE, Bleich SN. Feeding Low-Income Children during the Covid-19 Pandemic. N Engl J Med. 2020;382(18):e40.
34. Cicero A, Potter C, Kirk ST, Rivers C, Schoch-Spana M. Filling in the Blanks: National Research Needs to Guide Decisions about Reopening Schools in the United States. The Johns Hopkins Center for Health Security 2020 [Available from: https://www.centerforhealthsecurity.org/our-work/publications/filling-in-the-blanks-national-research-needs-to-guide-decisions-about-reopening-schools-in-the-united-states.
35. Weinberger DM, Malley R, Lipsitch M. Serotype replacement in disease after pneumococcal vaccination. Lancet. 2011;378(9807):1962-73.

36. Ribeiro GS, Reis JN, Cordeiro SM, Lima JB, Gouveia EL, Petersen M, et al. Prevention of Haemophilus influenzae type b (Hib) meningitis and emergence of serotype replacement with type a strains after introduction of Hib immunization in Brazil. Journal of Infectious Diseases. 2003;187(1):109-16.
37. National Health Commission of the People's Republic of China 2019 [Available from: http://www.nhc.gov.cn/.
38. Xing W, Liao Q, Viboud C, Zhang J, Sun J, Wu JT, et al. Hand, foot, and mouth disease in China, 2008-12: an epidemiological study. The Lancet Infectious diseases. 2014;14(4):308-18.
39. Bian L, Wang Y, Yao X, Mao Q, Xu M, Liang Z. Coxsackievirus A6: a new emerging pathogen causing hand, foot and mouth disease outbreaks worldwide. Expert Review of Anti-infective Therapy. 2015;13(9):1061-71.
40. Luo K, Gao L, Zou G, Liu J, Mo X, Wang L, et al. Spectrum of Enterovirus Serotypes Causing Uncomplicated Hand, Foot, and Mouth Disease and Enteroviral Diagnostic Yield of Different Clinical Samples. Clinical Infectious Diseases. 2018;67(11):1729-35.
41. Ji T, Han T, Tan X, Zhu S, Yan D, Yang Q, et al. Surveillance, epidemiology, and pathogen spectrum of hand, foot, and mouth disease in mainland of China from 2008 to 2017. Biosafety and Health. 2019.
42. Takahashi S, Liao Q, Van Boeckel TP, Xing W, Sun J, Hsiao VY, et al. Hand, Foot, and Mouth Disease in China: Modeling Epidemic Dynamics of Enterovirus Serotypes and Implications for Vaccination. PLOS Medicine. 2016;13(2):e1001958.
43. Yang B, Liu F, Liao Q, Wu P, Chang Z, Huang J, et al. Epidemiology of hand, foot and mouth disease in China, 2008 to 2015 prior to the introduction of EV-A71 vaccine. Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin. 2017;22(50):16-00824.
44. Stevens DA. Coccidioidomycosis. New England Journal of Medicine. 1995;332(16):1077-82.
45. Stewart ER, Thompson GR. Update on the Epidemiology of Coccidioidomycosis. Current Fungal Infection Reports. 2016;10(4):141-6.
46. National Center for Atmospheric Research Staff (Eds). The Climate Data Guide: PRISM High-Resolution Spatial Climate Data for the United States: Max/min temp, dewpoint, precipitation 2017 [Available from: https://climatedataguide.ucar.edu/climate-data/prism-high-resolution-spatial-climate-data-united-states-maxmin-temp-dewpoint.
47. Diffenbaugh NS, Swain DL, Touma D. Anthropogenic warming has increased drought risk in California. Proceedings of the National Academy of Sciences. 2015;112(13):3931-6.
48. United States Geologic Service. California Drought: What is drought? : USGS; 2021 [Available from: https://ca.water.usgs.gov/california-drought/what-is-drought.html.
49. Swain DL, Tsiang M, Haugen M, Singh D, Charland A, Rajaratnam B, et al. The extraordinary California drought of 2013/2014: Character, context, and the role of climate change. Bull Am Meteorol Soc. 2014;95(9):S3-S7.
50. Griffin D, Anchukaitis KJ. How unusual is the 2012–2014 California drought? Geophysical Research Letters. 2014;41(24):9017-23.
51. Mao Y, Nijssen B, Lettenmaier DP. Is climate change implicated in the 2013–2014 California drought? A hydrologic perspective. Geophysical Research Letters. 2015;42(8):2805-13.

52. United States Department of Agriculture. Unites States Drought Monitor 2020 [Available from: https://droughtmonitor.unl.edu/.
53. Centers for Disease Control and Prevention. Valley Fever Maps Atlanta, GA: CDC; 2019 [Available from: https://www.cdc.gov/fungal/diseases/coccidioidomycosis/maps.html.
54. Fisher FS, Bultman MW, Johnson SM, Pappagianis D, Zaborsky E. Coccidioides niches and habitat parameters in the southwestern United States: a matter of scale. Annals of the New York Academy of Sciences. 2007;1111:47-72.
55. Maddy KT. Ecological factors of the geographic distribution of Coccidioides immitis. Journal of the American Veterinary Medical Association. 1957;130(11):475-6.
56. Komatsu K, Vaz V, McRill C, Colman T. Increase in coccidioidomycosis-Arizona, 1998-2001. Jama. 2003;289(12):1500-.
57. Comrie AC. Climate factors influencing coccidioidomycosis seasonality and outbreaks. Environ Health Perspect. 2005;113(6):688-92.
58. Kolivras KN, Comrie AC. Modeling valley fever (coccidioidomycosis) incidence on the basis of climate conditions. Int J Biometeorol. 2003;47(2):87-101.
59. Weaver EA, Kolivras KN. Investigating the Relationship Between Climate and Valley Fever (Coccidioidomycosis). EcoHealth. 2018;15(4):840-52.
60. Gorris M, Cat L, Zender C, Treseder K, Randerson J. Coccidioidomycosis dynamics in relation to climate in the southwestern United States. GeoHealth. 2018;2(1):6-24.
61. Coopersmith EJ, Bell JE, Benedict K, Shriber J, McCotter O, Cosh MH. Relating coccidioidomycosis (valley fever) incidence to soil moisture conditions. GeoHealth. 2017;1:51-63.
62. Tamerius JD, Comrie AC. Coccidioidomycosis incidence in Arizona predicted by seasonal precipitation. PloS one. 2011;6(6):e21009.
63. Zender CS, Talamantes J. Climate controls on valley fever incidence in Kern County, California. Int J Biometeorol. 2006;50(3):174-82.
64. Talamantes J, Behseta S, Zender CS. Fluctuations in climate and incidence of coccidioidomycosis in Kern County, California: a review. Annals of the New York Academy of Sciences. 2007;1111:73-82.
65. Comrie AC, Glueck MF. Assessment of climate-coccidioidomycosis model: model sensitivity for assessing climatologic effects on the risk of acquiring coccidioidomycosis. Annals of the New York Academy of Sciences. 2007;1111:83-95.
66. Azar A, Smith K. Epidemiologic Summary of Coccidioidomycosis in California, 2017. Sacramento, California: California Department of Public Health; 2017.
67. California Department of Public Health. Coccidioidomycosis in California Provisional Monthly Report January - December 2018. Sacramento, CA: CDPH; 2018.
68. Centers for Disease Control and Prevention. Case definitions for infectious conditions under public health surveillance, coccidioidomycosis. 2008.
69. California Department of Public Health. CDPH IDB Guidance for Managing Select Communicable Diseases: Coccidioidomycosis Sacramento: CDPH; 2018 [Available from: https://www.cdph.ca.gov/Programs/CID/DCDC/CDPH%20Document%20Library/IDBGuidanceforCALHJs-Cocci.pdf.
70. ESRI. ArcGIS Server REST API: Release 10. In: Institute ESR, editor. Redlands, CA: Environmental Systems Research Institute; 2020.

71. Chaney NW, Wood EF, McBratney AB, Hempel JW, Nauman TW, Brungard CW, et al. POLARIS: A 30-meter probabilistic soil series map of the contiguous United States. Geoderma. 2016;274:54-67.
72. Multi Resolution Land Characteristics Consortium. National Land Cover Database. 2011.
73. United States Geologic Service. USGS National Elevation Database 2018 [Available from: https://catalog.data.gov/dataset/usgs-national-elevation-dataset-ned.
74. Guo C, Yang J, Guo Y, Ou Q-Q, Shen S-Q, Ou C-Q, et al. Short-term effects of meteorological factors on pediatric hand, foot, and mouth disease in Guangdong, China: a multi-city time-series analysis. BMC infectious diseases. 2016;16(1):524.
75. Gasparrini A, Armstrong B, Kenward MG. Multivariate meta-analysis for non-linear and other multi-parameter associations. Statistics in medicine. 2012;31(29):3821-39.
76. Liu C, Yin P, Chen R, Meng X, Wang L, Niu Y, et al. Ambient carbon monoxide and cardiovascular mortality: a nationwide time-series analysis in 272 cities in China. The Lancet Planetary Health. 2018;2(1):e12-e8.
77. Peng RD, Dominici F, Louis TA. Model choice in time series studies of air pollution and mortality. Journal of the Royal Statistical Society: Series A (Statistics in Society). 2006;169(2):179-203.
78. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2015.
79. Polley E, Van Der Laan M. Super Learner in prediction. UC Berkeley Division of Biostatistics Working Paper Series. Working Paper 266, May 2010, http://biostats. bepress. com/ucbbiostat/paper266; 2010.
80. Wright MN, Ziegler A. ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. Journal of Statistical Software. 2017;77.i01.
81. Baumgardner DJ, Paretsky DP, Baeseman ZJ, Schreiber A. Effects of season and weather on blastomycosis in dogs: Northern Wisconsin, USA. Medical Mycology. 2011;49(1):49-55.
82. Greene DR, Koenig G, Fisher MC, Taylor JW. Soil isolation and molecular identification of Coccidioides immitis. Mycologia. 2000;92(3):406-10.
83. Friedman L, Smith CE, Pappagianis D, Berman R. Survival of Coccidioides immitis under controlled conditions of temperature and humidity. American Journal of Public Health and the Nations Health. 1956;46(10):1317-24.
84. Crits-Christoph A, Robinson CK, Barnum T, Fricke WF, Davila AF, Jedynak B, et al. Colonization patterns of soil microbial communities in the Atacama Desert. Microbiome. 2013;1(1):1-13.
85. Lynch R, King A, Farías ME, Sowell P, Vitry C, Schmidt S. The potential for microbial life in the highest-elevation (> 6000 masl) mineral soils of the Atacama region. Journal of Geophysical Research: Biogeosciences. 2012;117(G2).
86. Kerr JR. Bacterial inhibition of fungal growth and pathogenicity. Microbial Ecology in Health and Disease. 1999;11(3):129-42.
87. Kathiravan MK, Salake AB, Chothe AS, Dudhe PB, Watode RP, Mukta MS, et al. The biology and chemistry of antifungal agents: A review. Bioorganic & Medicinal Chemistry. 2012;20(19):5678-98.

88. Lauer A, Baal JD, Mendes SD, Casimiro KN, Passaglia AK, Valenzuela AH, et al. Valley Fever on the Rise-Searching for Microbial Antagonists to the Fungal Pathogen Coccidioides immitis. Microorganisms. 2019;7(2):31.
89. Köberl M, Müller H, Ramadan EM, Berg G. Desert Farming Benefits from Microbial Potential in Arid Soils and Promotes Diversity and Plant Health. PloS one. 2011;6(9):e24452.
90. Kollath DR, Teixeira MM, Funke A, Miller KJ, Barker BM. Investigating the Role of Animal Burrows on the Ecology and Distribution of Coccidioides spp. in Arizona Soils. Mycopathologia. 2020;185(1):145-59.
91. Prugh LR, Deguines N, Grinath JB, Suding KN, Bean WT, Stafford R, et al. Ecological winners and losers of extreme drought in California. Nature Climate Change. 2018;8(9):819-24.
92. Stephens RB, Rowe RJ. The underappreciated role of rodent generalists in fungal spore dispersal networks. Ecology. 2020;101(4):e02972.
93. Buckee C, Noor A, Sattenspiel L. Thinking clearly about social aspects of infectious disease transmission. Nature. 2021;595(7866):205-13.
94. Schrieks T, Botzen WJW, Wens M, Haer T, Aerts JCJH. Integrating Behavioral Theories in Agent-Based Models for Agricultural Drought Risk Assessments. Frontiers in Water. 2021;3(104).
95. Neelin JD, Langenbrunner B, Meyerson JE, Hall A, Berg N. California winter precipitation change under global warming in the Coupled Model Intercomparison Project phase 5 ensemble. Journal of Climate. 2013;26(17):6238-56.
96. Diffenbaugh NS, Giorgi F. Climate change hotspots in the CMIP5 global climate model ensemble. Climatic Change. 2012;114(3):813-22.
97. Bedsworth L, Cayan D, Franco G, Fisher L, Ziaja S. California's Fourth Climate Change Assessment. 2019.
98. Liang X, Lettenmaier DP, Wood EF, Burges SJ. A simple hydrologically based model of land surface water and energy fluxes for general circulation models. Journal of Geophysical Research: Atmospheres. 1994;99(D7):14415-28.
99. Tsang CA, Anderson SM, Imholte SB, Erhart LM, Chen S, Park BJ, et al. Enhanced surveillance of coccidioidomycosis, Arizona, USA, 2007-2008. Emerg Infect Dis. 2010;16(11):1738-44.
100. California Department of Public Health. Valley Fever Prevention Tips Sacramento, CA: CDPH; 2021 [Available from: https://www.cdph.ca.gov/Programs/CID/DCDC/Pages/ValleyFeverPrevention.aspx.
101. Hurd-Kundeti G, Cooksey GLS, Jain S, Vugia DJ. Valley Fever (Coccidioidomycosis) Awareness—California, 2016–2017. Morbidity and Mortality Weekly Report. 2020;69(42):1512.
102. Eulalio KD, de Macedo RL, Cavalcanti MA, Martins LM, Lazera MS, Wanke B. Coccidioides immitis isolated from armadillos (Dasypus novemcinctus) in the state of Piaui, northeast Brazil. Mycopathologia. 2001;149(2):57-61.
103. Cordeiro Rde A, e Silva KR, Brilhante RS, Moura FB, Duarte NF, Marques FJ, et al. Coccidioides posadasii infection in bats, Brazil. Emerg Infect Dis. 2012;18(4):668-70.
104. Egeberg RO, Elconin AE, Egeberg MC. Effect of salinity and temperature on Coccidioides immitis and three antagonistic soil saprophytes. Journal of bacteriology. 1964;88(2):473-6.
105. Lauer A, Talamantes J, Castanon Olivares LR, Medina LJ, Baal JD, Casimiro K, et al. Combining forces--the use of Landsat TM satellite imagery, soil parameter information, and

multiplex PCR to detect Coccidioides immitis growth sites in Kern County, California. PloS one. 2014;9(11):e111921.
106. Baptista-Rosas RC, Hinojosa A, Riquelme M. Ecological niche modeling of Coccidioides spp. in western North American deserts. Annals of the New York Academy of Sciences. 2007;1111(1):35-46.
107. Swatek F, Plunkett O. Ecological studies on Coccidioides immitis: Experimental infection of wild rodents and animals other than mammals. Proceedings of Symposium of Coccidioidomycosis. 1957:161-7.
108. Catalan-Dibene J, Johnson SM, Eaton R, Romero-Olivares AL, Baptista-Rosas RC, Pappagianis D, et al. Detection of coccidioidal antibodies in serum of a small rodent community in Baja California, Mexico. Fungal biology. 2014;118(3):330-9.
109. Egeberg R, Ely AF. Coccidioides immitis in the soil of the southern San Joaquin Valley. American Journal of Medical Sciences. 1956;231(2):151-4.
110. Head JR, Sondermeyer-Cooksey G, Heaney AK, Yu AT, Jones I, Bhattachan A, et al. Influence of meteorological factors and drought on coccidioidomycosis incidence in California, 2000–2020. medRxiv. 2022:2022.02.03.22270412.
111. Studier EH, Baca TP. Atmospheric conditions in artificial rodent burrows. The Southwestern Naturalist. 1968:401-10.
112. Bennett N, Jarvis J, Davies K. Daily and seasonal temperatures in the burrows of African rodent moles. African Zoology. 1988;23(3):189-95.
113. Gurney CM, Prugh LR, Brashares JS. Restoration of native plants is reduced by rodent-caused soil disturbance and seed removal. Rangeland Ecology & Management. 2015;68(4):359-66.
114. Mun H-T, Whitford WG. Factors affecting annual plants assemblages on banner-tailed kangaroo rat mounds. Journal of Arid Environments. 1990;18(2):165-73.
115. Whitford WG, Kay FR. Biopedturbation by mammals in deserts: a review. Journal of Arid Environments. 1999;41(2):203-30.
116. Endicott R, Dillard D, Barnes M, Chang H, Daniels D, Hoang N, et al. Carrizo Plain Ecosystem Project. Berkeley, CA: University of Califonia Berkeley; 2017.
117. Miranda V, Rothen C, Yela N, Aranda-Rickert A, Barros J, Calcagno J, et al. Subterranean Desert Rodents (Genus Ctenomys) Create Soil Patches Enriched in Root Endophytic Fungal Propagules. Microbial Ecology. 2019;77(2):451-9.
118. Corten D. Counting kangaroo rats from space Berkeley: Berkeleyan; 2008 [Available from: https://www.berkeley.edu/news/berkeleyan/2008/10/22_rats.shtml.
119. Wang A, Arah OA. G-computation demonstration in causal mediation analysis. Eur J Epidemiol. 2015;30(10):1119-27.
120. Chow NA, Kangiser D, Gade L, McCotter OZ, Hurst S, Salamone A, et al. Factors Influencing Distribution of Coccidioides immitis in Soil, Washington State, 2016. mSphere. 2021;6(6):e0059821-e.
121. Maddy K, Crecelius H, editors. Establishment of Coccidiodies immitis in negative soil following burial of infected animal tissues. Second Symposium on Coccidioidomycosis1965; 1965.
122. Pappagianis D. Marked increase in cases of coccidioidomycosis in California: 1991, 1992, and 1993. Clin Infect Dis. 1994;19 Suppl 1:S14-8.

123. Kollath D, Barker B, editors. Modeling the discharge of infectious arthroconidia of the fungal pathogen Coccidioides Posadasii growing in soil. 66th Annual Coccidioidomycosis Study Group Meeting; 2022; Bakersfield.
124. Das R, McNary J, Fitzsimmons K, Dobraca D, Cummings K, Mohle-Boetani J, et al. Occupational coccidioidomycosis in California: outbreak investigation, respirator recommendations, and surveillance findings. Journal of occupational and environmental medicine. 2012;54(5):564-71.
125. Sondermeyer Cooksey GL, Wilken JA, McNary J, Gilliss D, Shusterman D, Materna BL, et al. Dust Exposure and Coccidioidomycosis Prevention Among Solar Power Farm Construction Workers in California. American journal of public health. 2017;107(8):1296-303.
126. Litvinova M, Liu Q-H, Kulikov ES, Ajelli M. Reactive school closure weakens the network of social interactions and reduces the spread of influenza. Proceedings of the National Academy of Sciences. 2019;116(27):13174-81.
127. Eames KT, Tilston NL, Brooks-Pollock E, Edmunds WJ. Measured dynamic social contact patterns explain the spread of H1N1v influenza. PLoS computational biology. 2012;8(3).
128. Feehan D, Mahmud A. Quantifying interpersonal contact in the United States during the spread of COVID-19: first results from the Berkeley Interpersonal Contact Study. medRxiv. 2020:2020.04.13.20064014.
129. Jarvis CI, Van Zandvoort K, Gimma A, Prem K, Klepac P, Rubin GJ, et al. Quantifying the impact of physical distance measures on the transmission of COVID-19 in the UK. BMC medicine. 2020;18:1-10.
130. Zhang J, Litvinova M, Liang Y, Wang Y, Wang W, Zhao S, et al. Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. Science. 2020;368(6498):1481-6.
131. Zhang J, Litvinova M, Liang Y, Wang Y, Wang W, Zhao S, et al. Age profile of susceptibility, mixing, and social distancing shape the dynamics of the novel coronavirus disease 2019 outbreak in China. medRxiv. 2020:2020.03.19.20039107.
132. Couzin-Frankel J. School openings across globe suggest ways to keep coronavirus at bay, despite outbreaks. Science. 2020.
133. Cohen JA, Mistry D, Kerr CC, Klein DJ. Schools are not islands: Balancing COVID-19 risk and educational benefits using structural and temporal countermeasures. medRxiv. 2020:2020.09.08.20190942.
134. Espana G, Cavany S, Oidtman RJ, Barbera C, Costello A, Lerch A, et al. Impacts of K-12 school reopening on the COVID-19 epidemic in Indiana, USA. medRxiv. 2020:2020.08.22.20179960.
135. Falk A, Benda A, Falk P, Steffen S, Wallace Z, Høeg TB. COVID-19 Cases and Transmission in 17 K–12 Schools—Wood County, Wisconsin, August 31–November 29, 2020. Morbidity and Mortality Weekly Report. 2021;70(4):136.
136. Panovska-Griffiths J, Kerr CC, Stuart RM, Mistry D, Klein DJ, Viner RM, et al. Determining the optimal strategy for reopening schools, the impact of test and trace interventions, and the risk of occurrence of a second COVID-19 epidemic wave in the UK: a modelling study. The Lancet Child & Adolescent Health. 2020;4(11):817-27.
137. Zimmerman KO, Akinboyo IC, Brookhart MA, Boutzoukas AE, McGann K, Smith MJ, et al. Incidence and secondary transmission of SARS-CoV-2 infections in schools. Pediatrics. 2021.

138. Riley S, Ainslie KEC, Eales O, Walters CE, Wang H, Atchison C, et al. High prevalence of SARS-CoV-2 swab positivity and increasing R number in England during October 2020: REACT-1 round 6 interim report. medRxiv. 2020:2020.10.30.20223123.
139. UK Department of Education. Face coverings in Education United Kingdom2021 [Available from: https://www.gov.uk/government/publications/face-coverings-in-education/face-coverings-in-education#main-changes-to-previous-guidance.
140. Lewis D. What new COVID variants mean for schools is not yet clear. Nature. 2021.
141. Viner RM, Mytton OT, Bonell C, Melendez-Torres GJ, Ward J, Hudson L, et al. Susceptibility to SARS-CoV-2 Infection Among Children and Adolescents Compared With Adults: A Systematic Review and Meta-analysis. JAMA Pediatrics. 2020.
142. Dattner I, Goldberg Y, Katriel G, Yaari R, Gal N, Miron Y, et al. The role of children in the spread of COVID-19: Using household data from Bnei Brak, Israel, to estimate the relative susceptibility and infectivity of children. medRxiv. 2020.
143. Park YJ, Choe YJ, Park O, Park SY, Kim Y-M, Kim J, et al. Contact Tracing during Coronavirus Disease Outbreak, South Korea, 2020. Emerging infectious diseases. 2020;26(10).
144. Laxminarayan R, Wahl B, Dudala SR, Gopal K, Mohan C, Neelima S, et al. Epidemiology and transmission dynamics of COVID-19 in two Indian states. medRxiv. 2020.
145. Fateh-Moghadam P, Battisti L, Molinaro S, Fontanari S, Dallago G, Binkin N, et al. Contact tracing during Phase I of the COVID-19 pandemic in the Province of Trento, Italy: key findings and recommendations. medRxiv. 2020:2020.07.16.20127357.
146. Lopez AS, Hill M, Antezano J, Vilven D, Rutner T, Bogdanow L, et al. Transmission dynamics of COVID-19 outbreaks associated with child care facilities—Salt Lake City, Utah, April–July 2020. Morbidity and Mortality Weekly Report. 2020;69(37):1319.
147. Forbes H, Morton CE, Bacon S, McDonald HI, Minassian C, Brown JP, et al. Association between living with children and outcomes from COVID-19: an OpenSAFELY cohort study of 12 million adults in England. medRxiv. 2020:2020.11.01.20222315.
148. Fontanet A, Tondeur L, Madec Y, Grant R, Besombes C, Jolly N, et al. Cluster of COVID-19 in northern France: A retrospective closed cohort study. . MedRXiv. 2020.
149. Ismail SA, Saliba V, Lopez Bernal J, Ramsay ME, Ladhani SN. SARS-CoV-2 infection and transmission in educational settings: a prospective, cross-sectional analysis of infection clusters and outbreaks in England. The Lancet Infectious Diseases.
150. Centers for Disease Control and Prevention. COID Data Tracker: Variant Proportions Atlanta, GA: CDC; 2021 [Available from: https://covid.cdc.gov/covid-data-tracker/#variant-proportions.
151. Andrejko KL, Pry J, Myers JF, Jewell NP, Openshaw J, Watt J, et al. Prevention of COVID-19 by mRNA-based vaccines within the general population of California. medRxiv. 2021:2021.04.08.21255135.
152. Baden LR, El Sahly HM, Essink B, Kotloff K, Frey S, Novak R, et al. Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine. New England Journal of Medicine. 2020;384(5):403-16.
153. Haas EJ, Angulo FJ, McLaughlin JM, Anis E, Singer SR, Khan F, et al. Impact and effectiveness of mRNA BNT162b2 vaccine against SARS-CoV-2 infections and COVID-19 cases, hospitalisations, and deaths following a nationwide vaccination campaign in Israel: an observational study using national surveillance data. The Lancet. 2021;397(10287):1819-29.

154. Thompson MG, Burgess JL, Naleway AL, Tyner HL, Yoon SK, Meece J, et al. Interim estimates of vaccine effectiveness of BNT162b2 and mRNA-1273 COVID-19 vaccines in preventing SARS-CoV-2 infection among health care personnel, first responders, and other essential and frontline workers—eight US locations, December 2020–March 2021. Morbidity and Mortality Weekly Report. 2021;70(13):495.
155. Bates D, Maechler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models using lme4. Journal of Statistical Software. 2015;67(1):1-49.
156. Cauchemez S, Bhattarai A, Marchbanks TL, Fagan RP, Ostroff S, Ferguson NM, et al. Role of social networks in shaping disease transmission during a community outbreak of 2009 H1N1 pandemic influenza. Proceedings of the National Academy of Sciences. 2011;108(7):2825-30.
157. Perkins TA, Paz-Soldan VA, Stoddard ST, Morrison AC, Forshey BM, Long KC, et al. Calling in sick: impacts of fever on intra-urban human mobility. Proceedings Biological sciences. 2016;283(1834).
158. Ferguson NM, Laydon D, Nedjati-Gilan G, Imai N, Ainslie K, Baguelin M, et al. Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand Imperial College London: Imperial College COVID-19 Response Team; 2020 [Available from: https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-COVID19-NPI-modelling-16-03-2020.pdf.
159. He X, Lau EH, Wu P, Deng X, Wang J, Hao X, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. Nature medicine. 2020;26(5):672-5.
160. Diekmann O, Heesterbeek JA, Roberts MG. The construction of next-generation matrices for compartmental epidemic models. J R Soc Interface. 2010;7(47):873-85.
161. Kucharski AJ, Russell TW, Diamond C, Liu Y, Edmunds J, Funk S, et al. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. The lancet infectious diseases. 2020.
162. Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. Jama. 2020;323(13):1239-42.
163. Davies NG, Klepac P, Liu Y, Prem K, Jit M, M R, et al. Age-dependent Effects in the Transmission and Control of COVID-19 Epidemics. Nature medicine. 2020.
164. Jing Q-L, Liu M-J, Zhang Z-B, Fang L-Q, Yuan J, Zhang A-R, et al. Household secondary attack rate of COVID-19 and associated determinants in Guangzhou, China: a retrospective cohort study. The Lancet Infectious Diseases. 2020.
165. Laxminarayan R, Wahl B, Dudala SR, Gopal K, Mohan B C, Neelima S, et al. Epidemiology and transmission dynamics of COVID-19 in two Indian states. Science. 2020;370(6517):691-7.
166. Sachdev DD, Brosnan HK, Reid MJA, Kirian M, Cohen SE, Nguyen TQ, et al. Outcomes of Contact Tracing in San Francisco, California—Test and Trace During Shelter-in-Place. JAMA Internal Medicine. 2020.
167. California Department of Public Health. Tracking variants Sacramento, CA: CDPH; 2021 [Available from: https://www.cdph.ca.gov/Programs/CID/DCDC/Pages/COVID-19/COVID-Variants.aspx.
168. Centers for Disease Control and Prevention. Deta variant: What we know about the science Atlanta, GA: CDC; 2021 [Available from: https://www.cdc.gov/coronavirus/2019-ncov/variants/delta-variant.html.

169. Vermund SH, Pitzer VE. Asymptomatic transmission and the infection fatality risk for COVID-19: Implications for school reopening. Clinical Infectious Diseases. 2020.
170. Higdon MM, Wahl B, Jones CB, Rosen JG, Truelove SA, Baidya A, et al. A systematic review of COVID-19 vaccine efficacy and effectiveness against SARS-CoV-2 infection and disease. medRxiv. 2021:2021.09.17.21263549.
171. Lopez Bernal J, Andrews N, Gower C, Gallagher E, Simmons R, Thelwall S, et al. Effectiveness of Covid-19 Vaccines against the B.1.617.2 (Delta) Variant. New England Journal of Medicine. 2021.
172. Self WH. Comparative effectiveness of Moderna, Pfizer-BioNTech, and Janssen (Johnson & Johnson) vaccines in preventing COVID-19 hospitalizations among adults without immunocompromising conditions—United States, March–August 2021. MMWR Morbidity and mortality weekly report. 2021;70.
173. Fuller T, Baker M, Hubler S, Fink S. Santa Clara County: First Known U.S. Coronavirus Death Occurred on Feb. 6. The New York Times. 2020.
174. Henderson J, McCullough E, Treuhaft S. A Profile of Frontline Workers in the Bay Area. Bay Area Equity Atlas 2020 [Available from: https://bayareaequityatlas.org/essential-workers.
175. Prem K, Cook AR, Jit M. Projecting social contact matrices in 152 countries using contact surveys and demographic data. PLoS computational biology. 2017;13(9):e1005697.
176. Watson J, Whiting PF, Brush JE. Interpreting a covid-19 test result. Bmj. 2020;369:m1808.
177. Eikenberry SE, Mancuso M, Iboi E, Phan T, Eikenberry K, Kuang Y, et al. To mask or not to mask: Modeling the potential for face mask use by the general public to curtail the COVID-19 pandemic. Infectious Disease Modelling. 2020;5:293-308.
178. Liang M, Gao L, Cheng C, Zhou Q, Uy JP, Heiner K, et al. Efficacy of face mask in preventing respiratory virus transmission: A systematic review and meta-analysis. Travel medicine and infectious disease. 2020;36:101751.
179. van der Sande M, Teunis P, Sabel R. Professional and home-made face masks reduce exposure to respiratory infections among the general population. PloS one. 2008;3(7):e2618.
180. Uchida M, Kaneko M, Hidaka Y, Yamamoto H, Honda T, Takeuchi S, et al. Effectiveness of vaccination and wearing masks on seasonal influenza in Matsumoto City, Japan, in the 2014/2015 season: An observational study among all elementary schoolchildren. Preventive medicine reports. 2017;5:86-91.
181. Centers for Disease Control and Prevention. COVID Data Tracker: Vaccinations by County Atlanta, GA: CDC; 2021 [Available from: https://covid.cdc.gov/covid-data-tracker/#vaccinations-county-view.
182. USAFACTS. Coronavirus Locations: COVID-19 Map by County and State. 2021 [Available from: https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/.
183. California Hospitals: An Evolving Environment.: California Healthcare Foundation; 2015.
184. Jay J, Bor J, Nsoesie EO, Lipson SK, Jones DK, Galea S, et al. Neighbourhood income and physical distancing during the COVID-19 pandemic in the United States. Nature human behaviour. 2020:1-9.
185. Chamie G, Marquez C, Crawford E, Peng J, Petersen M, Schwab D, et al. SARS-CoV-2 Community Transmission During Shelter-in-Place in San Francisco. medRxiv. 2020:2020.06.15.20132233.

186. Di Domenico L, Pullano G, Sabbatini CE, Boëlle P-Y, Colizza V. Modelling safe protocols for reopening schools during the COVID-19 pandemic in France. Nature Communications. 2021;12(1):1073.
187. Bilinski A, Salomon JA, Giardina J, Ciaranello A, Fitzpatrick MC. Passing the Test: A Model-based analysis of safe school-reopening strategies. medRxiv. 2021:2021.01.27.21250388.
188. Szablewski CM. SARS-CoV-2 Transmission and Infection Among Attendees of an Overnight Camp—Georgia, June 2020. MMWR Morbidity and mortality weekly report. 2020;69.
189. Vlachos J, Hertegård E, B. Svaleryd H. The effects of school closures on SARS-CoV-2 among parents and teachers. Proceedings of the National Academy of Sciences. 2021;118(9):e2020834118.
190. Honein MA, Barrios LC, Brooks JT. Data and Policy to Guide Opening Schools Safely to Limit the Spread of SARS-CoV-2 Infection. Jama. 2021.
191. Centers for Disease Control and Prevention. Operational Strategy for K-12 Schools through Phased Mitigation Atlanta: CDC; 2021 [12]. Available from: https://www.cdc.gov/coronavirus/2019-ncov/community/schools-childcare/operation-strategy.html.
192. Fontanet A, Grant R, Tondeur L, Madec Y, Grzelak L, Cailleau I, et al. SARS-CoV-2 infection in primary schools in northern France: A retrospective cohort study in an area of high transmission. medRxiv. 2020.
193. Keeling MJ, Tildesley MJ, Atkins BD, Penman B, Southall E, Guyver-Fletcher G, et al. The impact of school reopening on the spread of COVID-19 in England. medRxiv. 2020:2020.06.04.20121434.
194. Goldstein E, Lipsitch M, Cevik M. On the effect of age on the transmission of SARS-CoV-2 in households, schools and the community. The Journal of Infectious Diseases. 2020.
195. California Department of Public Health. COVID-19 Public Health Guidance for K-12 Schools in California, 2021-22 School Year Sacramento, CA: State of California; 2021 [Available from: https://www.cdph.ca.gov/Programs/CID/DCDC/Pages/COVID-19/K-12-Guidance-2021-22-School-Year.aspx.
196. Centers for Disease Control and Prevention. Guidance for COVID-19 Prevention in Kindergarten (K)-12 Schools Atlanta, GA: CDC; 2021 [Available from: https://www.cdc.gov/coronavirus/2019-ncov/community/schools-childcare/k-12-guidance.html.
197. Chernozhukov V, Kasahara H, Schrimpf P. The association of opening K–12 schools with the spread of COVID-19 in the United States: County-level panel data analysis. Proceedings of the National Academy of Sciences. 2021;118(42):e2103420118.
198. Gold JA, Gettings JR, Kimball A, Franklin R, Rivera G, Morris E, et al. Clusters of SARS-CoV-2 infection among elementary school educators and students in one school district—Georgia, December 2020–January 2021. Morbidity and Mortality Weekly Report. 2021;70(8):289.
199. Lam-Hine T, McCurdy SA, Santora L, Duncan L, Corbett-Detig R, Kapusinszky B, et al. Outbreak Associated with SARS-CoV-2 B. 1.617. 2 (Delta) Variant in an Elementary School—Marin County, California, May–June 2021. Morbidity and Mortality Weekly Report. 2021;70(35):1214.

200. Stein-Zamir C, Abramson N, Shoob H, Libal E, Bitan M, Cardash T, et al. A large COVID-19 outbreak in a high school 10 days after schools' reopening, Israel, May 2020. Eurosurveillance. 2020;25(29):2001352.
201. Meuris C, Kremer C, Geerinck A, Locquet M, Bruyère O, Defêche J, et al. Transmission of SARS-CoV-2 After COVID-19 Screening and Mitigation Measures for Primary School Children Attending School in Liège, Belgium. JAMA Network Open. 2021;4(10):e2128757-e.
202. Chung E, Chow EJ, Wilcox NC, Burstein R, Brandstetter E, Han PD, et al. Comparison of Symptoms and RNA Levels in Children and Adults With SARS-CoV-2 Infection in the Community Setting. JAMA Pediatrics. 2021.
203. SafeGraph COVID-19 Data Consortium 2020 [Available from: https://www.safegraph.com/covid-19-data-consortium.
204. Kretzschmar ME, Rozhnova G, Bootsma MCJ, van Boven M, van de Wijgert JHHM, Bonten MJM. Impact of delays on effectiveness of contact tracing strategies for COVID-19: a modelling study. The Lancet Public Health. 2020;5(8):e452-e9.
205. Yang S, Wu J, Ding C, Cui Y, Zhou Y, Li Y, et al. Epidemiological features of and changes in incidence of infectious diseases in China in the first decade after the SARS outbreak: an observational trend study. The Lancet Infectious Diseases. 2017;17(7):716-25.
206. Zheng Y, Jit M, Wu JT, Yang J, Leung K, Liao Q, et al. Economic costs and health-related quality of life for hand, foot and mouth disease (HFMD) patients in China. PloS one. 2017;12(9):e0184266.
207. Mao Q-y, Wang Y, Bian L, Xu M, Liang Z. EV71 vaccine, a new tool to control outbreaks of hand, foot and mouth disease (HFMD). Expert Review of Vaccines. 2016;15(5):599-606.
208. Li R, Liu L, Mo Z, Wang X, Xia J, Liang Z, et al. An Inactivated Enterovirus 71 Vaccine in Healthy Children. New England Journal of Medicine. 2014;370(9):829-37.
209. Zhu F, Xu W, Xia J, Liang Z, Liu Y, Zhang X, et al. Efficacy, Safety, and Immunogenicity of an Enterovirus 71 Vaccine in China. New England Journal of Medicine. 2014;370(9):818-28.
210. Wei M, Meng F, Wang S, Li J, Zhang Y, Mao Q, et al. 2-Year Efficacy, Immunogenicity, and Safety of Vigoo Enterovirus 71 Vaccine in Healthy Chinese Children: A Randomized Open-Label Study. Journal of Infectious Diseases. 2017;215(1):56-63.
211. Zhu FC, Meng FY, Li JX, Li XL, Mao QY, Tao H, et al. Efficacy, safety, and immunology of an inactivated alum-adjuvant enterovirus 71 vaccine in children in China: a multicentre, randomised, double-blind, placebo-controlled, phase 3 trial. Lancet. 2013;381(9882):2024-32.
212. Xinhua. Hand, foot and mouth disease on rise in Beijing. ChinaDailycom. 2018.
213. Zhang J. [Trend of epidemics and variation of pathogens of hand, foot and mouth disease in China: a dynamic series analysis, 2008-2017]. Zhonghua liu xing bing xue za zhi = Zhonghua liuxingbingxue zazhi. 2019;40(2):147-54.
214. Liu L, Zhao X, Yin F, Lv Q. Spatio-temporal clustering of hand, foot and mouth disease at the county level in Sichuan province, China, 2008-2013. Epidemiol Infect. 2015;143(4):831-8.
215. Sichuan SBo. Sichuan Statistical yearbook 2012-2018. Beijing: China Statistics Press; 2018.
216. Wang J, Cao Z, Zeng DD, Wang Q, Wang X, Qian H. Epidemiological Analysis, Detection, and Comparison of Space-Time Patterns of Beijing Hand-Foot-Mouth Disease (2008–2012). PloS one. 2014;9(3):e92745.

217. Yang W, Li Z, Lan Y, Wang J, Ma J, Jin L, et al. A nationwide web-based automated system for outbreak early detection and rapid response in China. Western Pacific surveillance and response journal : WPSAR. 2011;2(1):10-5.
218. China Center of Disease Control and Prevention. Guideline of HFMD diagnosis and treatment in 2008. China CDC; 2009.
219. Omaña-Cepeda C, Martínez-Valverde A, del Mar Sabater-Recolons M, Jané-Salas E, Marí-Roig A, López-López J. A literature review and case report of hand, foot and mouth disease in an immunocompetent adult. BMC research notes. 2016;9(1):165.
220. Ventarola D, Bordone L, Silverberg N. Update on hand-foot-and-mouth disease. Clinics in dermatology. 2015;33(3):340-6.
221. Wang Y, Feng Z, Yang Y, Self S, Gao Y, Longini IM, et al. Hand, foot, and mouth disease in China: patterns of spread and transmissibility. Epidemiology. 2011;22(6):781-92.
222. Schuck-Paim C, Taylor RJ, Alonso WJ, Weinberger DM, Simonsen L. Effect of pneumococcal conjugate vaccine introduction on childhood pneumonia mortality in Brazil: a retrospective observational study. The Lancet Global Health. 2019;7(2):e249-e56.
223. Baker JM, Tate JE, Steiner CA, Haber MJ, Parashar UD, Lopman BA. Longer-term direct and indirect effects of infant rotavirus vaccination across all ages in the US; 2000 - 2013: analysis of a large hospital discharge dataset. Clinical Infectious Diseases. 2018:ciy580-ciy.
224. do Carmo GM, Yen C, Cortes J, Siqueira AA, de Oliveira WK, Cortez-Escalante JJ, et al. Decline in diarrhea mortality and admissions after routine childhood rotavirus immunization in Brazil: a time-series analysis. PLoS Med. 2011;8(4):e1001024.
225. Bergmeir C, Benítez JM. On the use of cross-validation for time series predictor evaluation. Information Sciences. 2012;191:192-213.
226. Roberts DR, Bahn V, Ciuti S, Boyce MS, Elith J, Guillera-Arroita G, et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography. 2017;40(8):913-29.
227. Breiman L. Random Forests. Machine Learning. 2011;45:5-32.
228. Meinshausen N. Quantile regression forests. Journal of Machine Learning Research. 2006;7(Jun):983-99.
229. Kürüm E, Warren JL, Schuck-Paim C, Lustig R, Lewnard JA, Fuentes R, et al. Bayesian Model Averaging with Change Points to Assess the Impact of Vaccination and Public Health Interventions. Epidemiology. 2017;28(6):889-97.
230. Bernal JL, Cummins S, Gasparrini A. Interrupted time series regression for the evaluation of public health interventions: a tutorial. Int J Epidemiol. 2017;46(1):348-55.
231. Wang W, Song J, Wang J, Li Y, Deng H, Li M, et al. Cost-effectiveness of a national enterovirus 71 vaccination program in China. PLOS Neglected Tropical Diseases. 2017;11(9):e0005899.
232. Takahashi S, Metcalf CJE, Arima Y, Fujimoto T, Shimizu H, Rogier van Doorn H, et al. Epidemic dynamics, interactions and predictability of enteroviruses associated with hand, foot and mouth disease in Japan. Journal of The Royal Society Interface. 2018;15(146).
233. Araki S, Shima M, Yamamoto K. Spatiotemporal land use random forest model for estimating metropolitan NO2 exposure in Japan. The Science of the total environment. 2018;634:1269-77.

234. Hansen V, Oren E, Dennis LK, Brown HE. Infectious disease mortality trends in the United States, 1980-2014. Jama. 2016;316(20):2149-51.
235. Binder S, Levitt AM. Preventing emerging infectious diseases: a strategy for the 21st century: overview of the updated CDC plan. 1998.
236. Gardy JL, Loman NJ. Towards a genomics-informed, real-time, global pathogen surveillance system. Nature Reviews Genetics. 2018;19(1):9-20.
237. Halliday JEB, Meredith AL, Knobel DL, Shaw DJ, Bronsvoort BMdC, Cleaveland S. A framework for evaluating animals as sentinels for infectious disease surveillance. Journal of the Royal Society, Interface. 2007;4(16):973-84.
238. Pauvolid-Corrêa A, Gonçalves Dias H, Marina Siqueira Maia L, Porfírio G, Oliveira Morgado T, Sabino-Santos G, et al. Zika Virus Surveillance at the Human-Animal Interface in West-Central Brazil, 2017-2018. Viruses. 2019;11(12).
239. Wallinga J, Teunis P, Kretzschmar M. Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents. American journal of epidemiology. 2006;164(10):936-44.
240. Bi Q, Wu Y, Mei S, Ye C, Zou X, Zhang Z, et al. Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. The Lancet Infectious Diseases. 2020.
241. Guan W-j, Ni Z-y, Hu Y, Liang W-h, Ou C-q, He J-x, et al. Clinical Characteristics of Coronavirus Disease 2019 in China. New England Journal of Medicine. 2020;382(18):1708-20.
242. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. New England Journal of Medicine. 2020;382(13):1199-207.
243. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. Annals of Internal Medicine. 2020;172(9):577-82.
244. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The lancet. 2020;395(10223):497-506.
245. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus–Infected Pneumonia in Wuhan, China. Jama. 2020;323(11):1061-9.
246. Lewnard JA, Liu VX, Jackson ML, Schmidt MA, Jewell BL, Flores JP, et al. Incidence, clinical outcomes, and transmission dynamics of severe coronavirus disease 2019 in California and Washington: prospective cohort study. BMJ. 2020;369:m1923.
247. Goldstein E, Lipsitch M, Cevik M. On the effect of age on the transmission of SARS-CoV-2 in households, schools and the community. medRxiv. 2020.
248. Bunyavanich S, Do A, Vicencio A. Nasal Gene Expression of Angiotensin-Converting Enzyme 2 in Children and Adults. Jama. 2020;323(23):2427-9.
249. Mizumoto K, Omori R, Nishiura H. Age specificity of cases and attack rate of novel coronavirus disease (COVID-19). medRxiv. 2020:2020.03.09.20033142.
250. Li W, Zhang B, Lu J, Liu S, Chang Z, Peng C, et al. Characteristics of Household Transmission of COVID-19. Clinical Infectious Diseases. 2020.
251. Posfay-Barbe KM, Wagner N, Gauthey M, Moussaoui D, Loevy N, Diana A, et al. COVID-19 in Children and the Dynamics of Infection in Families. Pediatrics. 2020:e20201576.

252. Danis K, Epaulard O, Bénet T, Gaymard A, Campoy S, Botelho-Nevers E, et al. Cluster of Coronavirus Disease 2019 (COVID-19) in the French Alps, February 2020. Clin Infect Dis. 2020;71(15):825-32.
253. Wu Q, Xing Y, Shi L, Li W, Gao Y, Pan S, et al. Coinfection and Other Clinical Characteristics of COVID-19 in Children. Pediatrics. 2020;146(1):e20200961.
254. Gudbjartsson DF, Helgason A, Jonsson H, Magnusson OT, Melsted P, Norddahl GL, et al. Spread of SARS-CoV-2 in the Icelandic Population. New England Journal of Medicine. 2020;382(24):2302-15.
255. Dong Y, Mo X, Hu Y, Qi X, Jiang F, Jiang Z, et al. Epidemiology of COVID-19 Among Children in China. Pediatrics. 2020;145(6):e20200702.
256. Zhu Y, Bloxham CJ, Hulme KD, Sinclair JE, Tong ZWM, Steele LE, et al. Children are unlikely to have been the primary source of household SARS-CoV-2 infections. medRxiv. 2020:2020.03.26.20044826.
257. Jones TC, Mühlemann B, Veith T, Biele G, Zuchowski M, Hoffmann J, et al. An analysis of SARS-CoV-2 viral load by patient age. medRxiv. 2020.
258. Stoye J. A critical assessment of some recent work on COVID-19. arXiv preprint arXiv:200510237. 2020.
259. Lennon NJ, Bhattacharyya RP, Mina MJ, Rehm HL, Hung DT, Smole S, et al. Comparison of viral levels in individuals with or without symptoms at time of COVID-19 testing among 32,480 residents and staff of nursing homes and assisted living facilities in Massachusetts. medRxiv. 2020:2020.07.20.20157792.
260. Brotons P, Launes C, Buetas E, Fumado V, Henares D, de Sevilla MF, et al. Susceptibility to Sars-COV-2 Infection Among Children And Adults: A Seroprevalence Study of Family Households in the Barcelona Metropolitan Region, Spain. Clinical Infectious Diseases. 2020.
261. California Department of Public Health. California Open Data Portal 2020 [Available from: https://data.ca.gov/dataset/covid-19-hospital-data/resource/42d33765-20fd-44b8-a978-b083b7542225
262. Ng D, Goldgof G, Shy B, Levine A, Balcerek J, Bapat SP, et al. SARS-CoV-2 seroprevalence and neutralizing activity in donor and patient blood from the San Francisco Bay Area. medRxiv. 2020:2020.05.19.20107482.
263. Bendavid E, Mulaney B, Sood N, Shah S, Ling E, Bromley-Dulfano R, et al. COVID-19 Antibody Seroprevalence in Santa Clara County, California. medRxiv. 2020:2020.04.14.20062463.
264. Census Bureau. U.S. Census Bureau Quickfacts: Oakland City, California 2020 [Available from: https://www.census.gov/quickfacts/oaklandcitycalifornia.
265. Bay Area Census - City of Oakland 2020 [Available from: http://www.bayareacensus.ca.gov/cities/Oakland.htm.
266. Healthy Alameda County. Healthy Alameda County :: Indicators :: Single-Parent 2020 [Available from: http://www.healthyalamedacounty.org/indicators/index/view?indicatorId=411&localeId=238.
267. Census Bureau. Households With Grandparents Living With Own Grandchildren Under 18 Years By Responsibility For Own Grandchildren And Presence Of Parent Of Grandchildren 2020 [Available from:

https://data.census.gov/cedsci/table?q=B10063&tid=ACSDT5Y2017.B10063&hidePreview=true
.
268. Bui Q, Miller CC. The age gap that women have babies: How a gap divides America. New York Times. 2018.
269. California Department of Education. Enrollment by school - School & school data files 2020 [Available from: https://www.cde.ca.gov/ds/sd/sd/filesenr.asp.
270. Ed-Data.org. EdData - County summary - Alameda 2020 [Available from: http://www.ed-data.org/district/Alameda/.
271. Prem K, Liu Y, Russell TW, Kucharski AJ, Eggo RM, Davies N, et al. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. The Lancet Public Health. 2020.
272. Verity R, Okell LC, Dorigatti I, Winskill P, Whittaker C, Imai N, et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. The Lancet infectious diseases. 2020.
273. Ver Hoef JM, Boveng PL. Quasi-Poisson vs. negative binomial regression: how should we model overdispersed count data? Ecology. 2007;88(11):2766-72.

Supplemental Information for Chapter 1

Justification of model choice

Our analysis makes use of several modelling techniques, which are each suited to obtain different target parameters of interest. Here, we describe the justification for each technique in greater detail.

Table 7. Model choice depended on the research question and associated target parameter.

Purpose of analysis	Analysis used	Target parameter	Justification for analysis
Estimate the nonlinear relationships between lagged temperature and precipitation and coccidioidomycosis incidence.	Distributed lag generalized additive model at the county-level, with meta-analysis	Incidence Rate Ratio expressing incidence rate of coccidioidomycosis at some temperature or precipitation level to the incidence rate at another level.	Here, the target parameter of interest can be directly expressed from the model coefficients on lagged temperature and precipitation. This approach provides the level of flexibility needed, while permitting Wald-based inference in the form of standard errors. Splines permit the modelling of non-linear relationships. Modelling at the county-level and creating a pooled relationship using meta-analysis permits examination of the factors that explain heterogeneity across space.
Estimate how distal climatic factors modify the effect of more recent climatic factors on coccidioidomycosis incidence.	Distributed lag generalized additive model with interaction terms	Model coefficient on the interaction term, reflecting the ratio of incidence rate ratios.	Here the target parameter of interest can be directly expressed from the model coefficients on the interaction term. This approach provides the level of flexibility needed, while permitting Wald-based inference in the form of standard errors.
Estimate the number of cases averted by or attributable to droughts in California between 2000 and 2020.	Ensemble prediction model	The absolute difference in cases between observed drought conditions, and counterfactual non-drought conditions, expressed as $E(Y) - E(Y0)$.	Here, the target parameter of interest is the difference in model predictions, rather than a simple model coefficient. Since we modelled the full time series, a greater level of flexibility in model form was needed to allow for different effects of climatic lags across seasons. This approach provides the most flexibility, while enabling robust predictions. The expense is that model-based inference is limited.

Multilevel distributed lag nonlinear models

County-specific models

We estimated relationships using a quasi-Poisson likelihood approach, with monthly cases as the outcome variable and the log of each census tract's population as an offset term, so that model coefficients reflected the log incidence rate ratio. The primary exposure variables were lagged total precipitation and mean temperature. These were modeled with natural cubic spline functions of smoothed three-month averages, with lags spanning 1 to 36 months prior to estimated date of onset. For each of the 36 lags, we controlled for other lags every three (precipitation) or six (temperature) months. This approach allowed us to generate distributed lag models in which we examined the lagged effect of precipitation and temperature across all 36 months, while accounting for the historical precipitation and temperature history over the time period of interest in a way that avoided over fitting. To determine the location of knots for the cubic spline, we systematically varied the location of internal knots placed at precipitation or temperatures corresponding to average percentiles across counties [75], selecting the model that minimized the sum of Q-AIC across all counties, where the Q-AIC is a modification of the Akaike information criterion (AIC) for quasi-likelihood models [77]. We also included a natural cubic spline for soil type (percent sand), as it may be correlated with precipitation or temperature and the outcome but not on the causal pathway between the exposure and outcome (e.g., as vegetation might be). A cubic spline on year was also included to account for reporting and other secular trends not due to climate. Model formulas expressing $Y_{i,t}$, the number of cases in census tract *i* in month *t*, for rainfall (equation 1) and temperature (equation 2) were:

For $j \in [1,36]$:

$$g(\mu_{i,t}) = \log\left(population_{i,t}\right) + \beta_0 + s(year_t) + s(\%sand_i) + s\left(precipitation_{i,t-j}\right) + \sum_{k=1}^{6} s(temperature_{i,t+3-6k}) + \sum_{k=1}^{12} s\left(precipitation_{i,t-3k}\right) \quad [1]$$

$$g(\mu_{i,t}) = \log\left(population_{i,t}\right) + \beta_0 + s(year_t) + s(\%sand_i) + s\left(temperature_{i,t-j}\right) + \sum_{k=1}^{6} s(temperature_{i,t+3-6k}) + \sum_{k=1}^{12} s\left(precipitation_{i,t-3k}\right) \quad [2]$$

Where $g()$ is a log function of the expectation $\mu_{i,t} = E\left(Y_{i,t}\right)$.

Pooled effects

Then, we used a fixed-effects meta-analysis to pool estimates of county-specific associations of precipitation and temperature with incidence [75]. We examined the overall shape of the associations between precipitation, temperature and incidence at each of the 36 monthly lags, and calculated the pooled IRR associated with an increase of precipitation or temperature from the 25th percentile to the 75th percentile (i.e., one interquartile range; IQR) over each lag.

Heterogeneity in the temperature-incidence and precipitation-incidence relationships across counties was additionally assessed using multivariate meta-regression models. This approach extends the fixed-effects meta-analysis to include meta-predictors, that summarize county-level information (e.g., mean county total precipitation, mean county temperature). The significance of the meta-predictors is determined using Wald tests [75]. For significant meta-predictors, we plotted separate exposure-response relationships across our set of meta-predictor values.

Effect modification

To examine potential effect modification of the wet winter period (which was determined by the fixed-effects meta-analysis to be the strongest predictor of fall coccidioidomycosis incidence) by antecedent conditions, we created a nonlinear interaction term between winter precipitation and antecedent conditions by multiplying the basis function for precipitation at a nine-month lag by a binary indicator for whether or not the census tract had a wetter or drier than average winter in the two years prior to exposure, three years prior to exposure, or both. The target parameter, the exponentiated coefficient on the modelled interaction term, is thus expressed as the ratio of the IRR for an IQR increase in total monthly winter precipitation following a dry year to the IRR for winter precipitation following a wet year. Deviations from mean were determined by calculating the percent difference of the monthly census tract value to the monthly county mean across the full time period examined. We used a single-stage distributed lag generalized additive model that adjusted for county as a fixed effect, and used cluster robust standard errors to adjust for the repeated observations at the census tract-level. All other model terms were the same as in the two stage analysis. The model formulas to examine effect modification by a drier than average winter two years ago, and two and three years ago, respectively, were:

$$g(\mu_{i,t}) = \log(population_{i,t}) + \beta_0 + s(year_t) + s(\%sand_i) + s(precipitation_{i,t-9}) + \beta_1 \mathbb{I}(precip.deviation_{i,t-21} < 0) + s(precipitation_{i,t-9}) * \mathbb{I}(precip.deviation_{i,t-21} < 0) + \sum_{k=1}^{6} s(temperature_{i,t+3-6k}) + \sum_{k=1}^{12} s(precipitation_{i,t-3k}) + \sum_{k=1}^{13} \mathbb{I}(county_i = k) \quad [3]$$

$$g(\mu_{i,t}) = \log(population_{i,t}) + \beta_0 + s(year_t) + s(\%sand_i) + s(precipitation_{i,t-9}) + \beta_1 \mathbb{I}(precip.deviation_{i,t-21} < 0 \,\&precip.deviation_{i,t-30} < 0) + s(precipitation_{i,t-9}) * \mathbb{I}(precip.deviation_{i,t-21} < 0 \,\&precip.deviation_{i,t-30} < 0) + \sum_{k=1}^{6} s(temperature_{i,t+3-6k}) + \sum_{k=1}^{12} s(precipitation_{i,t-3k}) + \sum_{k=1}^{13} \mathbb{I}(county_i = k) \quad [4]$$

All statistical analyses were conducted in R, version 3.3.1 (R Foundation for Statistical Computing), using the splines and dlnm package for fitting distributed lag generalized additive models and the mvmeta package for performing fixed-effect meta-regression models [75, 78].

Ensemble modeling

We estimated cases attributable to—or averted because of—major droughts in California between 2000 and 2020 using a powerful, flexible ensemble modelling approach to predict incidence under counterfactual scenarios with regard to the presence or absence of drought [17]. For each county, we modelled the monthly incidence (for all months, January through February) at the census tract-level using five generalized linear models (GLMs) and one random forest algorithm. Table 2 summarizes the variables included in each of the six models. Model formulas expressing $Y_{i,t}$, the number of cases in census tract *i* in month *t*, for the five GLMs were:

Model 1

$$g(\mu_{i,t}) = \log\left(population_{i,t}\right) + \beta_0 + s(year_t) + \sum_{k=1}^{3} \beta_k \mathbb{I}(season_t = k) \quad [5]$$

Where $g()$ is a log function of the expectation $\mu_{i,t} = E\left(Y_{i,t}\right)$

Model 2

$$g(\mu_{i,t}) = \log\left(population_{i,t}\right) + \beta_0 + s(year_t) + \sum_{k=1}^{3} \beta_k \mathbb{I}(season_t = k) + \beta_4(\%sand_i) + \beta_5(impervious\_surface_i) + \beta_6(elevation_i) + \beta_7\left(temperature_{i,t-1}\right) + \beta_8\left(precipitation_{i,t-1}\right) + \sum_{k=1}^{12} \alpha_k(temperature_{i,t-3*k}) + \sum_{k=1}^{12} \gamma_k\left(precipitation_{i,t-3*k}\right) \quad [6]$$

Model 3

$$g(\mu_{i,t}) = \log\left(population_{i,t}\right) + \beta_0 + s(year_t) + \sum_{k=1}^{3} \beta_k \mathbb{I}(season_t = k) + \beta_4(\%sand_i) + \beta_5(impervious\_surface_i) + \beta_6(elevation_i) + \beta_7\left(temperature_{i,t-1}\right) + \beta_8\left(precipitation_{i,t-1}\right) + \sum_{k=1}^{12} \alpha_k(temperature_{i,t-3*k}) + \sum_{k=1}^{12} \gamma_k\left(precipitation_{i,t-3k}\right) + \beta_9 \mathbb{I}\left(1yr\_post\_drought_{i,t}\right) + \beta_{10} \mathbb{I}\left(2yr\_post\_drought_{i,t}\right) + \sum_{k=1}^{4} \theta_k\left(precipitation_{i,t-3k}\right) * \mathbb{I}\left(1yr\_post\_drought_{i,t}\right) + \sum_{k=1}^{4} \pi_k(precipitation_{i,t-3*k}) * \mathbb{I}\left(2yr\_post\_drought_{i,t}\right) \quad [7]$$

Model 4

$$g(\mu_{i,t}) = \log\left(population_{i,t}\right) + \beta_0 + s(year_t) + \sum_{j=1}^{3} \beta_j \mathbb{I}(season_t = j) + \beta_4(\%sand_i) + \beta_5(impervious\_surface_i) + \beta_6(elevation_i) + \beta_7\left(temperature_{i,t-1}\right) + \beta_8\left(precipitation_{i,t-1}\right) + \sum_{k=1}^{4} \alpha_k(temperature_{i,t-3k}) + \sum_{k=1}^{4} \gamma_k\left(precipitation_{i,t-3k}\right) + \sum_{j=1}^{2} \sum_{k=1}^{4} \theta_{jk}\left(precipitation_{i,t-3k}\right) * \mathbb{I}\left(precipitation.dev_{i,t-(3k-12j)} \leq -0.5\right) + \sum_{j=1}^{2} \sum_{k=1}^{4} \theta_{jk}\left(precipitation_{i,t-3k}\right) * \mathbb{I}\left(days.above.30C_{i,t-(3k-12j)} \geq 7\right) \quad [8]$$

Model 5

$$g(\mu_{i,t}) = \log\left(population_{i,t}\right) + \beta_0 + s(year_t) + \sum_{j=1}^{3} \beta_j \mathbb{I}(season_t = j) + \beta_4(\%sand_i) + \beta_5(impervious\_surface_i) + \beta_6(elevation_i) + \beta_7\left(temperature_{i,t-1}\right) + \beta_8\left(precipitation_{i,t-1}\right) + \sum_{k=1}^{12} \alpha_k(temperature_{i,t-3k}) + \sum_{k=1}^{12} \gamma_k\left(precipitation_{i,t-3k}\right) + \beta_9 \mathbb{I}\left(1yr\_post\_drought_{i,t}\right) + \beta_{10} \mathbb{I}\left(2yr\_post\_drought_{i,t}\right) + \sum_{j=1}^{3} \sum_{k=1}^{12} \theta_{jk}\left(precipitation_{i,t-3k}\right) * \mathbb{I}(season_t = j) + \sum_{j=1}^{3} \sum_{k=1}^{12} \pi_{jk}\left(temperature_{i,t-3k}\right) * \mathbb{I}(season_t = j) + \sum_{j=1}^{3} \beta_{k+10} \mathbb{I}(season_t = j) * \left(precipitation_{i,t-1}\right) + \sum_{j=1}^{3} \beta_{k+13} \mathbb{I}(season_t = j) * \left(temperature_{i,t-1}\right) + \sum_{k=1}^{3} \beta_{k+16} \mathbb{I}(season_t = j) * \mathbb{I}\left(1yr\_post\_drought_{i,t}\right) + \sum_{j=1}^{3} \beta_{k+19} \mathbb{I}(season_t = j) * \mathbb{I}\left(2yr\_post\_drought_{i,t}\right) \quad [9]$$

Table 8. Variables included in the 6 models used in the ensemble model. Orange cells are used as main effects, and green cells were used as interactions. Models 1-5 were generalized linear models (GLM) and model 6 was random forest (RF)

	1 GLM	2 GLM	3 GLM	4 GLM	5 GLM	6 RF
Population (as an offset)						
Year (as a natural spline)						
Season (as a factor)						
Percent sand						
Impervious surface						
Elevation						
Total rainfall						
Lag 1 month						
Lag 3 month						
Lag 6 month						
Lag 9 month						
Lag 12 month						
Lag 15 month						
Lag 18 month						
Lag 21 month						
Lag 24 month						
Lag 27 month						
Lag 30 month						
Lag 33 month						
Lag 36 month						
Average temperature						
Lag 1 month						
Lag 3 month						
Lag 6 month						
Lag 9 month						
Lag 12 month						
Lag 15 month						
Lag 18 month						
Lag 21 month						
Lag 24 month						
Lag 27 month						
Lag 30 month						
Lag 33 month						
Lag 36 month						
𝕀(One year post drought) (indicator)						
𝕀(Two years post drought) (indicator)						
𝕀(Rainfall Deviation < -0.5) (indicator)						
Lag 15 month						
Lag 18 month						
Lag 21 month						
Lag 24 month						
Lag 27 month						
Lag 30 month						
Lag 33 month						
Lag 36 month						
𝕀(Days above 30°C ≥ 7) (indicator)						
Lag 15 month						
Lag 18 month						
Lag 21 month						
Lag 24 month						
Lag 27 month						
Lag 30 month						
Lag 33 month						
Lag 36 month						

The sum of squared errors for each algorithm was calculated using leave-out-one-year cross-validation whereby the model was fit for all but one year of the time period, and then used to predict the out-of-sample cases in the left out year. We calculated a weight for each algorithm where the weights equaled the inverse of the cross-validated risk. Ensemble model predictions were then made for each census tract in each month as a weighted average of the individual model predictions.

We used the model to predict the total number of cases in a county over a certain time period under observed conditions, following the equation:

$$\hat{E}(Y) = \sum_{t=t0}^{t=T} \sum_{i=1}^{N} \hat{E}(Y_{i,t} | A_{i,t} = A_{i,t}, W_{i,t}) \quad [10]$$

where:

$A_{i,t}$ represents the observed values of temperature and precipitation in census tract *i* at time *t*

$W_{i,t}$ represents the observed values for the other covariates in the model in census tract *i* at time *t*

$Y_{i,t}$ is the number of cases in census tract *i* at time *t* given observed values of temperature and precipitation and covariates

N is the total number of census tracts in the county

$t \in (t0, T)$ represents the period of time of interest

The ensemble model was then used to predict the counterfactual number of cases, $\hat{E}(Y_0)$, that would have been observed during the study period if, possibly contrary to what was observed, monthly average temperature higher than the historical average and total monthly precipitation below the historical average were deterministically set to their monthly county-level means, following the equation:

$$\hat{E}(Y_0) = \sum_{t=t0}^{t=T} \sum_{i=1}^{N} \hat{E}(Y_{i,t} | A_{i,t} = a_{i,t}, W_{i,t}) \quad [11]$$

where:

$a_{i,,t}$ represents the counterfactual values for temperature and precipitation in census tract *i* at time *t*, obtained by setting any value for temperature above the mean equal to the mean, and any value for precipitation below the mean equal to the mean.

$W_{i,t}$ represents the observed values for the other covariates in the model in census tract *i* at time *t*

The incident cases attributable to—or averted by—the drought was estimated as the difference between predicted cases under the observed conditions and those predicted under

the counterfactual “average climate” scenario. Because antecedent conditions as far back as three years may carry influence, we examined the attributable incidence during the drought and in the two years following the end of the drought. We estimated the effects separately for the severe drought spanning May 2012 until October 2015, and for the less severe drought from March 2007 to November 2009. We calculated the target parameter, $\hat{\psi}$, as follows, using equations [4] and [5]:

$$\hat{\psi} = \hat{E}(Y) - \hat{E}(Y_0) \quad [12]$$

Because seasonality of coccidioidomycosis in California is such that incidence is lowest in March-April, with peaks occurring in the fall, we considered the change in incident cases “during drought” to include the period starting at the onset of the drought and extending until the end of the transmission season following the drought (e.g., March 31, 2010; March 31, 2016). The two years post drought encompassed the full epidemiological seasons following the drought (e.g., April 1, 2010 – March 31, 2012; April 1, 2016 – March 31, 2018).

Table S1. Incidence rate ratios (IRRs) for fall coccidioidomycosis incidence associated with a one-unit increase in the IQR of temperature or precipitation

Month lag	IRR (95% CI) associated with a one unit increase in precipitation IQR	IQR of precipitation (mm)	IRR (95% CI) associated with a one unit increase in temperature IQR	IQR of temperature (°C)
1	0.87 (0.80, 0.94)	[0.3, 11.2]	1.29 (1.16, 1.44)	[18.1, 23.8]
2	0.89 (0.85, 0.94)	[0, 2.5]	1.55 (1.32, 1.82)	[20.6, 25.6]
3	0.95 (0.92, 0.98)	[0, 2]	2.02 (1.84, 2.22)	[20.3, 25.8]
4	0.87 (0.81, 0.94)	[0.2, 8.3]	0.58 (0.5, 0.66)	[18.3, 24.5]
5	1.38 (1.29, 1.48)	[2.3, 20.7]	0.49 (0.44, 0.54)	[15.9, 21.3]
6	1.45 (1.36, 1.55)	[9.5, 41.1]	0.5 (0.46, 0.55)	[13.6, 17.8]
7	1.36 (1.28, 1.45)	[19.1, 57.4]	0.58 (0.54, 0.63)	[11.7, 15.1]
8	1.39 (1.29, 1.49)	[26.9, 71.3]	0.57 (0.51, 0.64)	[10.1, 13.1]
9	1.45 (1.36, 1.55)	[27.1, 73.2]	0.74 (0.69, 0.79)	[9.5, 12]
10	1.29 (1.20, 1.38)	[21.2, 64.4]	0.83 (0.76, 0.91)	[9.6, 13.4]
11	0.90 (0.84, 0.96)	[11.8, 44.4]	0.98 (0.89, 1.07)	[11.6, 17.2]
12	0.94 (0.87, 1.01)	[3.8, 24.7]	1.11 (1.02, 1.2)	[14.4, 20.6]
13	1.01 (0.94, 1.09)	[0.3, 11.2]	1.23 (1.11, 1.35)	[18.1, 23.8]
14	1.05 (1.01, 1.09)	[0, 2.5]	1.29 (1.11, 1.5)	[20.6, 25.6]
15	1.04 (1.00, 1.08)	[0, 2]	1.16 (0.95, 1.41)	[20.3, 25.8]
16	1.24 (1.15, 1.33)	[0.2, 8.3]	0.79 (0.67, 0.94)	[18.3, 24.5]
17	0.99 (0.93, 1.07)	[2.3, 20.7]	0.77 (0.68, 0.88)	[15.9, 21.3]
18	1.00 (0.93, 1.07)	[9.5, 41.1]	0.8 (0.73, 0.88)	[13.6, 17.8]
19	1.03 (0.96, 1.10)	[19.1, 57.4]	0.87 (0.79, 0.95)	[11.7, 15.1]
20	0.97 (0.92, 1.03)	[26.9, 71.3]	0.88 (0.78, 1)	[10.1, 13.1]
21	0.93 (0.88, 0.98)	[27.1, 73.2]	1.35 (1.26, 1.44)	[9.5, 12]
22	0.91 (0.86, 0.97)	[21.2, 64.4]	1.31 (1.19, 1.43)	[9.6, 13.4]
23	1.02 (0.95, 1.09)	[11.8, 44.4]	1.22 (1.11, 1.34)	[11.6, 17.2]
24	1.08 (1.01, 1.16)	[3.8, 24.7]	1.2 (1.11, 1.31)	[14.4, 20.6]
25	0.85 (0.79, 0.92)	[0.3, 11.2]	1.28 (1.16, 1.41)	[18.1, 23.8]
26	0.98 (0.94, 1.02)	[0, 2.5]	1.24 (1.08, 1.43)	[20.6, 25.6]
27	1.06 (1.01, 1.12)	[0, 2]	1.24 (1, 1.54)	[20.3, 25.8]
28	1.10 (1.02, 1.19)	[0.2, 8.3]	1.05 (0.88, 1.25)	[18.3, 24.5]
29	0.74 (0.69, 0.79)	[2.3, 20.7]	0.99 (0.88, 1.12)	[15.9, 21.3]
30	0.62 (0.58, 0.67)	[9.5, 41.1]	1.03 (0.94, 1.12)	[13.6, 17.8]
31	0.78 (0.73, 0.84)	[19.1, 57.4]	0.98 (0.9, 1.06)	[11.7, 15.1]
32	0.87 (0.82, 0.93)	[26.9, 71.3]	1.03 (0.92, 1.15)	[10.1, 13.1]
33	0.81 (0.77, 0.86)	[27.1, 73.2]	0.99 (0.92, 1.06)	[9.5, 12]
34	0.80 (0.76, 0.85)	[21.2, 64.4]	1.13 (1.04, 1.22)	[9.6, 13.4]
35	0.90 (0.84, 0.96)	[11.8, 44.4]	1.07 (0.98, 1.16)	[11.6, 17.2]
36	0.90 (0.84, 0.96)	[3.8, 24.7]	1.1 (1, 1.2)	[14.4, 20.6]

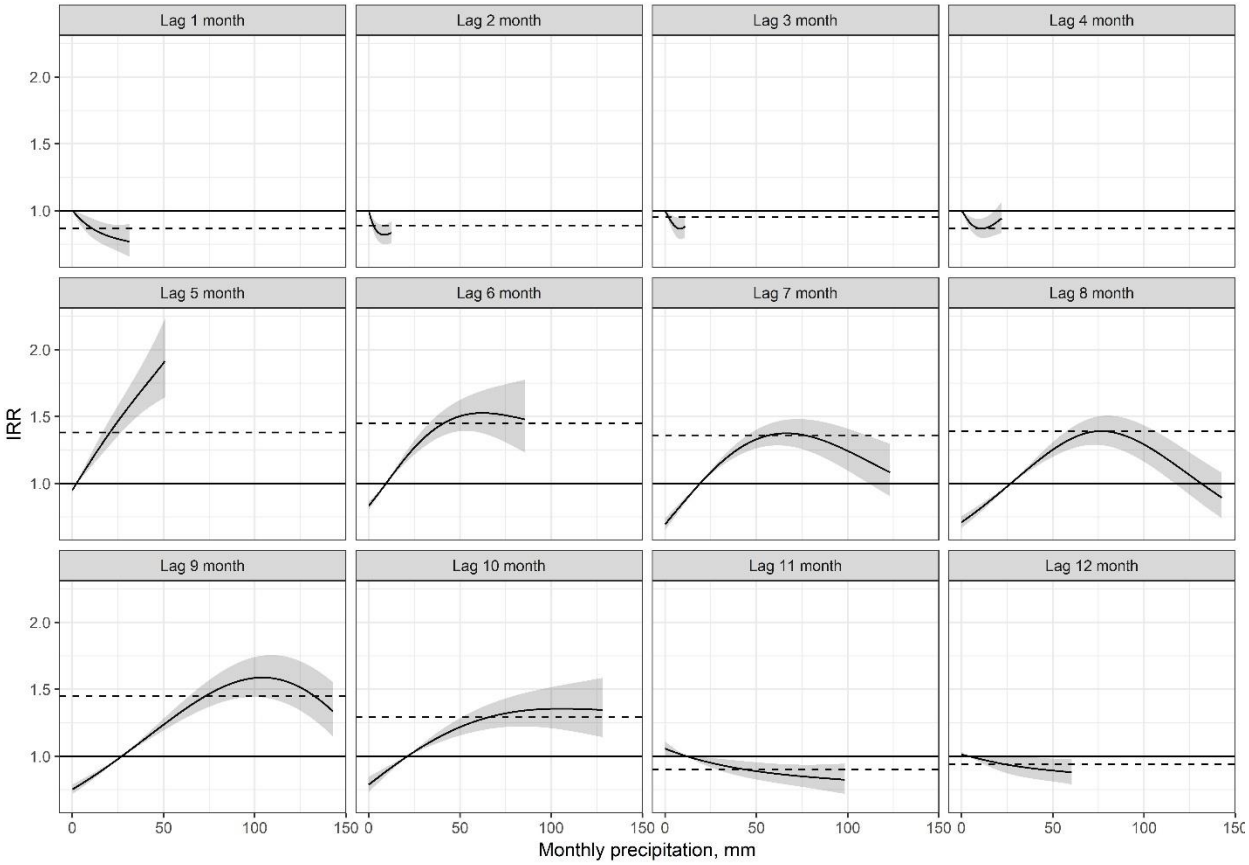


Figure S1. Each line represents a specific county, colored by their median annual precipitation. The lines depict where the n^{th} percentile of monthly total precipitation at a given lag falls along the pooled exposure-response relationship reflecting the relationship between lagged precipitation and coccidioidomycosis incidence. IRRs are scaled such that the reference is the 25^{th} percentile of precipitation at a given lag for the full study region.

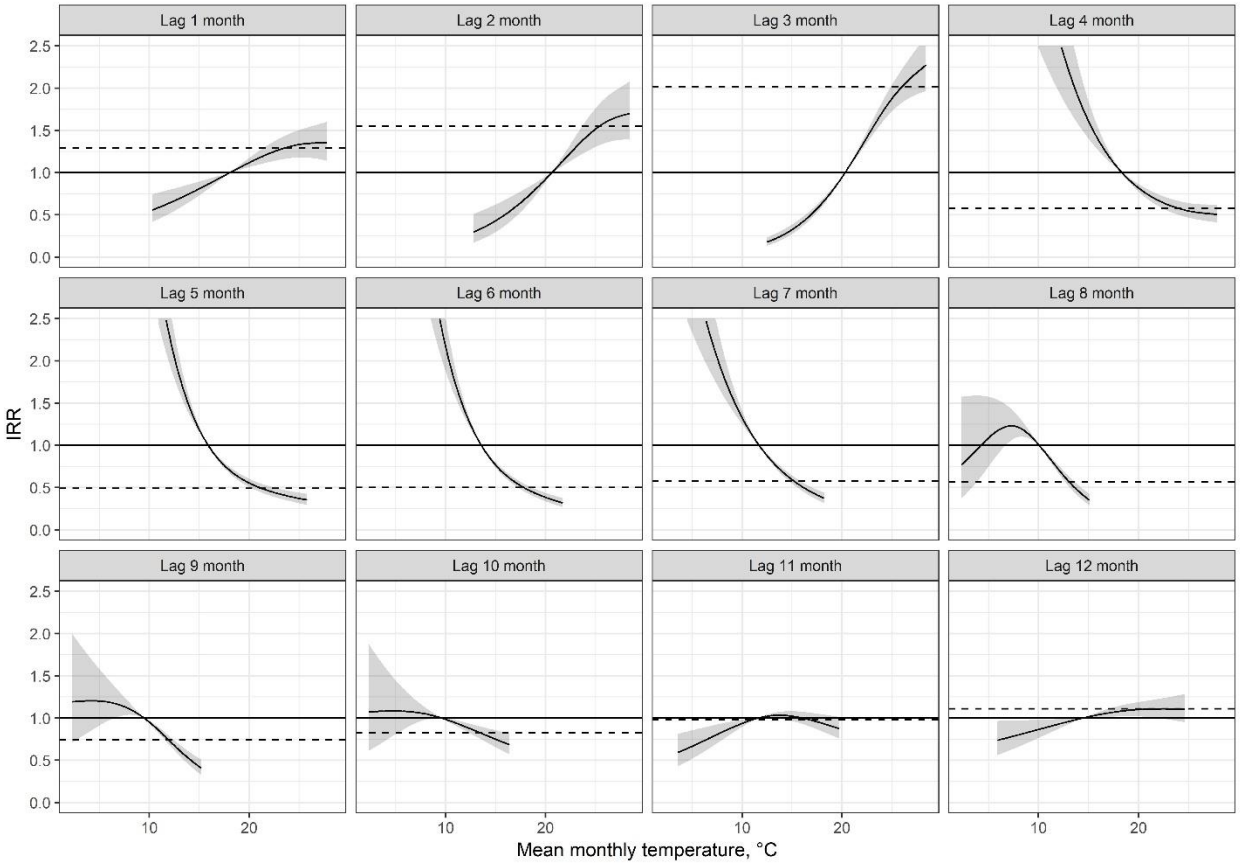


Figure S2. Each line represents a specific county, colored by their mean monthly temperatures. The lines depict where the n^{th} percentile of monthly temperature at a given lag falls along the pooled exposure-response relationship reflecting the relationship between lagged temperature and coccidioidomycosis incidence. IRRs are scaled such that the reference is the 25^{th} percentile of temperature at a given lag for the full study region.

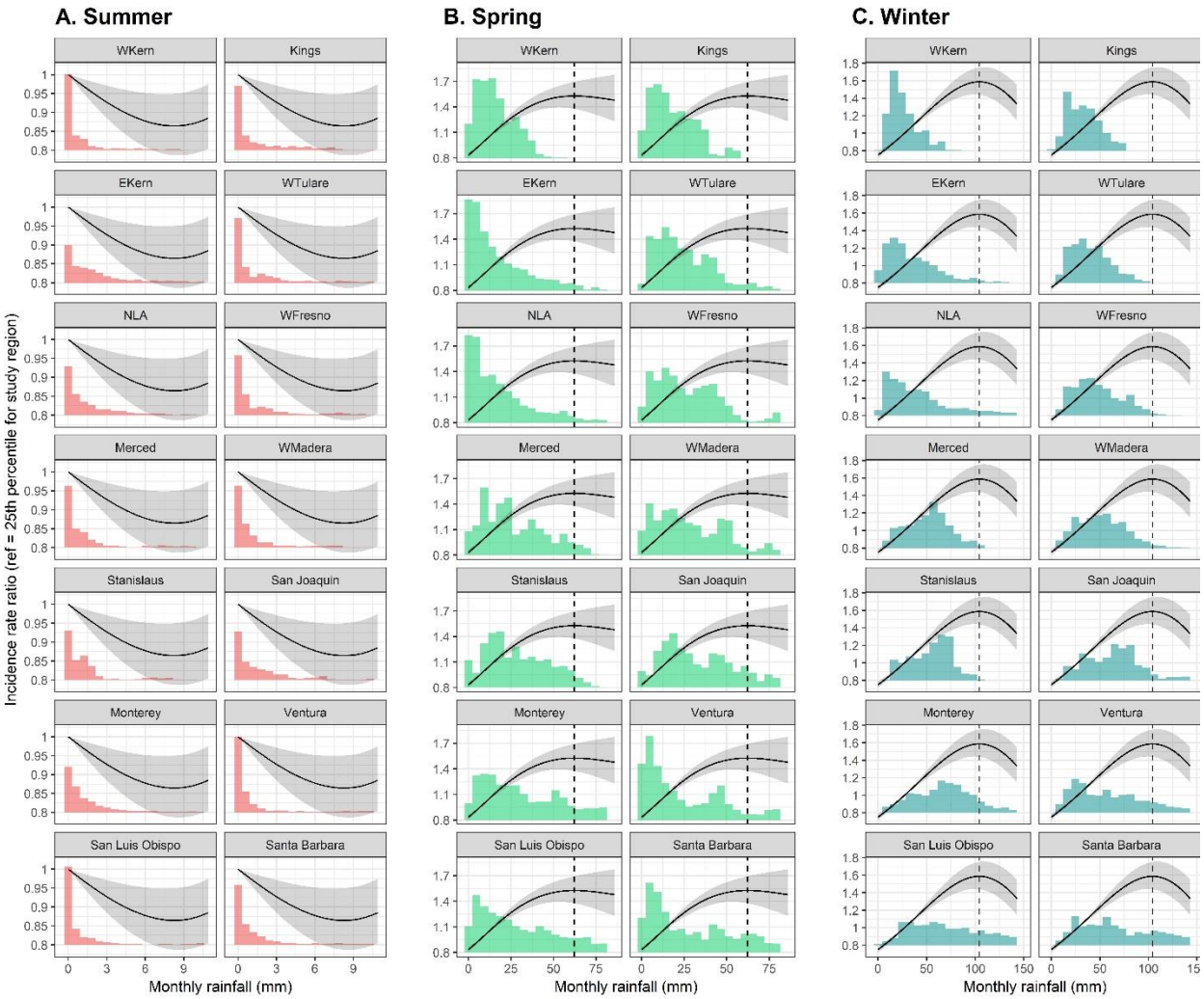


Figure S3. Pooled nonlinear relationship between change in temperature and precipitation at a 3, 6, and 9-month lags, corresponding to summer, spring, and winter, respectively. The IRR is centered at the 25th percentile for the study region, such that the black curve indicates the incidence rate for a given temperature or precipitation value compared to the incidence rate at the 25th percentile for the study region in that time period. Shaded gray regions reflect the 95% confidence interval. Precipitation corresponding to the maximum is indicated by the vertical dotted line. Histograms reflect the density of precipitation experienced by each county/sub-county in the study region.

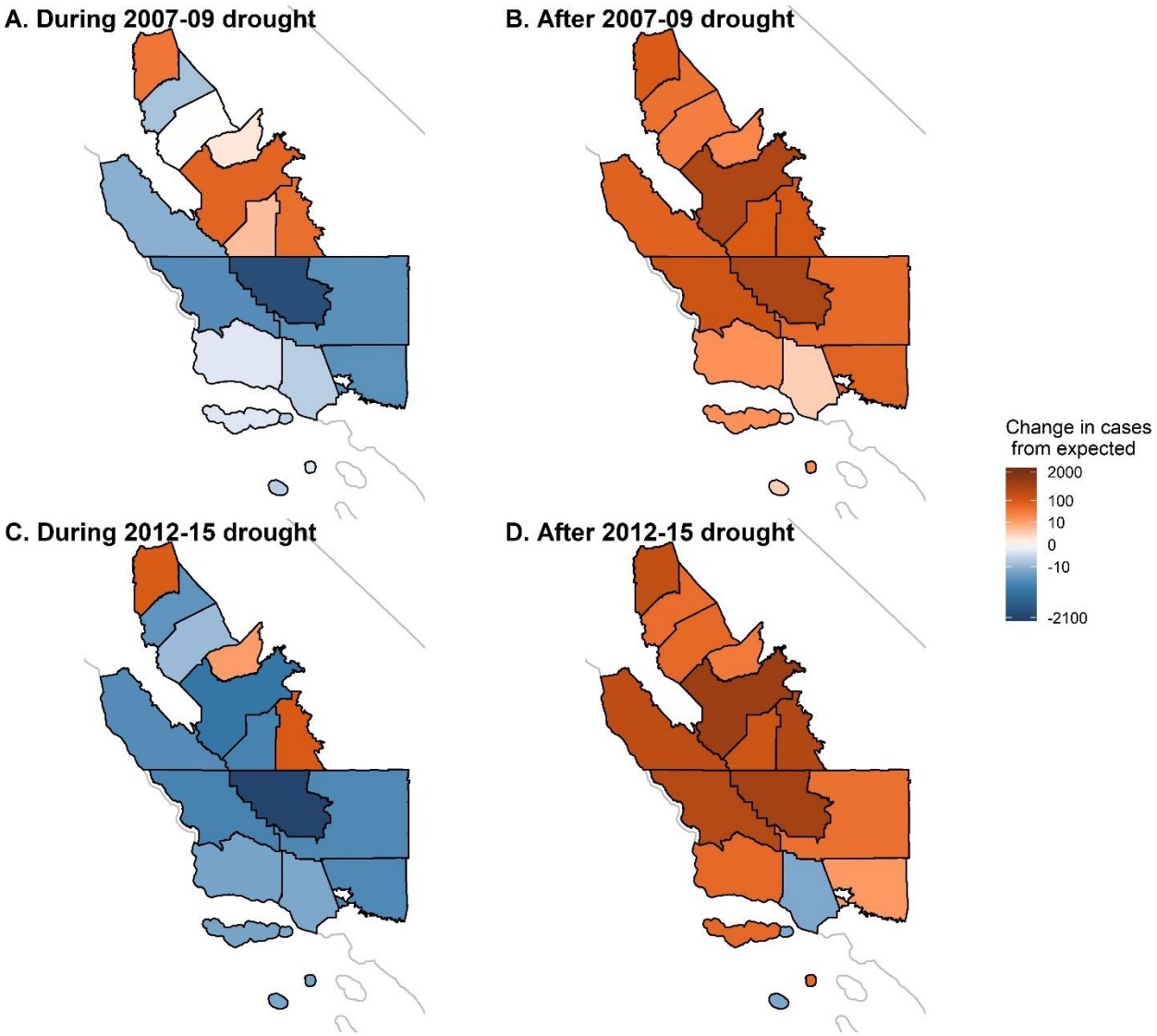
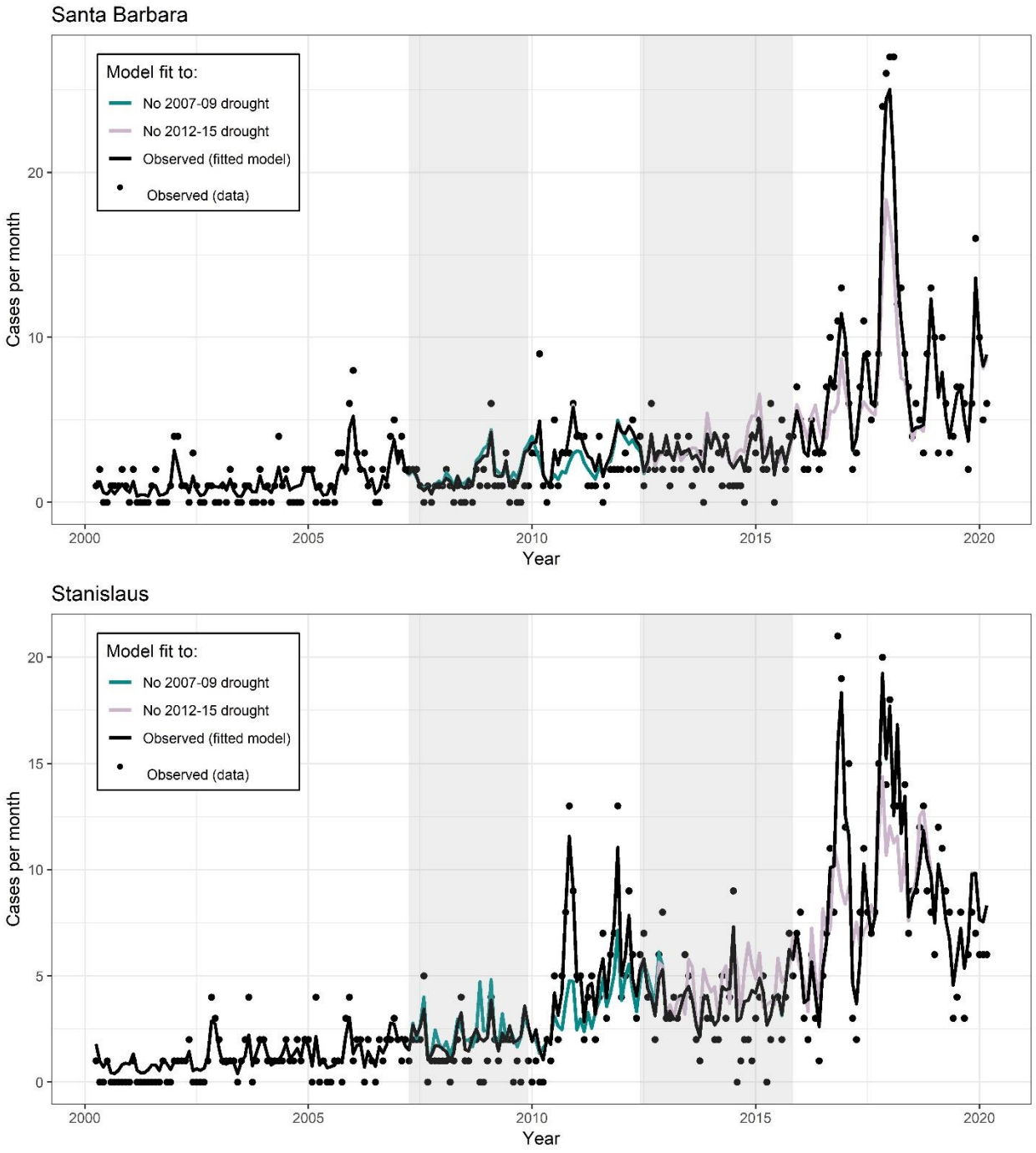
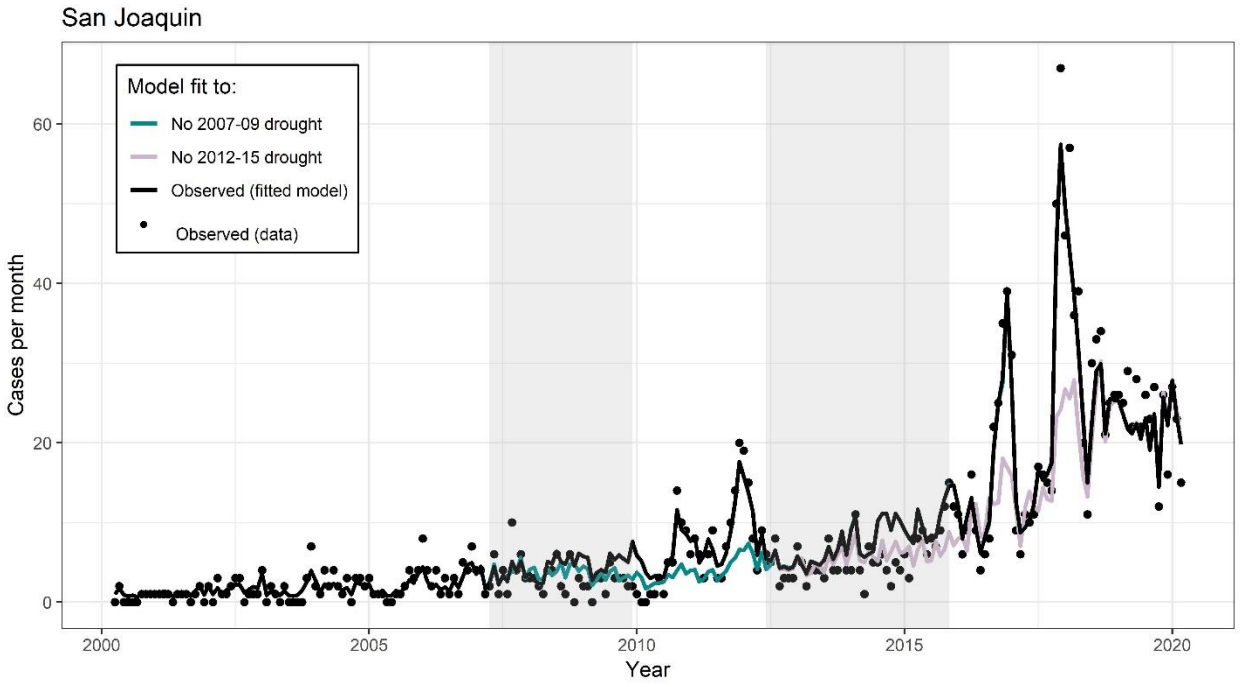
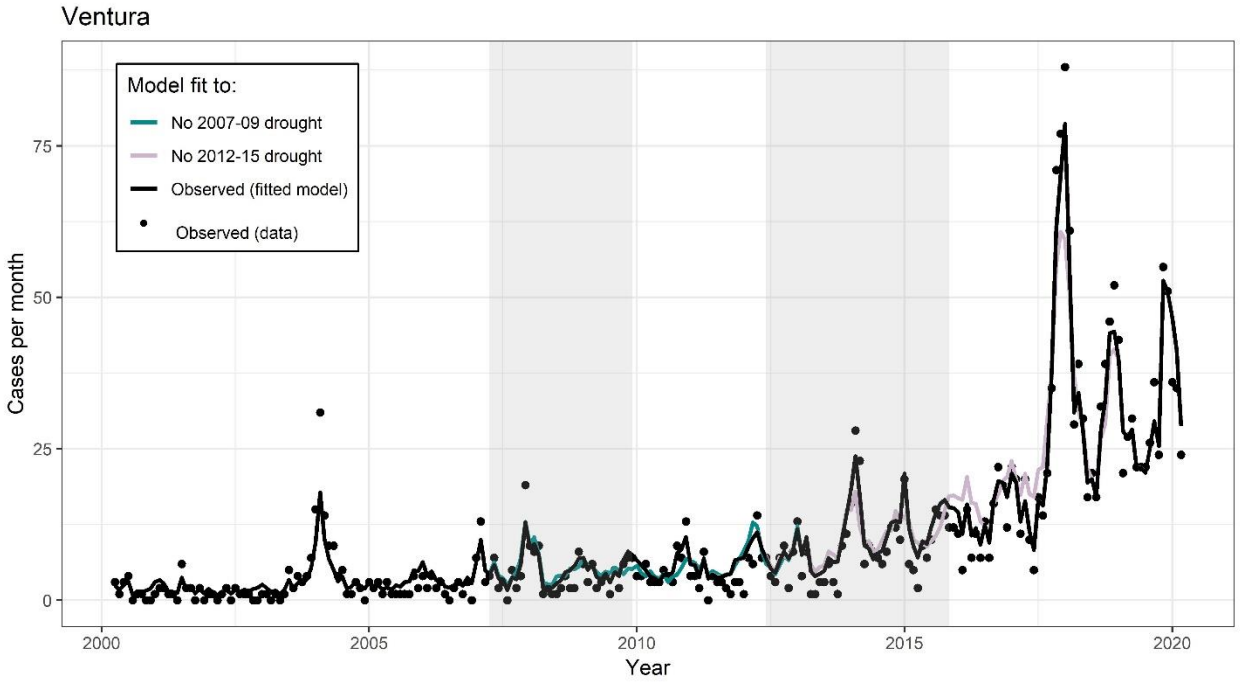


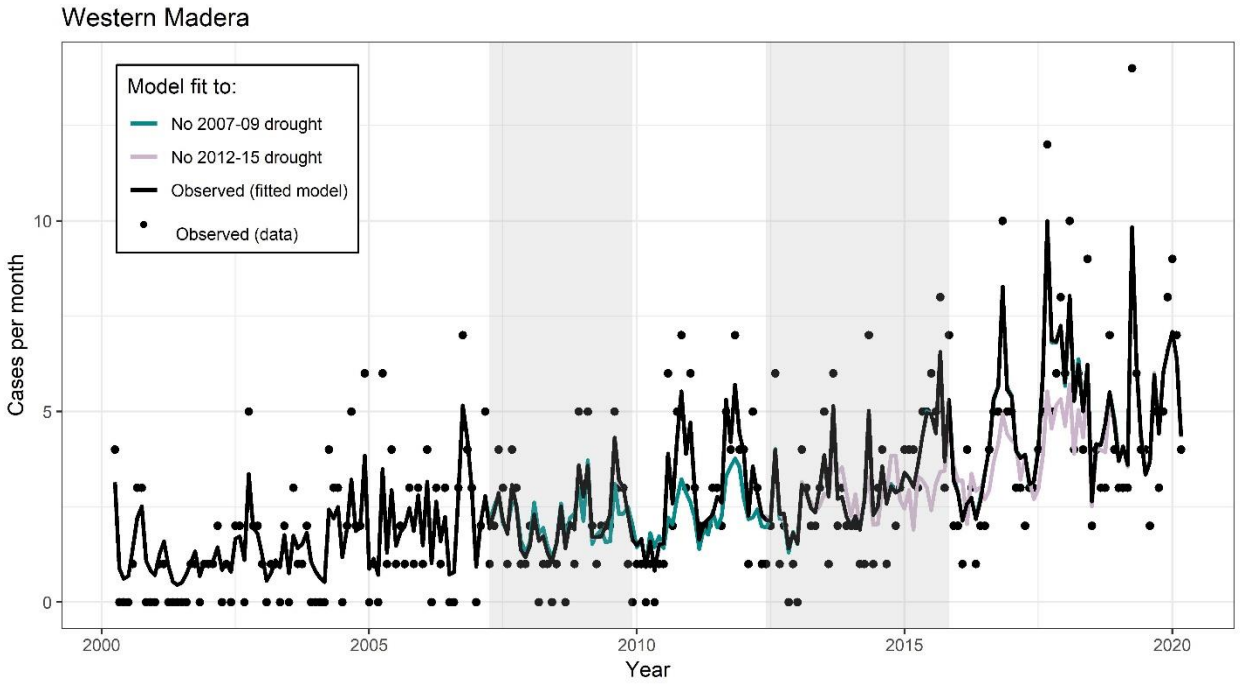
Figure S4. Estimated absolute excess (green) and averted (red) incident cases compared to the number expected in the absence of drought during (A and C) and in the two years following (B and D) the 2007-09 (A, B) and 2012-15 (C, D) droughts across the 14 counties in the study region. California state outline shown in light gray.

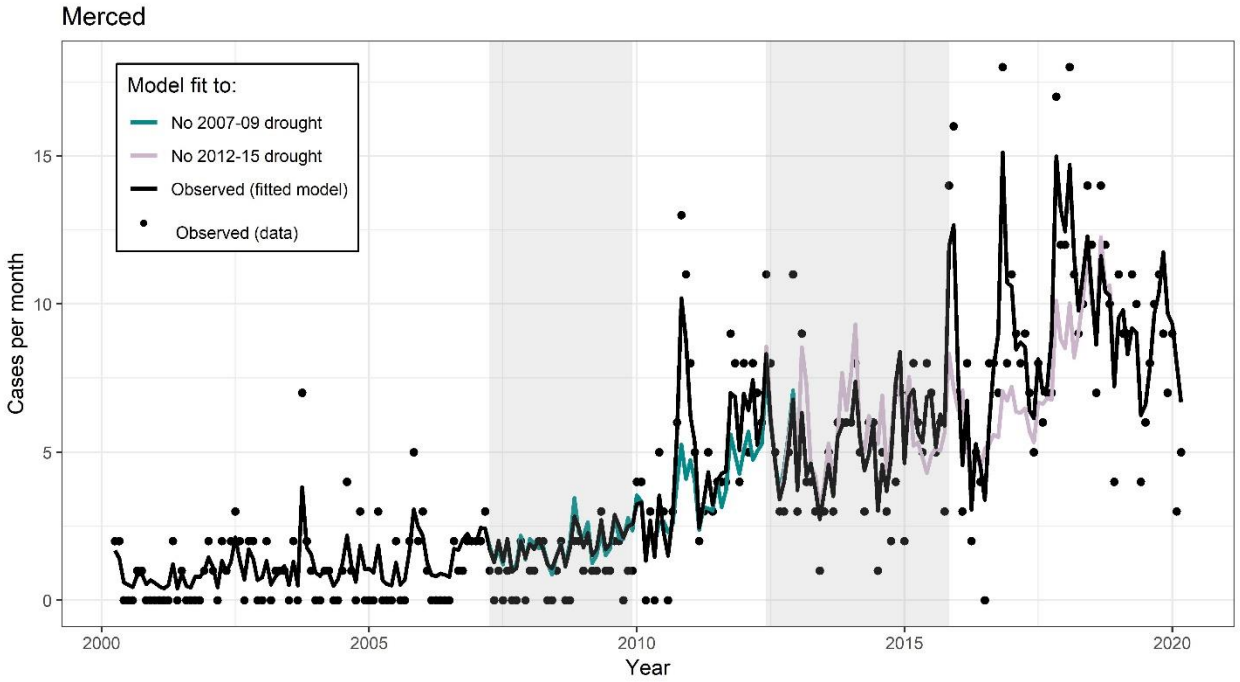
Figure S5-18. Observed incidence (black dots) by month for each county in the study region. Black line is the model fit under the observed environmental conditions. Color lines represent the expected incidence under the counterfactual intervention if the 2007-09 drought did not occur (cyan) or the 2012-15 drought did not occur (pink). Counterfactual scenarios were generated by setting temperatures observed to be higher than historical averages, and precipitation values observed to be below historical averages, deterministically to their average values.

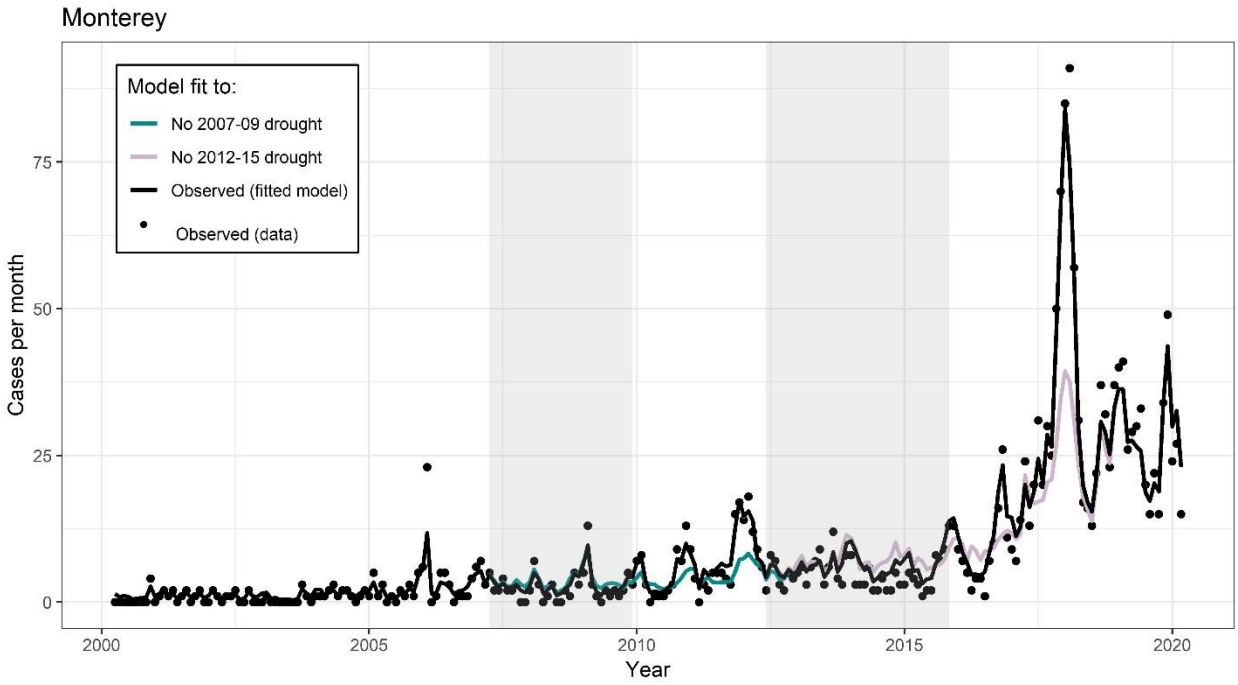


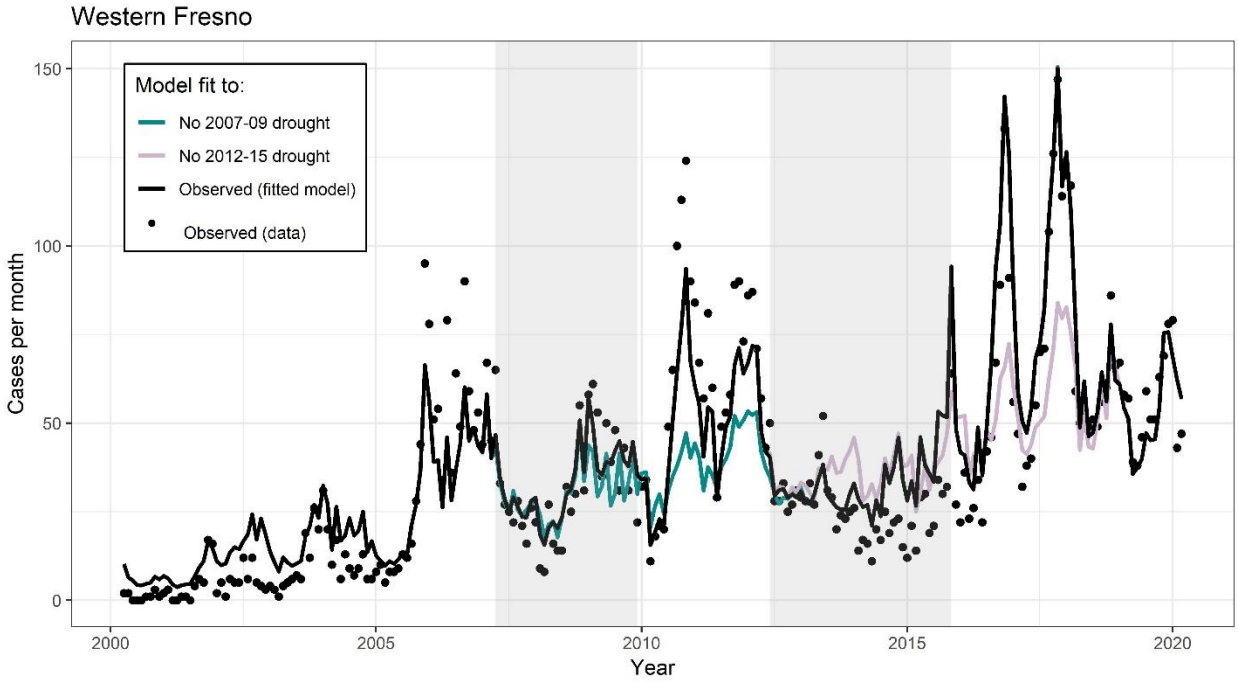

San Joaquin
Model fit to:
No 2007-09 drought
No 2012-15 drought
Observed (fitted model)
Observed (data)
Cases per month
0
20
40
60
2000
2005
2010
2015
2020
Year

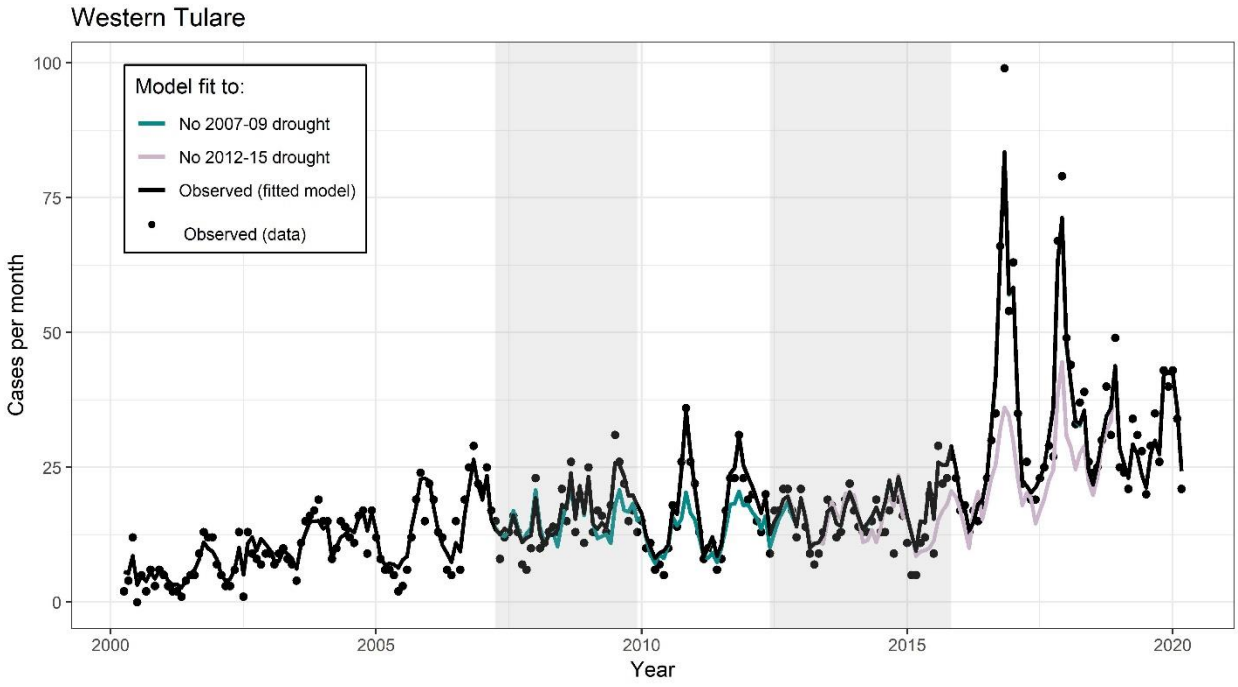

Ventura
Model fit to:
No 2007-09 drought
No 2012-15 drought
Observed (fitted model)
Observed (data)
Cases per month
0
25
50
75
2000
2005
2010
2015
2020
Year

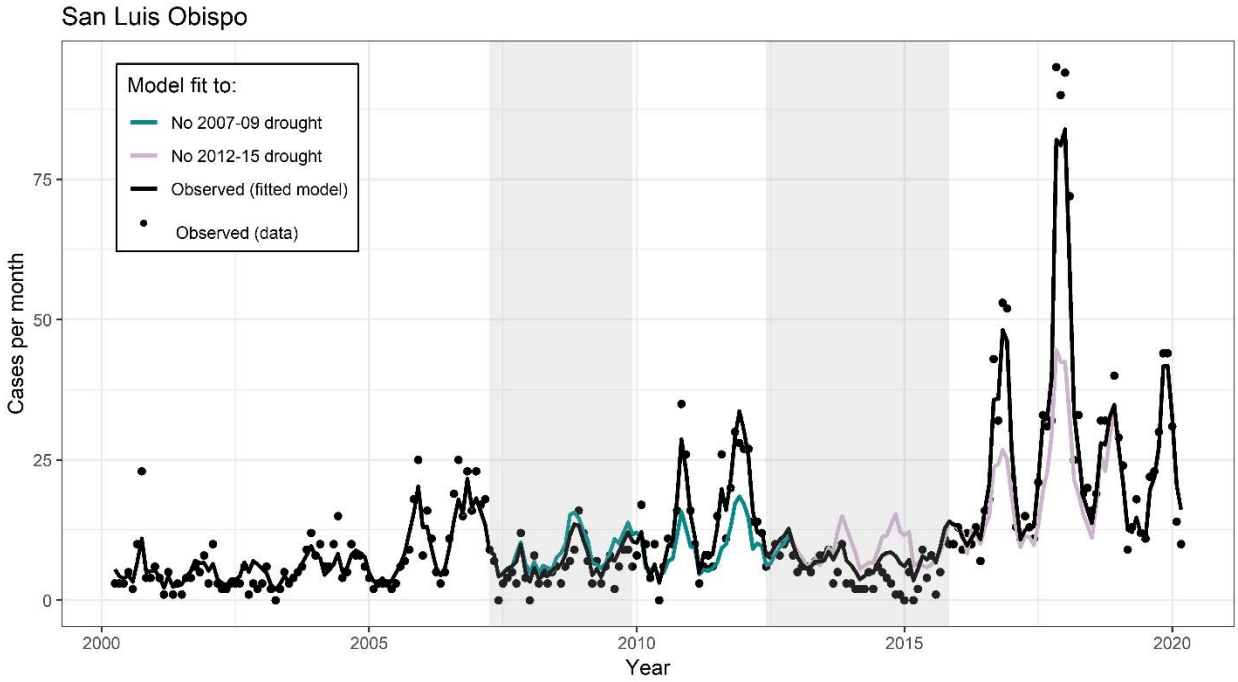

Western Madera
Model fit to:
No 2007-09 drought
No 2012-15 drought
Observed (fitted model)
Observed (data)
Cases per month
0
5
10
2000
2005
2010
2015
2020
Year

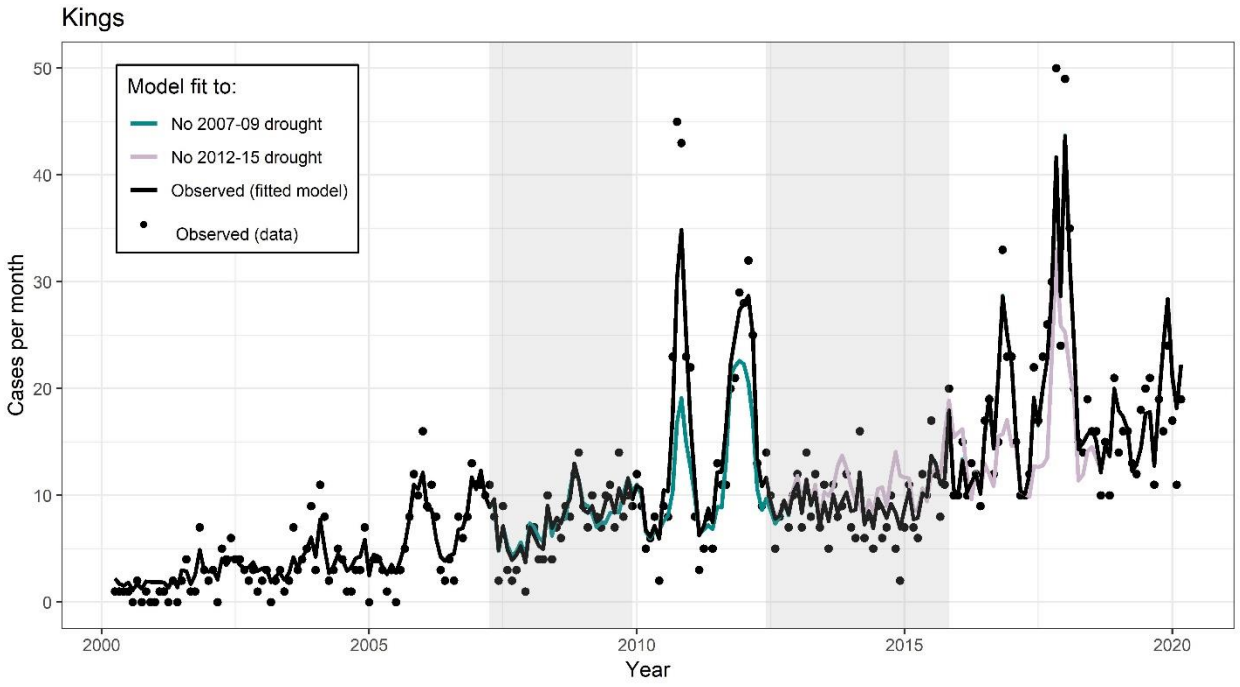

Merced
Model fit to:
No 2007-09 drought
No 2012-15 drought
Observed (fitted model)
Observed (data)
Cases per month
0
5
10
15
2000
2005
2010
2015
2020
Year

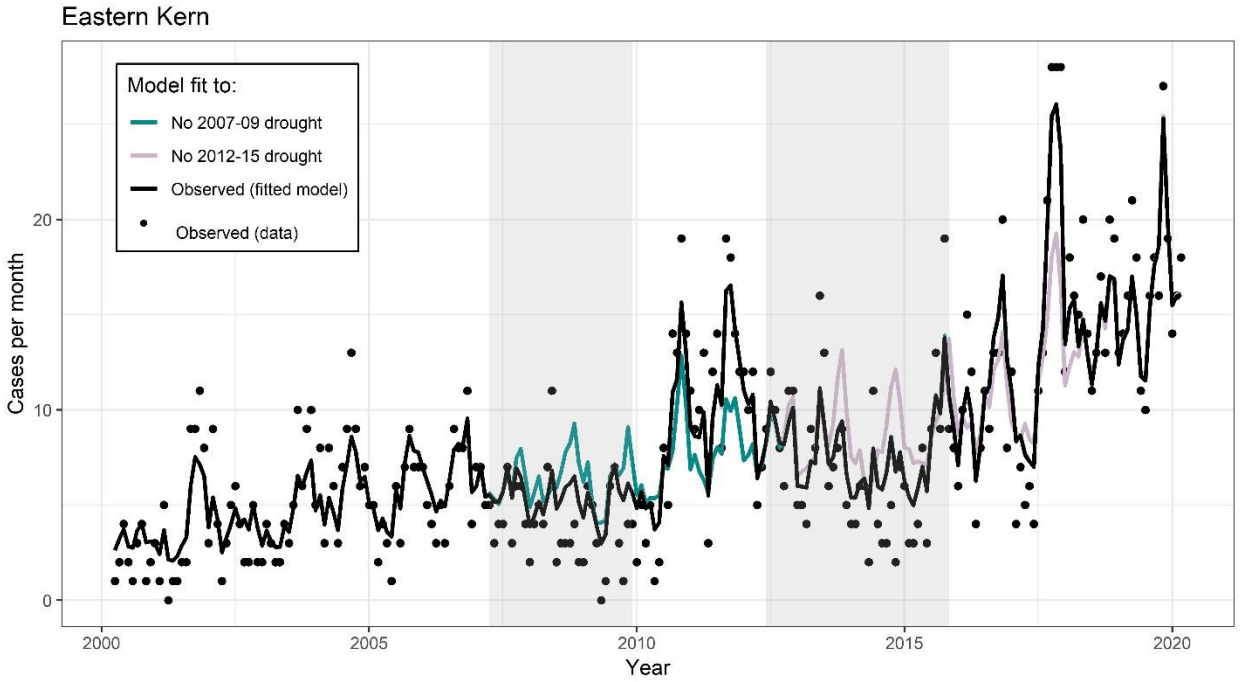

Monterey
Model fit to:
No 2007-09 drought
No 2012-15 drought
Observed (fitted model)
Observed (data)
Cases per month
0
25
50
75
2000
2005
2010
2015
2020
Year

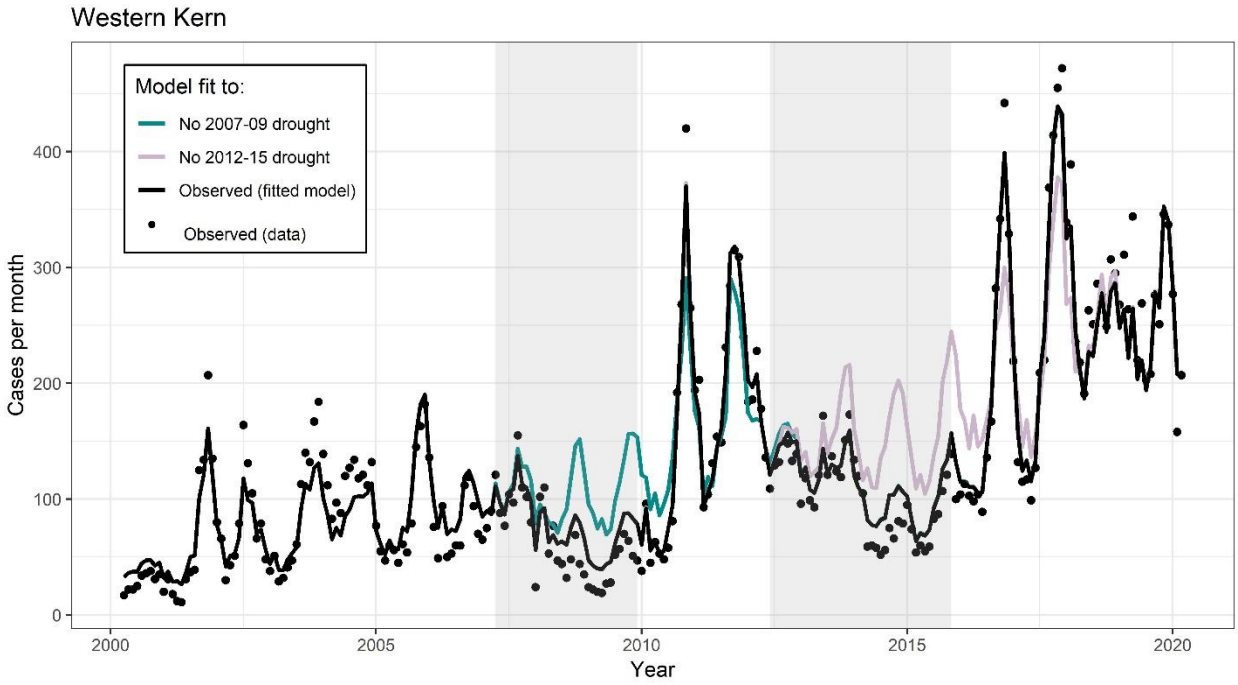

Western Fresno
Model fit to:
No 2007-09 drought
No 2012-15 drought
Observed (fitted model)
Observed (data)
Cases per month
0
50
100
150
2000
2005
2010
2015
2020
Year

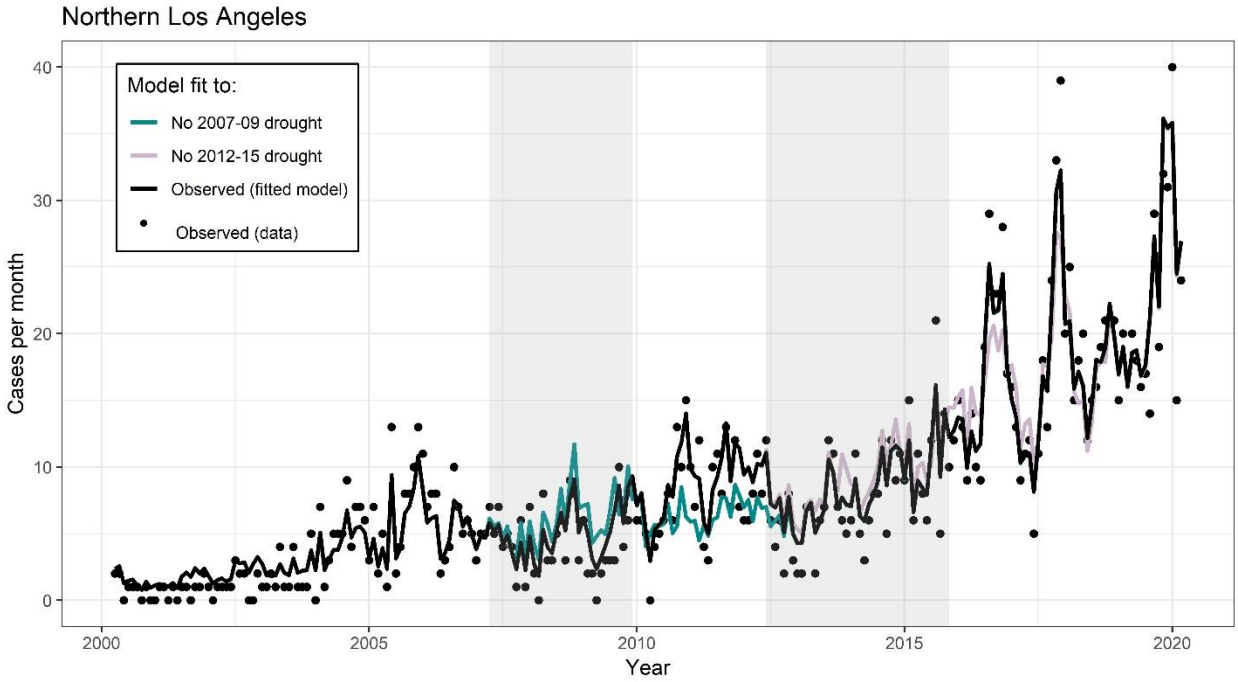

Western Tulare
Model fit to:
No 2007-09 drought
No 2012-15 drought
Observed (fitted model)
Observed (data)
Cases per month
Year
0
25
50
75
100
2000
2005
2010
2015
2020


San Luis Obispo
Model fit to:
No 2007-09 drought
No 2012-15 drought
Observed (fitted model)
Observed (data)
Cases per month
Year
0
25
50
75
2000
2005
2010
2015
2020


Kings
Model fit to:
No 2007-09 drought
No 2012-15 drought
Observed (fitted model)
Observed (data)
Cases per month
Year
0
10
20
30
40
50
2000
2005
2010
2015
2020


Eastern Kern
Model fit to:
No 2007-09 drought
No 2012-15 drought
Observed (fitted model)
Observed (data)
Cases per month
Year
0
10
20
2000
2005
2010
2015
2020


Western Kern
Model fit to:
No 2007-09 drought
No 2012-15 drought
Observed (fitted model)
Observed (data)
Cases per month
0
100
200
300
400
2000
2005
2010
2015
2020
Year


Northern Los Angeles
Model fit to:
No 2007-09 drought
No 2012-15 drought
Observed (fitted model)
Observed (data)
Cases per month
0
10
20
30
40
2000
2005
2010
2015
2020
Year

Supplemental Information for Chapter 3

Notice of approval for human research

Berkeley
UNIVERSITY OF CALIFORNIA
Committee for Protection of Human Subjects (CPHS)
Office for Protection of Human Subjects (OPHS)

1608 Fourth Street, Suite 220
Berkeley, CA 94710-5940
510 642-7461
ophs@berkeley.edu
cphs.berkeley.edu
FWA# 00006252

NOTICE OF APPROVAL FOR HUMAN RESEARCH

DATE:	*April 10, 2020*
TO:	*Justin Remais, SPH Administration* Jennifer Head, SPH Administration, Kristin Andrejko, SPH Administration
CPHS PROTOCOL NUMBER:	*2020-04-13180*
CPHS PROTOCOL TITLE:	*Role of children's social networks in shaping impact of school closures on COVID-19 transmission in the San Francisco Bay Area*
FUNDING SOURCE(S):	*NONE*

A(n) ***new*** application was submitted for the above-referenced protocol. Your submission has been reviewed by the Office for Protection of Human Subjects (OPHS) and granted exemption, as it satisfies the Federal and/or UC Berkeley requirements under category(ies) ***2*** .

Effective Date: *April 10, 2020*

Expiration Date: *April 09, 2030*

Amendments/Modifications: Any change in the design, conduct, or key personnel of this research must be approved by the OPHS ***prior*** to implementation. For more information, see Amend/Modify an Approved Protocol.

Please note that although your research has been deemed exempt from full committee and subcommittee review, you still have a responsibility to protect your subjects, and the research should be conducted in accordance with the principles of the Belmont Report. Download the Belmont Report at this link: www.hhs.gov/ohrp/humansubjects/guidance/belmont.html.

This approval is issued under University of California, Berkeley Federalwide Assurance #00006252.

Note: Exempt determinations are good for ten years. If the study continues beyond the ten-year period, the protocol must be cloned and re-submitted for an updated exempt determination.

If you have any questions about this matter, please contact the OPHS staff at 642-7461 or email ophs@berkeley.edu .

Sincerely,

Office for Protection of Human Subjects (OPHS)
UC Berkeley

Community Contact Survey

Section 0: Consent

DESCRIPTION: You are invited to participate in a research study to understand the effect of school closures in your community. Our team is assisting public health agencies to develop computational models that understand how school closures have affected the spread of COVID-19 in your community. These models will be useful in knowing when to re-open schools and when to close schools under future outbreaks. These models depend on knowing the contact patterns of children and their families following school closures. You will be asked to fill out a form describing the number of people and their ages that you have been within 6 feet of yesterday. We ask families with children in pre-school through 12th grade to also fill out information about their children's contacts.

PROJECT TEAM: We are a team of epidemiologists, mathematicians, and engineers at UC Berkeley School of Public Health who are assisting public health professionals in their COVID-19 planning and response efforts.

TIME COMMITMENT: Your participation will take approximately 5-10 minutes to provide information about your own contact history, and about 5 additional minutes per child to provide information about the contact history of your children.

RISKS AND BENEFITS: We foresee no risks associated with this study. The benefits which may reasonably be expected to result from this study are better epidemiological models that lead to more informed school closure policies. We cannot and do not guarantee that you will receive any benefits from this study.

PAYMENTS: This is a volunteer effort; no payments are involved. Thank you for your time!

PARTICIPANT'S RIGHTS: Your participation is voluntary and you have the right to withdraw your consent or discontinue participation at any time. The alternative is not to participate. Responses are confidential and anonymous. We do not collect personally identifying information and thus cannot identify you from your responses in the data. You have the right to refuse to answer particular questions. The results of this research study may be presented at scientific or professional meetings or published in scientific journals. Your individual privacy will be maintained in all published and written data resulting from the study.

CONTACT INFORMATION:
Questions: If you have any questions, concerns or complaints about this research, its procedures, risks and benefits, contact the UC Berkeley PI contact, Justin Remais: jvr@berkeley.edu

Independent Contact: If you are not satisfied with how this study is being conducted, or if you have any concerns, complaints, or general questions about the research or your rights as a participant, please contact the UC Berkeley Office for Protection of Human Subjects to speak to someone independent of the research team at (510)-642-7461, or by email ophs@berkeley.edu. The study was approved by UC Berkeley's Institutional Review Board with protocol ID 2020-04-13180.

Having read the information above, please select one of the two options below:

[] I CONSENT to take the survey
[] I DO NOT CONSENT to take the survey

Section 1. Screening

1. Do you have at least one child grade preK-12 in your household?
    - Yes
    - No

Section 2: Demographics

1. Choose one or more races that you identify as:
    - White
    - Black or African American
    - American Indian or Alaska Native

- Asian
- Native Hawaiian or Pacific Islander
- Prefer not to say
- Other

2. Do you identify as Hispanic, Latinx, or Spanish origin?
   - Yes
   - No

3. Information about income is very important to understand. Would you please give your best guess? Please indicate the answer that includes your entire household income between January 1, 2019 and December 31, 2019 before taxes.
   - Less than $19,999
   - $20,000 to $39,999
   - $40,000 to $59,999
   - $60,000 to $79,999
   - $80,000 to $99,999
   - $100,000 to $149,999
   - $150,000 or more
   - Prefer not to say

4. What is your zip code?
   [WRITE IN]

5. How did you hear about our survey?
   - My child's school
   - Online forum (e.g. Berkeley Parents Network, Nextdoor)
   - Social Media
   - Friend
   - Local public Health Department
   - Other

Section 3: State and County

1. In which state do you currently live?

▼Drop down with US States

2. In which county do you currently live?

▼Drop down with California counties, only display if State = California

3. Where do you live in [XX] County

▼Drop down with PUMS districts in California, only display if State = California

---

Section 4: Household Composition

1. How many people (including yourself) are in your household?

INCLUDE:

- everyone who is living or staying at this address for more than 2 months
- anyone else staying at this address who does not have another place to stay, even if they are at this address for 1 month or less (ex. college student who has returned home due to university/ dorm closure)

DO <u>NOT</u> INCLUDE :

- anyone who is living somewhere else for more than 2 months, such as a college student living away or someone in the Armed Forces on deployment
    - 1
    - 2
    - 3
    - 4
    - 5
    - 6
    - More than 6

1. Please fill out the following information about your household:

	Age (years)
Household member 1 (YOU!)	[WRITE IN]
Household member 2	[WRITE IN]
Household member 3	[WRITE IN]
Household member 4	[WRITE IN]
Household member 5	[WRITE IN]
Household member 6	[WRITE IN]
Household member 7	[WRITE IN]

2. In the past two weeks, have you, or anyone in your household, experienced a fever or dry cough?
    - Yes
    - No
    - Not sure/ prefer not to say

3. **BEFORE** COVID-19 related school closures, how many adults (18 years or older) typically spent the majority of school hours (8am - 3pm) at home?

INCLUDE anyone who typically works from home, is unemployed, or retired.

[WRITE IN NUMBER]

4. **AFTER** COVID-19 related school closures, how many adults (18 years or older) typically spend the majority of school hours (8am - 3pm) at home?

[WRITE IN NUMBER]

<u>Section 5: Adult Contact Diary</u>

The following questions will ask about yesterday [INSERT DATE]. We know it's hard to remember exactly what happened yesterday, but please give your best guess.

1. Where did you spend the majority of your day yesterday, [INSERT YESTERDAY'S DATE]?

- In my home
- At my place of work (if your place of work is your home during shelter in place, select 'In my home')
- At someone else's home who does not run a commercial daycare
- At a commercial daycare location
- At an outdoor leisure location

- Performing essential activities, such as grocery shopping, laundering clothes, or receiving health care

2. Think about people **that you do not live with** that you were **within 6 feet of** for **more than 5 seconds** yesterday ([YESTERDAY'S DATE]). How many of these people were infants, toddlers, or pre-school aged children (0-4 years)

[WRITE IN]

2a. [IF 2 > 0] In the boxes below, write the number of infants, toddlers, or pre-school aged children (0-4 years) that you were within 6 feet of for more than 5 seconds at each location

	Number of infants, toddlers, or pre-school aged children (0-4 years)
In my home	[WRITE IN]
At my place of work	[WRITE IN]
At someone else's home	[WRITE IN]
At a childcare center that has remained open during Shelter in Place	[WRITE IN]
At an outdoor leisure location	[WRITE IN]
Performing essential activities, such as grocery shopping, laundering clothes, or receiving health care	[WRITE IN]
Riding or waiting for public transit	[WRITE IN]
My child's school	[WRITE IN]
Other	[WRITE IN]

3. Think about people **that you do not live with** that you were **within 6 feet of** for **more than 5 seconds** yesterday ([YESTERDAY'S DATE]). How many of these people were young children (5-12 years)

[WRITE IN]

3a. [IF 3 > 0] In the boxes below, write the number of young children (5-12 years) that you were within 6 feet of for more than 5 seconds at each location

	Number of young children (5-12 years)
In my home	[WRITE IN]
At my place of work	[WRITE IN]
At someone else's home	[WRITE IN]
At a childcare center that has remained open during Shelter in Place	[WRITE IN]
At an outdoor leisure location	[WRITE IN]

Performing essential activities, such as grocery shopping, laundering clothes, or receiving health care	[WRITE IN]
Riding or waiting for public transit	[WRITE IN]
My child's school	[WRITE IN]
Other	[WRITE IN]

4. Think about people **that you do not live with** that you were **within 6 feet of** for **more than 5 seconds** yesterday ([YESTERDAY'S DATE]). How many of these people were teenagers (13-17 years)

[WRITE IN]

4a. [IF 4 > 0] In the boxes below, write the number of teenagers (13-17 years) that you were within 6 feet of for more than 5 seconds at each location

	Number of teenagers (13-17 years)
In my home	[WRITE IN]
At my place of work	[WRITE IN]
At someone else's home	[WRITE IN]
At a childcare center that has remained open during Shelter in Place	[WRITE IN]
At an outdoor leisure location	[WRITE IN]
Performing essential activities, such as grocery shopping, laundering clothes, or receiving health care	[WRITE IN]
Riding or waiting for public transit	[WRITE IN]
My child's school	[WRITE IN]
Other	[WRITE IN]

5. Think about people **that you do not live with** that you were **within 6 feet of** for **more than 5 seconds** yesterday ([YESTERDAY'S DATE]). How many of these people were young adults (18-39 years)

[WRITE IN]

5a. [IF 5 > 0] In the boxes below, write the number of young adults (18-39 years) that you were within 6 feet of for more than 5 seconds at each location

	Number of young adults (18-39 years)
In my home	[WRITE IN]
At my place of work	[WRITE IN]
At someone else's home	[WRITE IN]

At a childcare center that has remained open during Shelter in Place	[WRITE IN]
At an outdoor leisure location	[WRITE IN]
Performing essential activities, such as grocery shopping, laundering clothes, or receiving health care	[WRITE IN]
Riding or waiting for public transit	[WRITE IN]
My child's school	[WRITE IN]
Other	[WRITE IN]

6. Think about people **that you do not live with** that you were **within 6 feet of** for **more than 5 seconds** yesterday ([YESTERDAY'S DATE]). How many of these people were middle aged adults (40-64 years)

[WRITE IN]

6a. [IF 6 > 0] In the boxes below, write the number of middle aged adults (40-64 years) that you were within 6 feet of for more than 5 seconds at each location

	Number of middle aged adults (40-64 years)
In my home	[WRITE IN]
At my place of work	[WRITE IN]
At someone else's home	[WRITE IN]
At a childcare center that has remained open during Shelter in Place	[WRITE IN]
At an outdoor leisure location	[WRITE IN]
Performing essential activities, such as grocery shopping, laundering clothes, or receiving health care	[WRITE IN]
Riding or waiting for public transit	[WRITE IN]
My child's school	[WRITE IN]
Other	[WRITE IN]

7. Think about people **that you do not live with** that you were **within 6 feet of** for **more than 5 seconds** yesterday ([YESTERDAY'S DATE]). How many of these people were older adults (65 +)

[WRITE IN]

7a. [IF 7 > 0] In the boxes below, write the number of older adults (65+ years) that you were within 6 feet of for more than 5 seconds at each location

	Number of older adults (65 +)
In my home	[WRITE IN]

At my place of work	[WRITE IN]
At someone else's home	[WRITE IN]
At a childcare center that has remained open during Shelter in Place	[WRITE IN]
At an outdoor leisure location	[WRITE IN]
Performing essential activities, such as grocery shopping, laundering clothes, or receiving health care	[WRITE IN]
Riding or waiting for public transit	[WRITE IN]
My child's school	[WRITE IN]
Other	[WRITE IN]

8. [IF "At my place of work" is selected for ANY of 2a - 7a]: Where do you work?
    - Office building
    - Grocery store
    - Restaurant
    - Health care facility
    - Various locations, as a delivery driver or postal employee
    - Various locations, as a law enforcement officer
    - Construction site
    - Retail store
    - Public park
    - Gas station or garage
    - Child care/daycare center
    - School or tutoring agency
    - Food processing facility
    - Warehouse or manufacturing facility
    - Other

Section 6: Children Screening Questions

1. We are hoping to get information on all members of the household, especially children in pre-school - 12th grade. Are you willing to help by answering these questions for one or more of your children?

   Only answer YES if someone else in your household has not already filled out a survey for your children.
    - Yes
    - No
    - Someone else in my household has already completed the survey for my children
    - I do not have children in prek-12th grade in my household

[only display the next series of questions about kids if they answer YES above; this series will display for the number of children that they selected above]

2. How many children will you complete the survey for?
    - 1

- 2
- 3
- 4
- 5

3. Do you think school closures have helped reduce the number of covid-19 cases in your community (flatten the curve)?
    - Yes
    - No

4. Do you think school closures are necessary to flatten the curve?
    - Yes
    - No

5. Has your child missed any routine pediatrician appointments during the Shelter in Place order (ex. well-child check ups, yearly physical, routine childhood immunizations), either because you were unable to or unwilling to attend?

    - Yes-I was unable to attend a visit
    - Yes- I was unwilling to attend a visit
    - No- My child has not missed any appointments
    - No- My child has not had any pediatrician visits scheduled, but if they did, I would be willing to attend
    - No- My child has not had any pediatrician visits scheduled, but if they did, I would be unwilling to attend

---

Section 7: Children Contact Diary

Please answer these questions for the [first/second/third/fourth/fifth] of your school aged children.

1. How old is your child (in years?)
[Write in]

2. What type of school does your child attend?
    - Private
    - Public
    - Charter
    - Home-school
    - Other [write in]

3. Where did your child spend the majority of your day yesterday, [INSERT YESTERDAY'S DATE]?
    - In my home
    - At my place of work (if your place of work is your home, select 'In my house')
    - At someone else's home who does not run a commercial daycare
    - At a commercial daycare location
    - At an outdoor leisure location
    - Performing essential activities, such as grocery shopping, laundering clothes, or receiving health care

4. Think about people **that you do not live with** that your child was **within 6 feet of** for **more than 5 seconds** yesterday ([YESTERDAY'S DATE]). How many of these people were infants, toddlers, or pre-school aged children (0-4 years)

    [WRITE IN]

a. **[IF 4> 0]** In the boxes below, write the number of infants, toddlers, or pre-school aged children (0-4 years) that your child was within 6 feet of for more than 5 seconds at each location

	Number of infants, toddlers, or pre-school aged children (0-4 years)
In my home	[WRITE IN]
At my place of work	[WRITE IN]
At someone else's home	[WRITE IN]
At a childcare center that has remained open during Shelter in Place	[WRITE IN]
At an outdoor leisure location	[WRITE IN]
Performing essential activities, such as grocery shopping, laundering clothes, or receiving health care	[WRITE IN]
Riding or waiting for public transit	[WRITE IN]
My child's school	[WRITE IN]
Other	[WRITE IN]

5. Think about people **that you do not live with** that your child was **within 6 feet of** for **more than 5 seconds** yesterday ([YESTERDAY'S DATE]). How many of these people were young children (5-12 years)

[WRITE IN]

b. **[IF 5> 0]** In the boxes below, write the number of young children (5-12 years) that that your child was within 6 feet of for more than 5 seconds at each location

	Number of young children (5-12 years)
In my home	[WRITE IN]
At my place of work	[WRITE IN]
At someone else's home	[WRITE IN]
At a childcare center that has remained open during Shelter in Place	[WRITE IN]
At an outdoor leisure location	[WRITE IN]
Performing essential activities, such as grocery shopping, laundering clothes, or receiving health care	[WRITE IN]
Riding or waiting for public transit	[WRITE IN]
My child's school	[WRITE IN]
Other	[WRITE IN]

6. Think about people **that you do not live with** that your child was **within 6 feet of** for **more than 5 seconds** yesterday ([YESTERDAY'S DATE]). How many of these people were teenagers (13-17 years)

[WRITE IN]

6a. [IF 6 > 0] In the boxes below, write the number of teenagers (13-17 years) that that your child was within 6 feet of for more than 5 seconds at each location

	Number of teenagers (13-17 years)
In my home	[WRITE IN]
At my place of work	[WRITE IN]
At someone else's home	[WRITE IN]
At a childcare center that has remained open during Shelter in Place	[WRITE IN]
At an outdoor leisure location	[WRITE IN]
Performing essential activities, such as grocery shopping, laundering clothes, or receiving health care	[WRITE IN]
Riding or waiting for public transit	[WRITE IN]
My child's school	[WRITE IN]
Other	[WRITE IN]

7. Think about people **that you do not live with** that your child was **within 6 feet of** for **more than 5 seconds** yesterday ([YESTERDAY'S DATE]). How many of these people were young adults (18-39 years)

[WRITE IN]

7a. **[IF 57> 0]** In the boxes below, write the number of young adults (18-39 years) that your child was within 6 feet of for more than 5 seconds at each location

	Number of young adults (18-39 years)
In my home	[WRITE IN]
At my place of work	[WRITE IN]
At someone else's home	[WRITE IN]
At a childcare center that has remained open during Shelter in Place	[WRITE IN]
At an outdoor leisure location	[WRITE IN]
Performing essential activities, such as grocery shopping, laundering clothes, or receiving health care	[WRITE IN]
Riding or waiting for public transit	[WRITE IN]
My child's school	[WRITE IN]

Other	[WRITE IN]

8. Think about people **that you do not live with** that your child was **within 6 feet of** for **more than 5 seconds** yesterday ([YESTERDAY'S DATE]). How many of these people were middle aged adults (40-64 years)

[WRITE IN]

8a. [IF 8 > 0] In the boxes below, write the number of middle aged adults (40-64 years) that your child was within 6 feet of for more than 5 seconds at each location

	Number of middle aged adults (40-64 years)
In my home	[WRITE IN]
At my place of work	[WRITE IN]
At someone else's home	[WRITE IN]
At a childcare center that has remained open during Shelter in Place	[WRITE IN]
At an outdoor leisure location	[WRITE IN]
Performing essential activities, such as grocery shopping, laundering clothes, or receiving health care	[WRITE IN]
Riding or waiting for public transit	[WRITE IN]
My child's school	[WRITE IN]
Other	[WRITE IN]

9) Think about people **that you do not live with** that your child was **within 6 feet of** for **more than 5 seconds** yesterday ([YESTERDAY'S DATE]). How many of these people were older adults (65 +)

[WRITE IN]

9a. [IF 9 > 0] In the boxes below, write the number of older adults (65+ years) that your child was within 6 feet of for more than 5 seconds at each location

	Number of older adults (65 +)
In my home	[WRITE IN]
At my place of work	[WRITE IN]
At someone else's home	[WRITE IN]
At a childcare center that has remained open during Shelter in Place	[WRITE IN]
At an outdoor leisure location	[WRITE IN]

Performing essential activities, such as grocery shopping, laundering clothes, or receiving health care	[WRITE IN]
Riding or waiting for public transit	[WRITE IN]
My child's school	[WRITE IN]
Other	[WRITE IN]

[*Repeat Q's 1-9 in Section 7 depending on how many children said they would answer for]*

_____

Section 8: Thank you message

Thank you! Your response will help schools understand the impact of school closures on COVID-19 transmission in your community!

Do you have another family member who has not taken the survey? Please invite them to participate by sharing the link here: [custom referral link]

Do you know other families who have not taken the survey? Please invite them to participate by sharing the link here: [custom referral link]

FAQ:

What will you do with this data?
Leading epidemiologists are assisting public health agencies to develop computational models that understand how school closures have affected the spread of COVID-19 in your community. These models will be useful in knowing when to re-open schools and when to close schools under future outbreaks. These models depend on knowing the contact patterns of children and their families following school closures. We urgently need volunteers to help us understand the effect of school closures in your community.

Who is behind this project?
We are a team of epidemiologists, mathematicians, and engineers at UC Berkeley School of Public Health who are assisting federal and state officials in their COVID-19 planning and response efforts.

Inverse probability weighting of survey responses

To account for the fact that some respondents did not indicate the locations where they had contact with a given age group, we created a linear mixed model accounting for a random effect at the household level and fixed effects for race and income to model the probability that the individual filled out a location matrix. A binary indicator of whether the individual filled out the location matrix correctly was calculated—the individual was assigned a 0 (indicated incorrect) if the respondent indicated that they have more than zero contacts in a given age category but did not indicate the location where these contacts occurred. We applied a weight defined as the inverse probability of filling out a location matrix correctly when calculating the average number of contacts per location. Weights ranged from 0 to 24 and were not truncated. Figure S2 displays the weighted contact matrices by location.

Construction of synthetic population for transmission model

Household membership and age were drawn from a distribution based on census data on average household size (for households with and without children), proportion of households with children ages <18 years, proportion of single parent households, proportion of multi-generation households, and age of mother at first parity (Table S1). Individuals between 5-18 years old were assigned membership in a school, grade and class, using school district data on school and class sizes. Adults 18-65 years old were assigned membership in a workplace, using census data on employment, with some adults being assigned to schools (staff) and classes (teachers). College students were treated as belonging to non-essential workplaces. We validated the composition of the synthetic population by comparing household age-stratified contact patterns between our synthetic population, the 2018 one-year American Community Survey PUMS from the 9 Bay Area counties, and our household survey (Figure S1). The synthetic population had 16,000 individuals, such that each agent in the synthetic population represented 25 individuals in the real population.

Transmission model details

We developed a discrete-time, age-structured individual-based stochastic model to simulate COVID-19 transmission dynamics in the synthetic population (Figure 1A). At each point in time, representative of one day, each individual is associated with an epidemiological state: susceptible (S), exposed (E), asymptomatic (A), symptomatic with non-severe illness (C), symptomatic with severe illness (H1, D1) resulting in eventual hospitalization before recovery (H2) or hospitalization before death (D2), recovered (R), or dead (M). Model parameters are in Table S5.

The daily contact rate between individuals *i* and *j* on day *t*, $K_{ij,t}$, was estimated for pairs of individuals,

$$K_{ij,t} = \begin{cases} 1 & \text{for household interaction} \\ 5/7 \cdot \rho_{sch}(x_i, x_j, t) \cdot \rho_{int}(x_i, x_j, t) & \text{for class interaction} \\ 1/7 \cdot \rho_{sch}(x_i, x_j, t) \cdot \rho_{int}(x_i, x_j, t) & \text{for grade interaction} \\ 1/35 \cdot \rho_{sch}(x_i, x_j, t) \cdot \rho_{int}(x_i, x_j, t) & \text{for school interacion} \\ 5/7 \cdot \rho_{wrk}(x_i, x_j, t) \cdot \rho_{int}(x_i, x_j, t) & \text{for workplace interaction} \\ K(age_i, age_j)/N(age_j) \cdot \rho_{com}(x_i, x_j, t) & \text{for community interaction} \end{cases}$$

where the scaling ratios between classes, grades, and schools were obtained from previous study on transmission in various settings.[156] Community interaction represents the number of contacts expected between individuals from age groups of individuals *i* and *j* scaled by the number of individuals in the age group of individual *j*. $\rho_{int}(x_i, x_j, t)$ is a factor between 0 and 1 representing a social distancing intervention to reduce contact between individual pairs, and is equal to one under a no-intervention scenario. Because symptomatic individuals mix less with the community[157], we simulated a 100% reduction in daily school or work contacts and a 75% reduction in community contacts for a proportion (48%) of symptomatic individuals, and an additional proportion (50%) of their household members.[158] For these individuals, $\rho_{sch}(x_i, x_j, t)$ and $\rho_{wrk}(x_i, x_j, t)$ is equal to 0 and $\rho_{com}(x_i, x_j, t)$ is equal to 0.25, if: 1) either individual *i* or *j* is symptomatic (C, H1, or D1) on day *t* and isolates with some probability, or 2) either individual *i* or *j* is a household member of a symptomatic individual on day *t* and quarantines with some probability; and otherwise equal to 1. We assumed that individuals were in the infectious class for up to 3 days prior to observing symptoms[159], during which time they did not reduce their daily contacts.

Transmission was implemented probabilistically for contacts between susceptible (S) and infectious individuals in the asymptomatic (A) or symptomatic and non-hospitalized states (C, H1, D1). Movement of individual *i* on day *t* from a susceptible to exposed class is determined by a Bernoulli random draw with probability of infection per day given by the daily force of infection, $\lambda_{i,t}$:

$$\lambda_{i,t} = \alpha\beta_i \sum_{j=1}^{N} K_{ij,t} A_{j,t} + \beta_i \sum_{j=1}^{N} K_{ij,t}(C_{j,t} + H1_{j,t} + D1_{j,t}) \quad (1)$$

where α is the ratio of the force of infection between asymptomatic and symptomatic individuals; and β_i is calculated from $\bar{\beta}$, the population mean transmission rate of the pathogen. $\bar{\beta}$ is determined using the next-generation matrix method[160] as:

$$\bar{\beta} = \frac{R_0}{[d_I(p_C + \alpha p_a) + d_H(p_H + p_D)]\,\bar{K}} \quad (2)$$

where R_0 is the basic reproduction number (defined as the expected number of secondary cases from a single infected case in a completely susceptible population); p_s is the proportion of agents destined for state *s*; d_I is the average time between infection and recovery for tracks A and C; d_H is the average time between infection and hospitalization for tracks H and D; and $\bar{K}$ is

the mean number of contacts an individual makes daily under no interventions, weighted by their probability of being contacted.[239] We represent age-varying susceptibility[163] using an age-stratified β_i that incorporates the ratio of the susceptibility of adults to children and jointly solves equations (3) and (4):

$$\bar{\beta} = \beta_{i \geq age} \frac{n_{i \geq age}}{N} + \beta_{i < age} \frac{n_{i < age}}{N} \quad (3)$$

$$\beta_{i \geq age} = \beta_{i < age} \left(\frac{\text{Susceptibility of adults}}{\text{Susceptibility of children}} \right) \quad (4)$$

Using this method, we calculated the secondary attack rate among household members to be between 9.6% and 11.1%, in agreement with prior studies.[143, 144, 164, 240]

The duration of the latent period, d_L, for each individual transitioning from class E was drawn from a Weibull distribution with mean 5.4 days (95% CI: 2.4, 8.3).[241-243] Whether an individual remained asymptomatic, or was hospitalized, or died was determined via Bernoulli random draws from age-stratified conditional probabilities (Figure 1B, Table S5). The time to recovery for non-hospitalized cases (mean: 13.1 days, 95% CI: 8.3, 16.9)[244], the time to hospitalization for severe cases (mean: 10.3, 95% CI: 6.5, 13.3)[245], and time to recovery or death for hospitalized cases (mean: 14.4, 95% CI: 11.3, 16.6) were sampled from Weibull distributions (Table S5).[246] Simulations were initiated on January 17, two weeks before the first known case in Santa Clara County, assuming a fully susceptible population seeded with a random number (range: 5-10) of exposed individuals[173]. We averaged results over 1,000 independent realizations and estimated confidence intervals as the 2.5th and 97.5th percentile of all realizations.

The disease progression track followed by each individual after movement to the exposed state was assigned from Bernoulli random draws at the start of each simulation. Tracks for unvaccinated individuals were sampled from distributions specified by the age-stratified conditional probabilities, given in Table S1. Tracks for vaccinated individuals were sampled from distributions specified by the age-stratified conditional probabilities given, accounting for differential vaccine effectiveness against symptomatic and severe disease. The probability of death is conditional on hospitalization, so conditioning the probability of hospitalization on vaccination status also conditions the probability of death on vaccination status. Specifically, the probability of success for symptomatic and severe (e.g., requiring hospitalization) disease was updated as follows:

$$P(symptoms \mid age, infection, vaccination) = P(symptoms \mid age, infection) * \frac{\left(1 - VE_{symptomatic}\right)}{\left(1 - VE_{any\ infection}\right)} \quad (6)$$

$$P(hospital \mid age, symptoms, vaccination) = P(hospital \mid age, symptoms) * \frac{\left(1 - VE_{severe}\right)}{\left(1 - VE_{symptomatic}\right)} \quad (7)$$

Description of reopening strategies

1. Schools open without precautions

   In this scenario, schools are open under a business-as-usual scenario. For all interactions, $\rho_{int}(x_i, x_j, t)$ = 1. The average class size is 20 students, the average sizes of elementary (K - 5), middle (6-8), and high schools (9-12) are 383, 414, and 619 students.

2. Stable cohorts: *classroom groups are enforced, reducing other grade and school contacts by 50% (weak) or 75% (strong)*

   In this scenario, we assume that students reduce their contacts with other teachers and students outside of their class group (or cohort) by a given proportion. We model both reductions of outside-class contacts by 50% ("weak" cohort approach) or 75% ("strong" cohort approach). The size of the class group is 20 students, on average. This may be equivalent to reductions in lunchroom or recess contacts, while still permitting chance interactions in the hallways or bathrooms. Here, we update $\rho_{int}(x_i, x_j, t)$ such that for the weak cohort (2a):

$$\rho_{int}(x_i, x_j, t) = \begin{cases} 1 & \text{for class interaction} \\ 0.5 & \text{for grade interaction} \\ 0.5 & \text{for school interaction} \end{cases}$$

and for the strong cohort (2b):

$$\rho_{int}(x_i, x_j, t) = \begin{cases} 1 & \text{for class interaction} \\ 0.25 & \text{for grade interaction} \\ 0.25 & \text{for school interaction} \end{cases}$$

3. Staggered half classes: *Class sizes are cut in half, and each half attends two days a week*

   In this scenario, we assume that classes are halved, to average 10 students each. Half the class attends school two days a week, and the other half attends a different two days a week. Teachers and administrators attend four days a week. We group school, grade, and class interactions by whether or not they are within the same shift, and update $\rho_{int}(x_i, x_j, t)$ accordingly:

$$\rho_{int}(x_i, x_j, t) = \begin{cases} 2/5 & \text{for class interaction within shift} \\ 2/5 & \text{for grade interaction within shift} \\ 2/5 & \text{for school interaction within shift} \\ 0 & \text{for pairs in different shifts} \end{cases}$$

4. Staggered school days: *half the school attends two staggered days a week according to grade groups. Class size is maintained at regular levels*

In this scenario, we assume that grades in a school attend two days a week, and the other half attends a different two days a week. For instance, in elementary schools, grades K-2 attend Mondays and Tuesdays, and grades 3-5 attend Thursdays and Fridays. In middle schools, grades 6-8 attend Mondayes and Tuesdays, and grade 8 attends Thursdays and Fridays. In high schools, grades 9-10 attend Mondays and Tuesdays and grade 11-12 attends Thursdays and Fridays. Teachers only attend the two days in which their classroom is present. School administrators attend all four days a week. We group school, grade, and class interactions by whether or not they are within the same shift, and update $\rho_{int}(x_i, x_j, t)$ accordingly:

$$\rho_{int}(x_i, x_j, t) = \begin{cases} 2/5 & \text{for class interaction} \\ 2/5 & \text{for grade interaction} \\ 2/5 & \text{for school interaction within shift} \\ 0 & \text{for pairs in different shifts} \end{cases}$$

5. Students and faculty wear masks

In this scenario, we assume that both students and teachers wear masks while at school. We assume that the masks both reduce the likelihood of acquiring COVID-19, as well as the likelihood of transmitting it. We assume that the effectiveness of masks for elementary school children is 15%, the effectiveness for middle school children is 25%, the effectiveness for high school children is 35% and the effectiveness for teachers is 50%. Accordingly, for each school, grade, or class pair, we have:

$$\rho_{int}(x_i, x_j, t) = \big(1 - \eta(x_i)\big) \cdot \Big(1 - \eta\big(x_j\big)\Big),$$

where $\eta(x_i)$ represents the effectiveness of the mask for individual *i.* such that $\eta(x_i)$ = 0.15 if the individual is an elementary school student, $\eta(x_i)$ = 0.25 if the individual is a middle school student, $\eta(x_i)$ = 0.35 if the individual is a high school student, and $\eta(x_i)$ = 0.5 if the individual is a teacher or staff member.

6. Monthly/weekly testing of teachers and students (periodic test-trace-isolate, TTI): *Faculty and students are tested with 85% sensitivity on a weekly or monthly basis[42], and positive cases are isolated and their class quarantined for 14 days*

In this scenario, every 7 or 30 days, the state of the non-hospitalized agents are ascertained through a simulated test. We assumed that the test would detect individuals in a symptomatic or asymptomatic or pre-symptomatic state with 85% sensitivity and 100% specificity. If a truly positive case was simulated to test positive, the case would reduce their school contacts by 100% for 14 days and their community contacts by 75% for 14 days. Additionally, the students or teacher in the same class as the case would reduce their school contacts by 100% and their community contacts by 75% for 14 days. This is implemented though updating $\rho_{sch}(x_i, x_j, t)$ and $\rho_{comm}(x_i, x_j, t)$ as described. If a school administrator tested positive, only the administrator isolated for 14 days.

Choice of susceptibility parameters based on available literature

The impact of school closures depends critically on the relative susceptibility and infectiousness of children. What follows is a brief summary of key literature, emphasizing contact tracing studies where possible, as of July 17, 2020. We acknowledge that uncertainty remains in these parameters. While more studies from upper-income countries report a smaller role of children in the transmission of COVID-19 compared to adults, there is likely substantial selection bias owing to the increased likelihood of children to have less severe symptoms and the timing of studies during school closures when children had few non-household contacts. For these reasons we explore scenarios where children are half as susceptible to infection as adults, and scenarios where children are equally as susceptible to children as adults. A review by Goldstien, Lipsitch, and Cevik includes a more thorough discussion of this information [247].

Author	Journal	Country	Type of study	Findings
Bunyavanich, et al. [248]	JAMA	--	Laboratory	SARS-CoV2 uses the ACE2 receptor for host entry and the ACE2 gene expression increases linearly with age. Expression was lowest in younger children (2.40 mean $\log_2$ counts per million), and increased in older children (2.77 mean $\log_2$ counts per million), young adults (3.02 mean $\log_2$ counts per million), and older adults (3.09 mean $\log_2$ counts per million).
Zhang, et al. [131]	Preprint	China	Contact tracing	The odds of secondary attack in children was 0.34 (95% CI: 0.24 - 0.49) times that of adults. Children 0-14 years were 59% (95% CI 7-82%) less susceptible than individuals 65 years and over.
Mizumoto, et al. [249]	Preprint	Japan	Contact tracing	The odds of secondary attack in children was 0.21 (95% CI: 0.11 - 0.41) times that of adults.
Jing, et al.[164]	Preprint	China	Contact tracing	The secondary attack rate in children was 5.3%, compared to 12.6% overall; OR: 0.27 (0.13 - 0.55)
Li, et al. [250]	CID	China	Contact tracing	The secondary attack rate in children was 4%, compared to 20.5% in adults; OR = 0.16 (0.06-0.46)
Posfay-Barbe, et al. [251]	Pediatrics	Switzerland	Contact tracing	Adult household contacts had symptoms prior to or at the same time as the study child in 80% (31/39) of cases. In 8% (3/39) of households did the study child develop symptoms prior to any other household contact. 85% (75/88) of adult household contacts developed symptoms at some point, compared to 43% (10/23) of pediatric household contacts ($p<0.001$).
Danis, et al. [252]	CID	France	Contact tracing	A 9 year old child with co-occurring influenza and COVID 19 visited more than 80 children. Zero other students got COVID-19 but numerous other students had influenza

Wu, et al.[253]	Pediatrics	China	Contact tracing	Of 68 children with confirmed COVID-19 admitted to Qingdao Women's and Children's Hospital from January 20 to February 27, 2020, and with complete epidemiological data, 65 (95.59%) patients were household contacts of previously infected adults.
Dattner, et al.[142]	Preprint	Israel	Contact tracing	25% of children infected over all households, compared to 44% of adults infected over all households, excluding index cases. Using a modelling approach, estimated that the susceptibility of children (under 20 years old) is 45% [40%, 55%] of the susceptibility of adults, and the infectivity of children 85% [65%, 110%] relative to adults.
Bi, et al.[240]	Lancet ID	China	Contact tracing	The household secondary attack rate in children (7.4%) was similar to that for adults (7.9%). Children were similarly susceptible to infection compared to adults, but were less likely to have severe symptoms.
Laxminarayan, et al.[165]	Science	India	Contact tracing	One-third of the 33,584 confirmed cases were <30 years of age in two Indian states. After adjusting for the fact that most contacts involving children occurred in the household setting, there was not strong evidence of differential risk of acquiring or transmitting infection across ages.
Park, et al.[143]	EID	South Korea	Contact tracing	A total of 11.8% (95% CI 11.2%–12.4%) household contacts of index patients had COVID-19; in households with an index patient 10–19 years of age, 18.6% (95% CI 14.0%–24.0%) of contacts had COVID-19; in households with an index patient 0-9 years of age, 5.3% (1.3–13.7) of contacts had COVID-19.
Fatah-Moghadam, et al.[145]	Pre-Print	Italy	Contact tracing	The secondary attack rate in children 0-14 was 8.4%, compared to 9.2% in adults 25-29 and >15% in adults 30 and older. Children were estimated to be the most infectious compared to other age groups.
Fontanet et al.[148]	Pre-Print	France	Retrospective cohort	Study evaluated students, teachers, staff, and family members who attended a high school linked to a cluster of COVID-19 and found an infection attack rate of 40.9% in high school affiliated students, staff, and faculty.
Fontanet et al.[192]	Pre-print	France	Retrospective cohort	Students, teachers, and staff who were exposed to SARS-CoV-2 before school closures did not develop COVID-19 following exposure to three SARS-CoV-2 positive students
Gudbjartsson, et al.[254]	NEJM	Iceland	Screening	Screened 6% of the population for COVID-19 through both targeted and random means. 13.3% and 0.8% of the target and general population

				were positive, respectively, and 6.3% and 0% of the targeted and general population of individuals 10 and younger.
Dong, et al.[255]	Pediatrics	China	Surveillance	2135 pediatric patients with COVID-19 were analyzed. Children of all ages were represented in the data set as being equally susceptible, with 5.3% of cases severe.
Zhu, et al.[256]	Preprint	China, Singapore, US, Vietnam, South Korea	Systematic review of household clusters	In three (9.7%) of 31 published household clusters, the index case was a child compared to 30 (54%) of 56 published household clusters of influenza A (H5N1).
Davies, et al.[163]	Nature Medicine	China, South Korea, China, Italy, Singapore, and Canada	Modelling	SEIR models were fit to surveillance data from several countries. It is estimated that the susceptibility of children is half that of adults over 20 years old.
Jones et al.[257]	Preprint	Germany	Virology	Children and adults shed similar viral loads that likely represent infectivity.
Stoye [258]	Preprint	Germany	Virology	Re-analysis of Jones, et al. suggests that there is an increase in viral load by age, with higher viral load in older ages.
Lennon, et al. [259]	Preprint	USA	Virology	Researchers found similar distributions of viral load in patients with or without symptoms at the time of testing during the local peak of the epidemic; as the epidemic waned, individuals without symptoms at the time of testing had lower viral loads.
Viner, et al.[141]	JAMA Pediatrics	Multiple	Review	In systematic review and meta-analysis of 32 studies, individuals <20 years had 44% lower odds of secondary infection with SARS-CoV-2 compared with adults ≥20 years; in those younger than 10 years, odds were 48% lower; in those older than 14 years, odds were not different.
Brotons, et al. [260]	CID	Spain	Serology	A seroprevalence study of 381 households and 1,084 contacts in Spain found that infection seroprevalence rates were similar in children and adults, but more likely to have mild or asymptomatic cases. This supports evidence for equal susceptibility. Positive pediatric contacts were found to also be mild or asymptomatic.

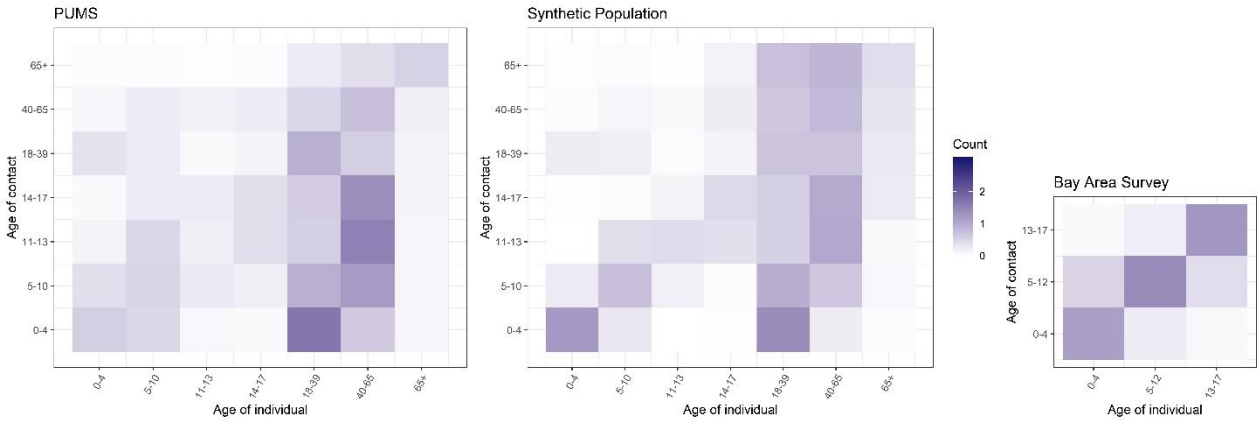


Figure S1. Validation of synthetic population.
To validate the household composition in our synthetic population, we compared the household contact matrix for individuals represented in Public Use Microdata Sample (PUMS) from the 2018 1-year American Community Survey (left) for 9 Bay area Counties (Alameda, San Francisco, Contra Costa, Marin, Napa, San Mateo, Santa Clara, Solano, and Sonoma) and one random draw of the synthetic population (middle). Similar patterns are reflected. Compared to PUMS, the number of household contacts of the same age groups within the synthetic population are elevated, which follows the pattern seen among our household Bay Area survey (right).

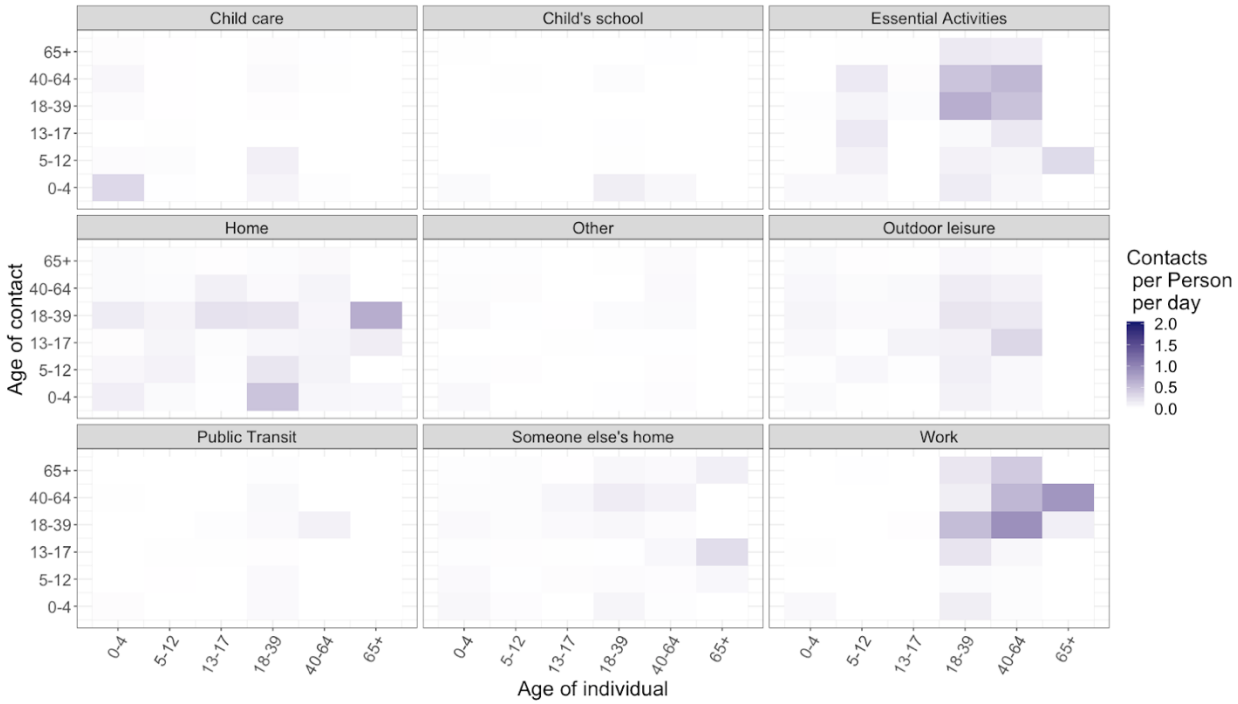


Figure S2. Location stratified contact matrices adjusted for non-response
We used inverse probability weighting to adjust for non-responses in location-specific contact rates. Weighted contact matrices did not differ substantially from unweighted matrices.

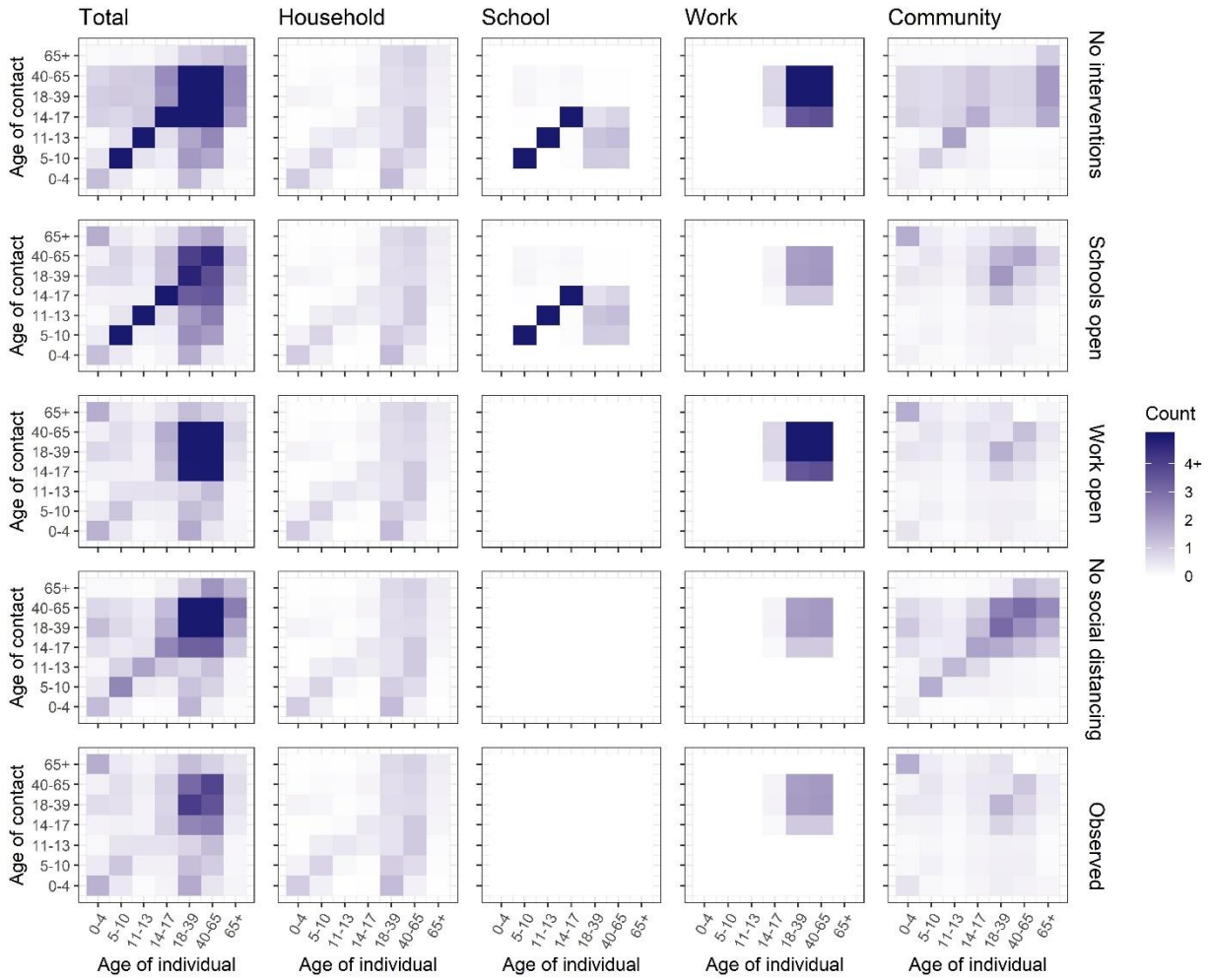


Figure S3. Age-specific contact matrices used for each counterfactual scenario.
Synthetic age-specific contact patterns across all locations, at home, in the workplace, in school, and at other locations during normal circumstances (i.e., under no intervention) are presented in the top row. Age-specific and location-specific contact matrices under the various counterfactual physical distancing interventions are presented in rows 2-4. Observed contact patterns are presented in the bottom row. Darker color intensities indicate higher proclivity of making the age-specific contact.

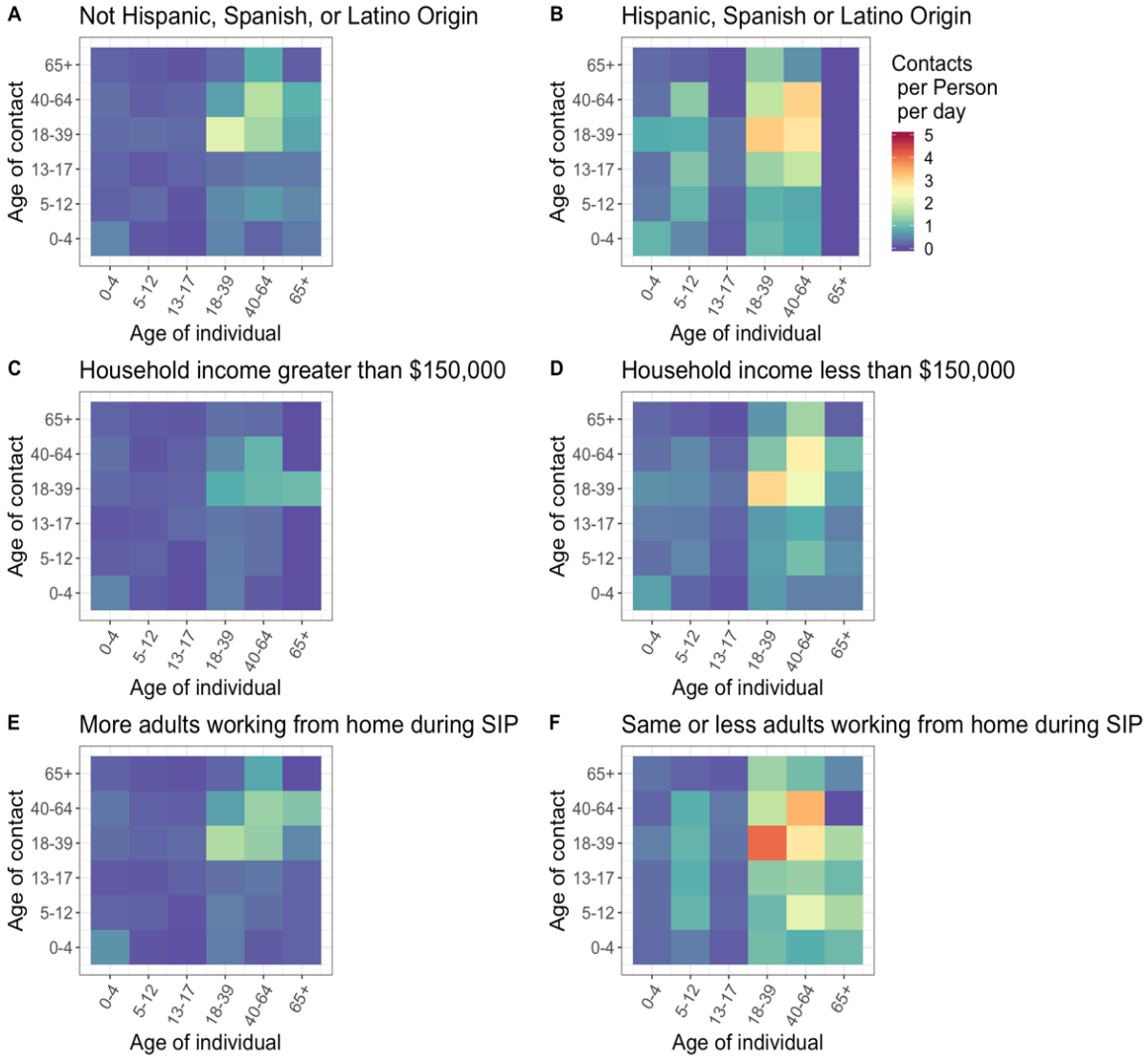


Figure S4. Community contact matrices by household characteristics
In multivariate adjusted regression modeling, Hispanic households had 2.32 (0.08, 4.5) more contacts than non-Hispanic households, households with an income less than $150,000 had 0.35 (-1.12, 1.8) less contacts compared to households with income less than $150,000, and households with less or the same number of adults working from home during Shelter in Place (SIP) had 1.85 (0.16, 3.52) more contacts than household with the same number or less adults working from home during SIP.
SIP: shelter-in-place

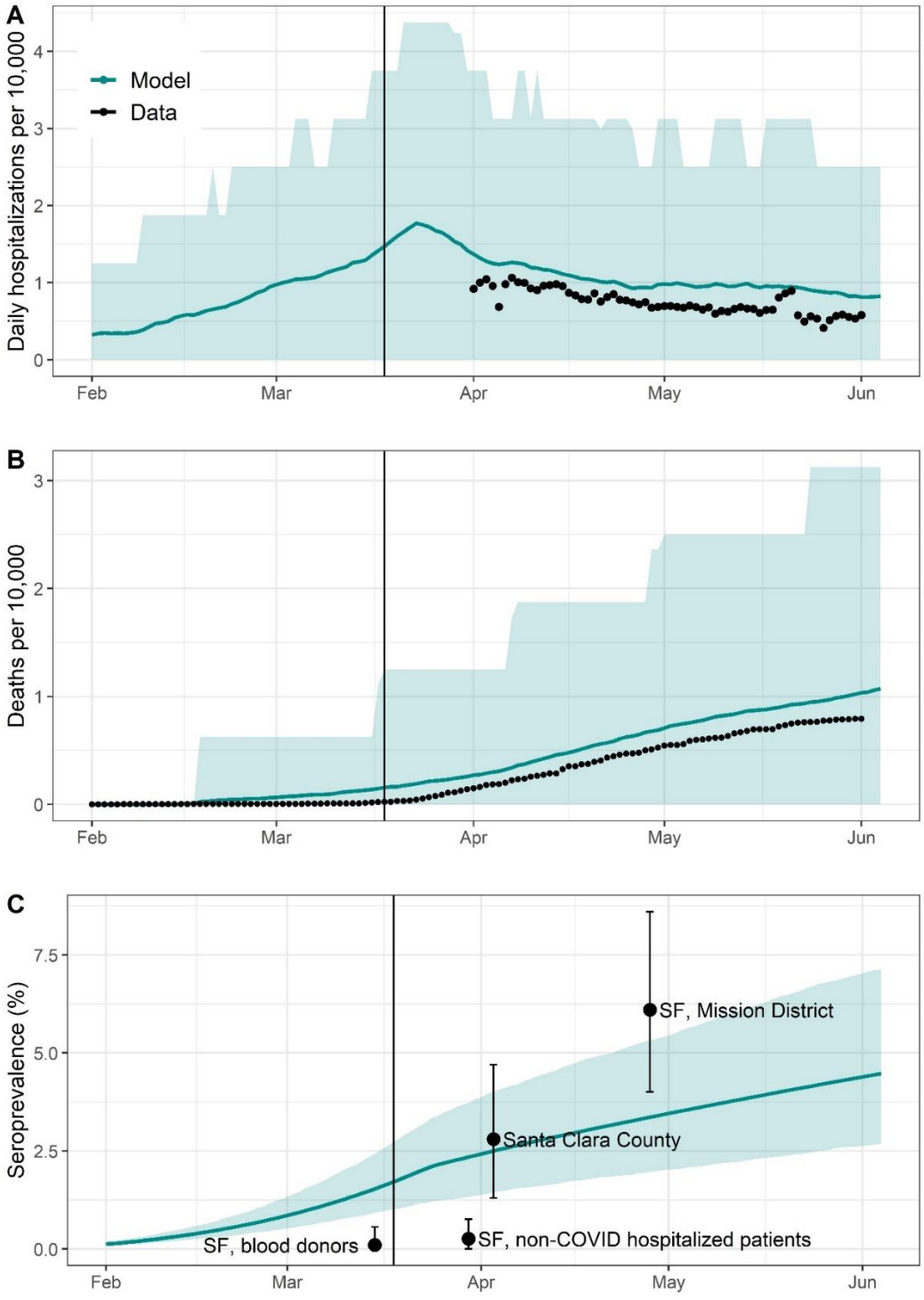


Figure S5. Comparison of model to observed data
Comparison of modelled (teal) to observed (black) data on daily hospitalizations per 10,000 population (A), cumulative deaths per 10,000 (B), and seroprevalence (C). The teal line represents the mean of 500 modelled simulations for the "observed" scenario, with the teal shaded region representing the 2.5th and 97.5th percentiles of model estimates.

A) Data on confirmed and suspected COVID-19 hospitalizations are downloadable from the California Department of Public Health open data portal.[261]
B) Reported deaths, and population per county are available from usafacts.org.[182] Since usafacts.org reports only confirmed deaths, we upweighted deaths by the time-varying ratio of confirmed to confirmed and suspected COVID-19 cases from ICU data.
C) Estimates of the seroprevalence of infection are obtained from studies conducted in various populations from the Bay Area: blood donors from the San Francisco Bay Area, patients hospitalized at a San Francisco hospital with confirmed negative test for COVID-19,[262] Santa Clara County,[263] and La Mission District,[185] a neighborhood in San Francisco with a high Latinx and essential worker population. We expect the seroprevalence of blood donors and patients hospitalized for non-COVID infections to be lower than in the source population given that blood donors tend to be healthier than the average population, and that the hospitalized population precluded capturing of current COVID-19 cases. We expect seroprevalence in La Mission District to be higher than the source population given the large proportion of essential workers in this neighborhood.

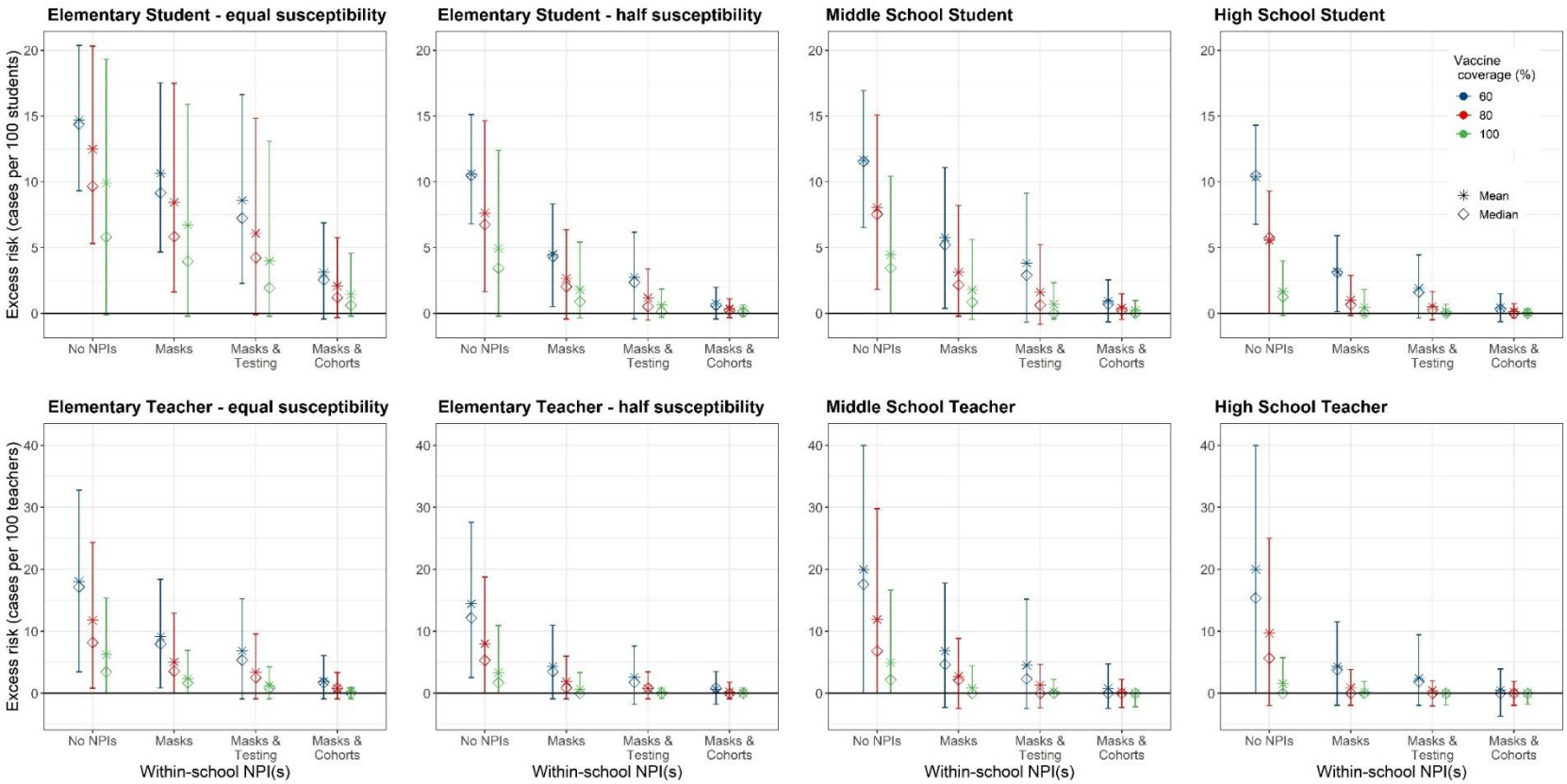


Figure S6. Effect of non-pharmaceutical interventions and vaccine coverages >70%. We examined the effect of three non-pharmaceutical interventions across three levels of community vaccination coverage (50%, 60%, 70%), assuming that vaccination coverage within school children 12+ and teachers matches that in the community and the vaccine effectiveness is 77% against infection, 85% against symptomatic infection, and 93% against severe infection. Masks indicate universal masks regardless of vaccination status. We calculated the mean (stars) and median (diamonds) of excess cases per 100 persons attributable to school transmission among population subgroups across 1,000 model realizations. Vertical lines reflect the 89thpercentile high probability density interval (HPDI)**.**

Discussion of interactive effect of multiple combined interventions

We found that the number of cases averted by closing a set of school levels is somewhat less than the summation of the averted cases achieved by closing each school level in isolation (Figure S6a). In other words, when one level of schooling is closed, each additional closure has a smaller marginal benefit. This is in part driven by the many households observed to have multiple school age children in different schooling levels. If a single schooling level alone (e.g., elementary only, middle only) is open, each child has the possibility to acquire an infection at school, and transmit at home, say, to a parent. If multiple school levels are open, the parent is placed at risk by multiple attending children, but only one case (the parent) can be averted. That is, if multiple grade levels were opened, and each child were infected at school, at most one onward transmission event could occur at home. Thus, closing multiple school levels does not avert cases at a rate proportional to the reduced risk.

In contrast, the number of excess cases associated with allowing social gatherings and in-person work, or allowing social gatherings and in-person school, was higher than the excess cases associated with either individually. (Figure S6b). This suggests that, by itself, social distancing is the least effective intervention; yet it becomes an important control measure when workplaces and schools are open. This can be explained as social gatherings act to spread infection across groups. Consider a social gathering of 10 individuals with 10 different workplaces. If their workplaces are all open, this gathering becomes significantly riskier as each individuals' exposure to infection is higher when their workplace is open. At the same time, if people are only interacting through contact at their workplaces and schools, outbreaks remain isolated and don't spread further throughout the network.

Finally, the effects of closing schools and workplaces were additive, in that the excess cases associated with schools alone being open added to the excess cases associated with workplaces alone being open equaled the excess cases if both were open, but social distancing maintained (Figure S6b). This is likely the result of a balancing of reductions in marginal benefits and increases in marginal benefits. On side that would have increasing marginal benefits, a child who experiences infection at school has the possibility to transmit the virus to their parent, who may then introduce infection to their workplace. The reverse is true in that an adult who acquires infection from work may pass the virus to a child, who brings it into a school environment. On side that would have decreasing marginal benefits, consider a household with one working adult and one school child. If schools only are open, and a child is infected at school, and that child transmits the virus to a working adult, it is most likely to be the working adult in their household. The reverse is true for if workplaces only are open, and the working parent is infected at the workplace and transmits the virus to their child. Therefore, if schools alone are open, one household transmission event can occur, and if workplaces alone are open, one household transmission event can occur. If both are open, there is still only one onward transmission event can occur at home (from child to parent, or parent to child).

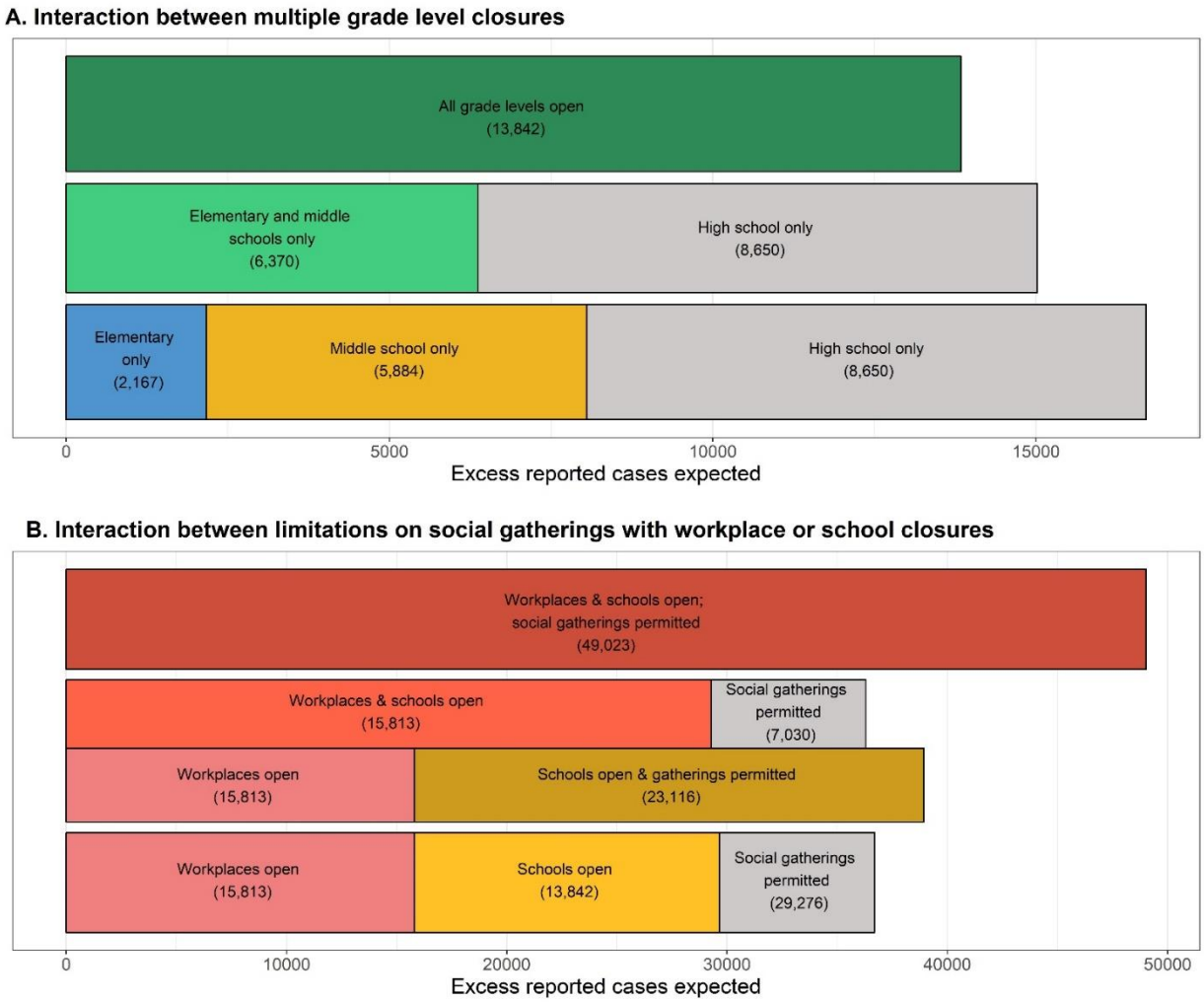


Figure S6. Interactive effect of multiple combined interventions.
We examined whether the effects of interventions were strictly additive or interactive when implemented in combination. Panel A demonstrates the effect of grade level closures. The excess cases associated with any set of grade levels being open was less than the sum of the excess cases associated with each individual grade level being open. Panel B demonstrates interaction between school closures, workplace closures, and limitations on social gatherings. School closures and workplace closures had an additive effect, meaning that the estimated excess cases associated with schools and workplaces being open was equal to the estimated excess cases of schools alone being open added to the estimated excess cases of workplaces along being open. Allowance of social gatherings interacted synergistically with workplaces or schools, whereby the excess cases associated with either social gatherings allowed and workplaces open, or social gatherings allowed and schools open, was greater than the sum of the individual effects.

Table S1: Synthetic population parameters

Parameter	Synthetic city (Oakland)	Source
Total population	429,082	US Census Bureau[264]
Proportion <5 years old	0.063	US Census Bureau[264]
Proportion 5-17 years old	0.135	US Census Bureau[264]
Proportion >65 years old	0.129	US Census Bureau[264]
Proportion of HH with kids <18 years old	0.251	Bay Area Census[265]
Average size of HH with kids	4.2	Bay Area Census[265]
Average size of HH without kids	2.5	Bay Area Census[265]
Proportion of single parent HHs	0.4	Healthy Alameda County[266]
Proportion of HHs that are intergenerational	0.04	US Census Bureau[267]
Mean age of mother at first parity	29	New York Times[268]
Median Age	33	Bay Area Census[265]
Number of public elementary schools	74	CA Dept. of Education[269]
Number of public middle schools	16	CA Dept. of Education[269]
Number of public high schools	18	CA Dept. of Education[269]
Proportion of kids in public school	0.87	Ed-Data.org[270]
Average class size (elementary school)	20	CA Dept. of Education[269]
Average class size (middle school)	19	CA Dept. of Education[269]
Average class size (high school)	19	CA Dept. of Education[269]
Average total size (elementary school)	383	CA Dept. of Education[269]
Average total size (middle school)	414	CA Dept. of Education[269]
Average total size (high school)	619	CA Dept. of Education[269]

Table S2. Representativeness of survey sample to Bay Area population

Number and proportion of individuals represented in the survey stratified by income and race sample before and after applying demographic weights. Public Use Microdata Sample (PUMS) from the 2018 1-year American Community Survey (ACMS) were used to calculate the expected distribution of income and race across for 9 Bay area Counties (Alameda, San Francisco, Contra Costa, Marin, Napa, San Mateo, Santa Clara, Solano, and Sonoma County). The expected distribution of race and income was compared against the demographic distribution of the 612 respondents in a web-based contact survey distributed across the Bay Area. Demographic weights were calculated by dividing the expected ACMS proportion by the proportion of the demographic represented in the survey sample.

		Un-weighted N (%)	Weighted N (%)	ACMS N (%)
Income				
	Less than $19,999	33 (5.4)	27.7 (4.9)	358345 (4.7)
	$20,000 to $39,999	51 (8.4)	46.3 (8.2)	589603 (7.8)
	$40,000 to $59,999	53 (8.7)	50.0 (8.8)	635775 (8.4)
	$60,000 to $79,999	61 (10.0)	49.8 (8.8)	641278 (8.5)
	$80,000 to $99,999	58 (9.5)	50.2 (8.8)	640969 (8.5)
	$100,000 to $149,999	106 (17.4)	96.2 (17)	1392802 (18.4)
	$150,000 or more	248 (40.7)	247 (43.6)	3124120 (41.3)
Race				
	White alone	341 (55.9)	286 (50.5)	3679854 (48.6)
	Black or African American alone	56 (9.2)	33.7 (5.9)	460135 (6.1)
	American Indian or Alaskan Native	2 (0.3)	0.4 (0.01)	11094 (0.15)
	Asian Alone	159 (26.1)	160 (28.2)	2077052 (27.4)
	Native Hawaiian or Pacific Islander Alone	2 (0.3)	0.39 (0.01)	39793 (0.50)
	Some other race alone	23 (3.8)	50.2 (8.9)	811534 (10.0)
	Two or more races	27 (4.4)	36.1 (6.4)	463094 (61)

Table S3. Characteristics of survey respondents

Characteristics of the 612 households who responded to the contact survey administered via Qualtrics between May 4, 2020 and June 1, 2020.

		n (%) (N = 612)
County		
	Alameda	218 (35.6)
	Contra Costa	121 (19.8)
	Marin	4 (0.7)
	Napa	5 (0.8)
	San Francisco	69 (11.3)
	San Mateo	42 (6.9)
	Santa Clara	108 (17.6)
	Sonoma	23 (3.8)
Race		
	White alone	341 (55.7)
	Asian alone	159 (26)
	Black or African American alone	56 (9.2)
	Two or more races	27 (4.4)
	Some other race alone	23 (3.8)
	American Indian or Alaska Native	4 (0.7)
	Native Hawaiian or Pacific islander alone	2 (0.3)
Household Income		
	Less than $19,999	34 (5.6)
	$20,000 to $39,999	51 (8.3)
	$40,000 to $59,999	53 (8.7)
	$60,000 to $79,999	61 (10)
	$80,000 to $99,999	58 (9.5)
	$100,000 to $149,999	107 (17.5)
	$150,000 or more	248 (40.5)
Single Parent Household		
	No	555 (90.7)
	Yes	57 (9.3)
Weekday of Reported Contacts		
	Monday	53 (8.7)
	Tuesday	102 (16.7)
	Wednesday	175 (28.6)
	Thursday	103 (16.8)
	Friday	77 (12.6)
	Saturday	59 (9.6)
	Sunday	43 (7)

Table S4. Composition of community matrices

The age-structured community matrix for analyses examining the effect of the spring semester closure (March 17 - June 1) was generated through a combination of POLYMOD and survey location data. Green boxes indicate the contact matrix was added to the overall community matrix, and red boxes indicate the contact matrix was subtracted from the overall community matrix.

	Survey				POLYMOD		Synthetic population	
Counterfactual scenario:	**Community***	**Daycare**	**Work**	**Transit**	**Community***	**Transit**	**Work**	**School**
Observed							**28% of workplaces**	
Schools open							**28% of workplaces**	**100%**
Workplaces open							**100%**	
Socializing permitted							**28% of workplaces + 10% of co-workers****	10% of classmates**
No interventions							**100%**	**100%**

*Community excludes school, work, and household contacts, but includes daycare, transit, essential activities, others' homes, and leisure.

**We assumed that in the absence of school, individuals would socialize would classmates and co-workers

Table S5. Parameters of the susceptible-exposed-infected-recovered model

Parameter	Ages (i)	Values	References
Basic reproduction number, R_0	all	5·0 – Delta variant 2·5 – Alpha	Kucharski, et al[161] Wu, et al[162] CDC [168]
Proportion of infections attributable to Delta variant	all	84%	CDPH [167]
Average incubation period, d_L (95% CI)	all	5.4 (2.4, 8.3)	Guan, et al[241] Li, et al[242] Lauer, et al[243]
Average duration of infection, non-hospitalized individuals, d_I (95% CI)	all	13.1 (8.3, 16.9)	Huang, et al[244]
Average time from infection to hospitalization, d_H (95% CI)	all	10.3 (6.5, 13.3)	Wang, et al[245]
Average duration of hospitalization, individuals who recover, d_R, or die, d_M (95% CI)	all	14.4 (11.3, 16.6)	Lewnard, et al[246]
Probability case is clinical, Pr(clinical|age)	$i < 20$ $i \geq 20$	0.21 0.69	Davies, et al[163]
Probability infection is acquired from subclinical transmission, α	all	0.50	Davies, et al[163] Prem, et al[271]
Probability of hospitalization among clinical cases, Pr(hospital|age)	$i < 10$ $10 \leq i < 20$ $20 \leq i < 30$ $30 \leq i < 40$ $40 \leq i < 50$ $50 \leq i < 60$ $60 \leq i < 70$ $70 \leq i < 80$ $i \geq 80$	0.00001 0.000408 0.0104 0.0343 0.0425 0.0816 0.118 0.166 0.184	Verity, et al[272]
Probability of death among hospitalized patients, Pr (death|age, hospital)	$i < 20$ $20 \leq i < 30$ $30 \leq i < 40$ $40 \leq i < 50$ $50 \leq i < 60$ $60 \leq i < 70$ $70 \leq i < 80$ $i \geq 80$	0.02 0.031 0.0475 0.0785 0.1215 0.186 0.301 0.4515	Lewnard et al[246]
Ratio of susceptibility among adults to susceptibility among children, $\beta_{i<20}/\beta_{i\geq20}$	all	0.50 or 1	Various; (see Supporting Information)
Vaccination coverage (for later analysis)	$i \geq 12$	50 – 95%	CDC Data Tracker [181]
Vaccine effectiveness against: - any infection - symptomatic infection - severe infection (requiring hospitalization)	$i \geq 12$	 77% 85% 93%	 Higdon, et al [170] Bernal, et al [171] Self, et al [172]

Table S6. Effect of reopening strategies: children half as susceptible, moderate community transmission.

Excess proportion of infections, symptomatic infections, hospitalizations, or deaths attributable to school re-openings, that would be experienced by each sub-group over a four-month period if schools were allowed to open under certain circumstances compared to if schools remained closed.

Interv.	Subgroup	Excess percent affected, % (95% CI)		Excess rate per 10,000 sub-pop. (95% CI)	
		Infection	Symptomatic infection	Hospitalizations	Deaths
None	Teachers (all)	14.83 (0.93, 29.25)	10.27 (0.47, 20.66)	40.5 (-46.95, 146.64)	2.97 (0, 47.17)
	HighSch teachers	30.5 (0, 64.17)	20.95 (0, 46.01)	76.53 (0, 384.62)	4.94 (0, 181.82)
	Middle teachers	19.53 (-2.33, 52.18)	13.44 (-2.27, 38.65)	55.01 (0, 444.44)	5.04 (0, 5.04)
	Elementary teachers	5.75 (-1.72, 15.79)	4.09 (-1.75, 11.97)	18.14 (-86.96, 172.45)	1.26 (0, 1.26)
	Students (all)	14.18 (1.63, 26.77)	2.98 (0.33, 5.83)	0.08 (0, 0.08)	0.01 (0, 0.01)
	HighSch students	25.75 (0.78, 55.99)	5.4 (0, 12.17)	0.17 (0, 0.17)	0 (0, 0)
	Middle students	17.21 (0, 46.34)	3.62 (0, 10.11)	0.11 (0, 0.11)	0.03 (0, 0.03)
	Elementary students	4.93 (0.22, 13.48)	1.04 (-0.1, 2.96)	0.01 (0, 0.01)	0.01 (0, 0.01)
	HH member	2.04 (-0.77, 5.07)	1.38 (-0.51, 3.73)	6.86 (-14.32, 30.11)	0.87 (-3.8, 7.48)
	Community member	1.16 (-0.9, 3.28)	0.79 (-0.7, 2.35)	4.2 (-7.33, 16.32)	0.54 (-2.73, 3.66)
Stable cohorts (weak)	Teachers (all)	3.16 (-1.42, 8.74)	2.18 (-1.41, 6.19)	8.46 (-47.39, 91.76)	0.61 (0, 0.61)
	HighSch teachers	5.65 (-3.85, 21.06)	3.88 (-3.7, 15.38)	16.62 (-185.19, 196.08)	0.2 (0, 0.2)
	Middle teachers	3.92 (-4.65, 17.03)	2.68 (-4.55, 13.04)	9.27 (0, 222.22)	0.85 (0, 0.85)
	Elementary teachers	1.72 (-3.51, 7.69)	1.21 (-2.7, 5.98)	4.32 (-86.96, 87.74)	0.71 (0, 0.71)
	Students (all)	2.92 (0.19, 6.96)	0.61 (-0.05, 1.62)	0.03 (0, 0.03)	0 (0, 0)
	HighSch students	4.58 (-0.46, 14.09)	0.96 (-0.17, 3.28)	0.08 (0, 0.08)	0 (0, 0)
	Middle students	3.41 (-0.6, 12.89)	0.71 (-0.4, 2.97)	0 (0, 0)	0 (0, 0)
	Elementary students	1.56 (-0.23, 4.53)	0.34 (-0.22, 1.22)	0.01 (0, 0.01)	0 (0, 0)
	HH member	0.5 (-1.23, 2.5)	0.35 (-0.94, 1.8)	2.19 (-15.29, 22.34)	0.32 (-3.83, 7.33)
	Community member	0.29 (-1.18, 1.8)	0.2 (-0.89, 1.27)	0.92 (-9.08, 11.86)	0.16 (-2.75, 2.75)
Stable cohorts (strong)	Teachers (all)	1.25 (-2.77, 5.16)	0.92 (-1.88, 4.25)	2.14 (-47.39, 47.85)	0.24 (0, 0.24)
	HighSch teachers	1.9 (-5.66, 10.72)	1.39 (-3.92, 9.44)	3.31 (-185.19, 188.68)	0.77 (0, 0.77)
	Middle teachers	1.65 (-6.52, 9.56)	1.2 (-4.55, 8.7)	2.34 (-222.22, 222.22)	0 (0, 0)
	Elementary teachers	0.79 (-3.51, 5.22)	0.6 (-2.7, 4.42)	1.59 (-86.96, 87.72)	0.46 (0, 0.46)
	Students (all)	1.3 (0.05, 3.41)	0.27 (-0.1, 0.81)	0 (0, 0)	0 (0, 0)
	HighSch students	1.64 (-0.65, 6.08)	0.35 (-0.32, 1.48)	0 (0, 0)	0 (0, 0)
	Middle students	1.53 (-0.65, 5.36)	0.31 (-0.41, 1.42)	0 (0, 0)	0 (0, 0)
	Elementary students	0.95 (-0.42, 3.12)	0.2 (-0.31, 0.88)	0 (0, 0)	0 (0, 0)
	HH member	0.22 (-1.55, 2.08)	0.15 (-1.19, 1.53)	0.73 (-17.97, 18.49)	0.17 (-3.8, 3.97)
	Community member	0.15 (-1.33, 1.54)	0.1 (-0.96, 1.14)	0.49 (-9.94, 10.04)	0.06 (-2.73, 2.76)
2-day half class shifts	Teachers (all)	0.7 (-2.38, 3.85)	0.49 (-2.34, 3.27)	-0.01 (-47.39, 47.85)	-0.18 (-0.18, 0)
	HighSch teachers	1.03 (-5.66, 9.09)	0.63 (-5.66, 6)	-0.83 (-185.19, 172.49)	-0.49 (-0.49, 0)
	Middle teachers	0.85 (-6.38, 8.89)	0.54 (-4.65, 6.67)	-1.47 (-222.22, 217.39)	-0.58 (-0.58, 0)
	Elementary teachers	0.48 (-3.6, 5.08)	0.4 (-2.72, 3.54)	0.9 (-87.72, 87.72)	0.11 (0, 0.11)
	Students (all)	0.4 (-0.44, 1.31)	0.09 (-0.19, 0.39)	0.01 (0, 0.01)	0 (0, 0)
	HighSch students	0.52 (-0.94, 2.24)	0.12 (-0.35, 0.75)	0 (0, 0)	0 (0, 0)
	Middle students	0.45 (-0.95, 2.4)	0.09 (-0.45, 0.67)	0.03 (0, 0.03)	0 (0, 0)
	Elementary students	0.29 (-0.71, 1.3)	0.06 (-0.32, 0.5)	0 (0, 0)	0 (0, 0)
	HH member	0.09 (-1.59, 1.8)	0.06 (-1.14, 1.3)	-0.05 (-18.38, 18.29)	0.06 (-3.8, 4.01)
	Community member	0.04 (-1.42, 1.55)	0.03 (-0.96, 1.08)	-0.03 (-10.03, 9.87)	0.01 (-2.74, 2.75)
2-day staggered grades	Teachers (all)	0.68 (-2.78, 4.13)	0.46 (-2.3, 3.24)	2.12 (-47.62, 47.85)	-0.05 (-0.05, 0)
	HighSch teachers	1.12 (-5.56, 9.43)	0.75 (-3.92, 5.88)	5.91 (-185.19, 192.31)	0.17 (0, 0.17)
	Middle teachers	0.92 (-6.53, 9.09)	0.6 (-4.65, 6.98)	4.02 (0, 222.22)	-0.01 (-0.01, 0)
	Elementary teachers	0.39 (-4.2, 4.43)	0.27 (-3.42, 4.27)	-0.35 (-87.72, 87.72)	-0.17 (-0.17, 0)
	Students (all)	0.55 (-0.32, 1.66)	0.12 (-0.19, 0.44)	0.01 (0, 0.01)	0 (0, 0)
	HighSch students	0.8 (-0.75, 3.04)	0.18 (-0.33, 0.82)	0 (0, 0)	0 (0, 0)
	Middle students	0.77 (-0.88, 3.42)	0.17 (-0.43, 1)	0.02 (0, 0.02)	0 (0, 0)
	Elementary students	0.26 (-0.67, 1.31)	0.05 (-0.33, 0.49)	0 (0, 0)	0 (0, 0)
	HH member	0.15 (-1.65, 1.92)	0.1 (-1.23, 1.35)	0.9 (-18.34, 18.7)	0.12 (-3.79, 7.24)
	Community member	0.09 (-1.48, 1.46)	0.06 (-0.98, 1.08)	0.18 (-9.98, 9.96)	0.04 (-2.76, 3.6)

Masks	Teachers (all)	1.73 (-2.32, 6.29)	1.22 (-1.89, 4.76)	4.2 (-47.39, 48.09)	0.44 (0, 0.44)
	HighSch teachers	4.38 (-3.92, 18.19)	2.94 (-3.85, 13.46)	10.2 (-181.82, 192.31)	0.77 (0, 0.77)
	Middle teachers	1.84 (-6.39, 11.63)	1.31 (-4.65, 8.89)	3.41 (-217.51, 222.22)	-0.29 (-0.29, 0)
	Elementary teachers	0.45 (-3.54, 5.08)	0.38 (-3.42, 4.31)	1.72 (-87.72, 87.74)	0.58 (0, 0.58)
	Students (all)	2.51 (0.05, 6.95)	0.53 (-0.05, 1.65)	0.07 (0, 0.01)	0 (0, 0)
	HighSch students	5.42 (-0.45, 18.37)	1.13 (-0.17, 4.21)	0.16 (0, 0.16)	0 (0, 0)
	Middle students	2.37 (-0.62, 9.43)	0.5 (-0.41, 2.36)	0.05 (0, 0.05)	0 (0, 0)
	Elementary students	0.64 (-0.55, 2.2)	0.14 (-0.31, 0.63)	0.01 (0, 0.01)	0 (0, 0)
	HH member	0.35 (-1.45, 2.34)	0.24 (-1.15, 1.72)	0.88 (-18.12, 18.47)	0.19 (-3.8, 3.83)
	Community member	0.21 (-1.42, 2.01)	0.14 (-1, 1.36)	0.8 (-9.91, 11.01)	0.06 (-2.73, 2.75)
Monthly testing w/ case isolation & class quarantine	Teachers (all)	12.9 (0.48, 26.64)	8.94 (0.47, 19.25)	34.98 (-46.95, 141.53)	3.28 (0, 47.17)
	HighSch teachers	27.06 (-1.79, 61.54)	18.51 (-1.82, 44.24)	72.19 (0, 384.8)	5.52 (0, 185.19)
	Middle teachers	16.61 (-2.33, 48.78)	11.64 (-2.27, 35.57)	46.7 (0, 434.78)	4.12 (0, 4.12)
	Elementary teachers	4.93 (-1.77, 14.17)	3.48 (-1.75, 11.02)	13.3 (-86.96, 90.09)	1.95 (0, 1.95)
	Students (all)	12.22 (1.25, 25.17)	2.55 (0.19, 5.51)	0.11 (0, 0.11)	0 (0, 0)
	HighSch students	22.57 (0.16, 52.71)	4.73 (0, 11.85)	0.25 (0, 0.25)	0 (0, 0)
	Middle students	14.62 (-0.2, 41.63)	3.04 (0, 9.33)	0.17 (0, 0.17)	0 (0, 0)
	Elementary students	4.17 (0.19, 11.22)	0.86 (-0.11, 2.57)	0 (0, 0)	0 (0, 0)
	HH member	1.76 (-0.83, 4.51)	1.19 (-0.6, 3.17)	6.6 (-14.82, 29.26)	0.92 (-3.76, 7.44)
	Community member	1.01 (-0.78, 2.97)	0.69 (-0.57, 2.08)	3.68 (-7.27, 15.54)	0.5 (-2.72, 3.68)
Stable cohorts (strong), masks, monthly testing	Teachers (all)	0.28 (-2.86, 3.32)	0.2 (-2.39, 2.86)	0.31 (-47.4, 47.62)	0 (0, 0)
	HighSch teachers	0.47 (-5.66, 7.55)	0.28 (-5.77, 5.88)	2.26 (-185.19, 185.27)	0 (0, 0)
	Middle teachers	0.39 (-6.52, 6.82)	0.26 (-4.65, 6.67)	-0.71 (-222.22, 212.88)	0 (0, 0)
	Elementary teachers	0.15 (-3.65, 4.35)	0.13 (-3.45, 3.51)	-0.22 (-86.98, 87.72)	0.23 (0, 0.23)
	Students (all)	0.45 (-0.38, 1.39)	0.09 (-0.2, 0.42)	0.02 (0, 0.02)	0 (0, 0)
	HighSch students	0.68 (-0.92, 2.97)	0.13 (-0.34, 0.82)	0.04 (0, 0.04)	0 (0, 0)
	Middle students	0.56 (-0.99, 2.93)	0.12 (-0.43, 0.83)	0.03 (0, 0.03)	0 (0, 0)
	Elementary students	0.25 (-0.65, 1.3)	0.05 (-0.41, 0.43)	0 (0, 0)	0 (0, 0)
	HH member	0.07 (-1.7, 1.79)	0.03 (-1.25, 1.33)	0.17 (-15.32, 18.61)	0.14 (-3.8, 3.82)
	Community member	0.05 (-1.4, 1.51)	0.03 (-1.08, 1.06)	0.37 (-10.81, 10.04)	0.07 (-2.74, 2.77)
2-day half class shifts + stable cohorts	Teachers (all)	0.24 (-2.84, 3.37)	0.19 (-2.37, 2.84)	-0.01 (-47.39, 47.85)	0.43 (0, 0.43)
	HighSch teachers	0.35 (-5.77, 7.41)	0.13 (-5.66, 5.77)	-0.83 (-185.19, 172.49)	0.45 (0, 0.45)
	Middle teachers	0.46 (-6.52, 6.98)	0.36 (-4.65, 6.67)	-1.47 (-222.22, 217.39)	-0.29 (-0.29, 0)
	Elementary teachers	0.1 (-4.31, 4.39)	0.14 (-2.68, 3.54)	0.9 (-87.72, 87.72)	0.69 (0, 0.69)
	Students (all)	0.17 (-0.47, 0.84)	0.04 (-0.2, 0.33)	0.01 (0, 0.01)	0 (0, 0)
	HighSch students	0.2 (-0.92, 1.4)	0.03 (-0.47, 0.53)	0 (0, 0)	0 (0, 0)
	Middle students	0.21 (-1.08, 1.71)	0.06 (-0.44, 0.64)	0.03 (0, 0.03)	0 (0, 0)
	Elementary students	0.14 (-0.71, 1.12)	0.03 (-0.32, 0.42)	0 (0, 0)	0 (0, 0)
	HH member	0.04 (-1.65, 1.7)	0.02 (-1.22, 1.26)	-0.05 (-18.38, 18.29)	0.16 (-3.8, 7.26)
	Community member	0.02 (-1.41, 1.44)	0.02 (-1, 1.03)	-0.03 (-10.03, 9.87)	-0.02 (-2.74, 2.74)
2-day staggered grades + stable cohorts	Teachers (all)	0.3 (-2.83, 3.32)	0.18 (-2.37, 2.8)	0.93 (-47.62, 47.62)	0.05 (0, 0.05)
	HighSch teachers	0.39 (-5.66, 7.41)	0.26 (-5.45, 5.77)	1.43 (-185.19, 188.68)	0.56 (0, 0.56)
	Middle teachers	0.45 (-6.52, 6.98)	0.28 (-4.65, 4.91)	3.36 (0, 222.22)	0.21 (0, 0.21)
	Elementary teachers	0.2 (-4.31, 4.35)	0.12 (-3.48, 3.48)	-0.3 (-87.72, 86.96)	-0.26 (-0.26, 0)
	Students (all)	0.29 (-0.49, 1.05)	0.06 (-0.23, 0.38)	0 (0, 0)	0 (0, 0)
	HighSch students	0.4 (-0.83, 2.15)	0.09 (-0.34, 0.68)	0.02 (0, 0.02)	0 (0, 0)
	Middle students	0.38 (-1, 2.01)	0.09 (-0.44, 0.7)	0 (0, 0)	0 (0, 0)
	Elementary students	0.17 (-0.78, 1.23)	0.03 (-0.34, 0.41)	0 (0, 0)	0 (0, 0)
	HH member	0.06 (-1.75, 1.74)	0.04 (-1.29, 1.34)	0.47 (-18.55, 18.74)	0.06 (-3.82, 3.83)
	Community member	0.03 (-1.45, 1.49)	0.02 (-1, 1.06)	-0.04 (-10.01, 9.11)	0 (-2.75, 2.74)
All interventions: 2-day staggered grades, cohorts, masks, monthly testing	Teachers (all)	0.1 (-2.84, 3.24)	0.04 (-2.37, 2.38)	-1.59 (-47.62, 47.62)	-0.23 (-0.23, 0)
	HighSch teachers	0.17 (-5.77, 5.88)	0.12 (-5.56, 5.66)	-0.09 (-185.19, 188.68)	-0.73 (-0.73, 0)
	Middle teachers	0.2 (-6.53, 6.82)	0.06 (-4.65, 6.52)	-3.79 (-222.22, 5.1)	-0.45 (-0.45, 0)
	Elementary teachers	0.03 (-4.27, 4.31)	-0.01 (-3.48, 3.48)	-1.4 (-86.96, 86.21)	0.09 (0, 0.09)
	Students (all)	0.1 (-0.53, 0.8)	0.02 (-0.25, 0.29)	0.01 (0, 0.01)	0 (0, 0)
	HighSch students	0.15 (-0.97, 1.33)	0.02 (-0.47, 0.49)	0.02 (0, 0.02)	0 (0, 0)
	Middle students	0.15 (-1.09, 1.55)	0.03 (-0.46, 0.61)	0.02 (0, 0.02)	0 (0, 0)
	Elementary students	0.04 (-0.81, 0.91)	0.01 (-0.35, 0.42)	0 (0, 0)	0 (0, 0)
	HH member	0.03 (-1.8, 1.78)	0.01 (-1.33, 1.35)	-0.06 (-18.29, 18.5)	0.02 (-3.84, 3.84)
	Community member	0.02 (-1.43, 1.61)	0.01 (-1.02, 1.12)	-0.02 (-10.01, 9.98)	-0.05 (-2.74, 2.74)

Table S7. Effect of reopening strategies: children half as susceptible, high community transmission.

Excess proportion of infections, symptomatic infections, hospitalizations, or deaths attributable to school re-openings, that would be experienced by each sub-group over a four-month period if schools were allowed to open under certain circumstances compared to if schools remained closed.

Interv.	Subgroup	Excess percent affected, % (95% CI)		Excess rate per 10,000 sub-pop. (95% CI)	
		Infection	Symptomatic infection	Hospitalizations	Deaths
None	Teachers (all)	26.68 (11.75, 39.29)	18.36 (7.69, 27.85)	75.71 (-46.6, 232.56)	5.86 (0, 47.62)
	HighSch teachers	48.21 (14.71, 71.54)	33.3 (11.11, 53.57)	140.19 (0, 566.04)	11.44 (0, 192.31)
	Middle teachers	35.61 (6.67, 60.47)	24.4 (4.31, 44.44)	96.72 (0, 454.55)	6.27 (0, 217.39)
	Elementary teachers	13.25 (1.76, 26.32)	9.1 (0.85, 18.97)	37.77 (-87.72, 178.57)	3.12 (0, 86.21)
	Students (all)	26.44 (12.92, 37.61)	5.55 (2.53, 8.32)	0.19 (0, 4.75)	0.01 (0, 0)
	HighSch students	43.9 (14.63, 63.25)	9.24 (2.93, 14.29)	0.29 (0, 0)	0.02 (0, 0)
	Middle students	33.17 (8.38, 54.64)	6.93 (1.52, 12.34)	0.34 (0, 0)	0.02 (0, 0)
	Elementary students	11.36 (3.56, 20.64)	2.39 (0.52, 4.73)	0.04 (0, 0)	0 (0, 0)
	HH member	5.27 (0.79, 9.59)	3.56 (0.43, 6.72)	18.55 (-11.17, 50.94)	1.95 (-7.27, 11.18)
	Community member	4.02 (0.27, 7.74)	2.73 (0.14, 5.26)	14.79 (-4.55, 34.02)	1.93 (-3.61, 7.31)
Stable cohorts (weak)	Teachers (all)	7.57 (0.22, 15.93)	5.24 (0, 11.4)	20.16 (-47.62, 96.4)	1.45 (0, 46.73)
	HighSch teachers	12.68 (-1.89, 32.71)	8.91 (-1.92, 23.19)	33.35 (-187.02, 200)	2.95 (0, 0)
	Middle teachers	9.44 (-4.4, 25.56)	6.63 (-4.44, 20)	28.57 (-222.22, 232.56)	1.97 (0, 0)
	Elementary teachers	4.5 (-3.42, 12.93)	3.01 (-3.42, 9.53)	10.81 (-88.5, 170.25)	0.58 (0, 0)
	Students (all)	7.17 (2.59, 13.52)	1.49 (0.38, 2.92)	0.02 (0, 0)	0 (0, 0)
	HighSch students	10.74 (1.68, 23.79)	2.24 (0.16, 5.25)	0.07 (0, 0)	0 (0, 0)
	Middle students	8.53 (0.98, 19.76)	1.78 (-0.21, 4.53)	0.02 (0, 0)	0 (0, 0)
	Elementary students	4.11 (0.51, 9.15)	0.86 (-0.21, 2.19)	-0.01 (0, 0)	0 (0, 0)
	HH member	1.62 (-1.88, 5.2)	1.09 (-1.4, 3.71)	6.15 (-21.86, 34.43)	0.58 (-7.49, 7.64)
	Community member	1.23 (-2.05, 4.3)	0.83 (-1.41, 2.97)	4.64 (-11.81, 21.32)	0.35 (-4.51, 5.43)
Stable cohorts (strong)	Teachers (all)	3.32 (-1.89, 8.94)	2.25 (-1.91, 7.02)	9.15 (-48.64, 94.79)	1.14 (0, 46.73)
	HighSch teachers	4.94 (-5.77, 16.98)	3.43 (-5.6, 13.21)	18.24 (-188.68, 200)	1.44 (0, 0)
	Middle teachers	3.84 (-8.51, 15.91)	2.57 (-6.73, 12.77)	8.98 (-222.22, 227.27)	1.1 (0, 0)
	Elementary teachers	2.38 (-4.39, 9.28)	1.58 (-3.52, 6.96)	5.07 (-88.5, 89.29)	1.01 (0, 0)
	Students (all)	3.34 (0.56, 6.56)	0.7 (0, 1.63)	0.03 (0, 0)	0 (0, 0)
	HighSch students	4.27 (0.15, 10.23)	0.87 (-0.32, 2.52)	0.05 (0, 0)	0 (0, 0)
	Middle students	3.8 (-0.45, 10.23)	0.81 (-0.62, 2.57)	0.06 (0, 0)	0 (0, 0)
	Elementary students	2.5 (-0.25, 6.1)	0.54 (-0.31, 1.57)	0 (0, 0)	0 (0, 0)
	HH member	0.86 (-2.27, 4.18)	0.58 (-1.57, 2.95)	3.11 (-25.73, 29.24)	0.31 (-7.45, 7.5)
	Community member	0.68 (-2.14, 3.48)	0.46 (-1.63, 2.39)	2.52 (-14.48, 18.29)	0.27 (-4.53, 4.56)
2-day half class shifts	Teachers (all)	1.88 (-3.28, 6.96)	1.29 (-2.78, 5.67)	4.36 (-92.17, 94.34)	1.15 (0, 46.82)
	HighSch teachers	2.86 (-7.07, 13.57)	1.99 (-5.81, 11.19)	8.12 (-190.13, 196.08)	1.23 (0, 0)
	Middle teachers	2.12 (-8.7, 13.33)	1.53 (-6.67, 11.11)	7.67 (-224.24, 227.27)	1.1 (0, 0)
	Elementary teachers	1.34 (-5.22, 7.76)	0.88 (-4.33, 6.25)	1.3 (-89.29, 89.29)	1.1 (0, 0)
	Students (all)	0.98 (-0.47, 2.64)	0.2 (-0.29, 0.75)	0.02 (0, 0)	0 (0, 0)
	HighSch students	1.29 (-0.93, 4.39)	0.28 (-0.63, 1.32)	0.03 (0, 0)	0 (0, 0)
	Middle students	1.1 (-1.35, 4.1)	0.21 (-0.81, 1.39)	0.04 (0, 0)	0 (0, 0)
	Elementary students	0.7 (-1.05, 2.82)	0.15 (-0.52, 0.88)	0 (0, 0)	0 (0, 0)
	HH member	0.25 (-2.85, 3.36)	0.16 (-2.22, 2.43)	1.68 (-25.34, 29.25)	0.17 (-7.47, 7.45)
	Community member	0.21 (-2.5, 3.08)	0.14 (-1.73, 2.15)	1.27 (-14.41, 17.12)	0.11 (-4.54, 4.57)
2-day staggered grades	Teachers (all)	1.74 (-3.27, 7.04)	1.24 (-2.85, 5.63)	4.72 (-48.08, 93.9)	0.26 (0, 0)
	HighSch teachers	2.82 (-7.48, 12.96)	2.04 (-5.66, 10)	6.65 (-190.58, 196.08)	1.28 (0, 0)
	Middle teachers	2.46 (-8.61, 13.49)	1.72 (-6.82, 11.11)	8.55 (-222.22, 227.27)	0.76 (0, 0)
	Elementary teachers	0.96 (-5.22, 7.6)	0.68 (-4.37, 6.14)	2.33 (-88.5, 89.29)	-0.38 (0, 0)
	Students (all)	1.43 (-0.23, 3.47)	0.29 (-0.25, 0.88)	0.01 (0, 0)	0 (0, 0)
	HighSch students	2.18 (-0.67, 5.97)	0.44 (-0.61, 1.54)	0.05 (0, 0)	0 (0, 0)
	Middle students	2.03 (-1.12, 6.37)	0.42 (-0.65, 1.77)	-0.02 (0, 0)	0 (0, 0)
	Elementary students	0.63 (-1.11, 2.51)	0.12 (-0.53, 0.85)	0 (0, 0)	0 (0, 0)
	HH member	0.34 (-2.57, 3.41)	0.22 (-1.92, 2.47)	1.83 (-25.89, 29.31)	0.32 (-7.47, 7.47)
	Community member	0.23 (-2.58, 3.03)	0.15 (-1.82, 2.11)	1.22 (-15.39, 17.81)	0.04 (-4.56, 4.58)

Masks	Teachers (all)	4.87 (-1.41, 11.84)	3.34 (-1.87, 8.76)	14.22 (-48.08, 96.15)	1.29 (0, 46.73)
	HighSch teachers	10.47 (-3.77, 27.27)	7.06 (-3.77, 20.52)	32.84 (-188.68, 363.64)	2.93 (0, 0)
	Middle teachers	5.7 (-6.82, 18.35)	3.97 (-4.76, 14.89)	14.25 (-222.22, 227.27)	1.51 (0, 0)
	Elementary teachers	1.97 (-5.1, 9.18)	1.38 (-3.6, 7.08)	5.6 (-88.5, 90.09)	0.42 (0, 0)
	Students (all)	6.29 (1.51, 12.68)	1.31 (0.28, 2.74)	0.03 (0, 0)	0 (0, 0)
	HighSch students	12.92 (1.96, 27.26)	2.71 (0.24, 6.36)	0.06 (0, 0)	0 (0, 0)
	Middle students	6.47 (0.12, 15.68)	1.34 (-0.43, 3.74)	0.04 (0, 0)	0 (0, 0)
	Elementary students	1.79 (-0.58, 4.48)	0.37 (-0.4, 1.23)	0 (0, 0)	0 (0, 0)
	HH member	1.37 (-2.11, 4.76)	0.92 (-1.55, 3.37)	4.78 (-22.63, 33.02)	0.49 (-7.45, 7.62)
	Community member	1.08 (-1.93, 4.18)	0.74 (-1.34, 2.89)	4.2 (-12.75, 20.1)	0.53 (-4.53, 5.48)
Monthly testing w/ case isolation & class quarantine	Teachers (all)	24.71 (11.01, 36.5)	17.07 (7.06, 26.23)	69.41 (-46.82, 191.39)	5.4 (0, 47.62)
	HighSch teachers	45.37 (14.81, 67.32)	31.38 (9.09, 51.79)	132.26 (0, 545.45)	10.48 (0, 188.68)
	Middle teachers	32.38 (5.69, 56.52)	22.36 (3.56, 42.35)	89.16 (-212.77, 454.55)	6.71 (0, 217.39)
	Elementary teachers	12.18 (1.78, 23.79)	8.41 (0.85, 17.07)	32.65 (-87.72, 176.99)	2.54 (0, 86.21)
	Students (all)	24.14 (11.22, 35.5)	5.08 (2.17, 7.58)	0.22 (0, 4.81)	0 (0, 0)
	HighSch students	40.33 (12.71, 60.65)	8.44 (2.34, 13.39)	0.38 (0, 0)	0.02 (0, 0)
	Middle students	29.7 (7.66, 51.44)	6.21 (1.25, 11.72)	0.43 (0, 0)	0 (0, 0)
	Elementary students	10.5 (3.39, 19.67)	2.25 (0.41, 4.51)	0.01 (0, 0)	0 (0, 0)
	HH member	4.78 (0.6, 8.89)	3.23 (0.26, 6.09)	16.73 (-14.86, 46.15)	1.51 (-7.34, 11.07)
	Community member	3.62 (-0.32, 7.13)	2.47 (-0.29, 4.93)	13.18 (-5.49, 32.39)	1.66 (-3.63, 7.31)
Stable cohorts (strong), masks, monthly testing	Teachers (all)	0.97 (-3.78, 6.07)	0.71 (-2.92, 4.73)	2.59 (-91.91, 93.02)	0.64 (0, 46.3)
	HighSch teachers	1.59 (-7.55, 11.63)	1.16 (-7.27, 9.43)	7.01 (-190.13, 196.08)	1.68 (0, 0)
	Middle teachers	1.1 (-8.89, 11.11)	0.81 (-6.98, 8.89)	0.09 (-227.27, 222.22)	-0.01 (0, 0)
	Elementary teachers	0.63 (-5.31, 6.98)	0.47 (-4.39, 5.41)	1.52 (-88.5, 89.29)	0.43 (0, 0)
	Students (all)	1.18 (-0.34, 2.98)	0.24 (-0.24, 0.77)	0.02 (0, 0)	0 (0, 0)
	HighSch students	1.85 (-0.8, 5.56)	0.4 (-0.5, 1.5)	0.05 (0, 0)	0 (0, 0)
	Middle students	1.29 (-1.38, 4.82)	0.25 (-0.76, 1.4)	0 (0, 0)	0 (0, 0)
	Elementary students	0.68 (-1.26, 2.67)	0.13 (-0.56, 0.84)	0.01 (0, 0)	0 (0, 0)
	HH member	0.26 (-2.9, 3.48)	0.17 (-2.11, 2.5)	1.45 (-25.67, 29.27)	0.17 (-7.41, 7.49)
	Community member	0.24 (-2.62, 3.11)	0.16 (-1.89, 2.14)	1.19 (-16.26, 17.61)	0.07 (-4.55, 4.58)
2-day half class shifts + stable cohorts	Teachers (all)	0.67 (-4.63, 5.19)	0.48 (-3.34, 4.41)	2.05 (-92.76, 93.2)	0.59 (0, 46.3)
	HighSch teachers	1.01 (-7.69, 10)	0.74 (-7.55, 7.84)	4.72 (-188.68, 192.31)	0.34 (0, 0)
	Middle teachers	0.94 (-8.89, 10.6)	0.72 (-6.82, 8.89)	2.32 (-222.22, 224.24)	1.03 (0, 0)
	Elementary teachers	0.42 (-6.09, 6.96)	0.28 (-5.17, 5.83)	0.79 (-89.29, 89.29)	0.52 (0, 0)
	Students (all)	0.39 (-0.86, 1.79)	0.08 (-0.38, 0.54)	-0.01 (0, 0)	0 (0, 0)
	HighSch students	0.42 (-1.59, 2.47)	0.09 (-0.65, 0.92)	-0.01 (0, 0)	0 (0, 0)
	Middle students	0.44 (-1.75, 2.88)	0.07 (-0.85, 1.04)	-0.02 (0, 0)	0 (0, 0)
	Elementary students	0.35 (-1.32, 2.11)	0.07 (-0.54, 0.7)	0 (0, 0)	0 (0, 0)
	HH member	0.15 (-3.09, 3.17)	0.1 (-2.26, 2.23)	1.12 (-26.16, 29.66)	0.16 (-7.42, 7.49)
	Community member	0.12 (-2.88, 2.91)	0.08 (-1.93, 2.05)	0.71 (-15.48, 18.09)	0.04 (-4.55, 4.53)
2-day staggered grades + stable cohorts	Teachers (all)	0.92 (-3.77, 5.7)	0.61 (-3.26, 4.34)	1.94 (-48.43, 93.02)	0.57 (0, 46.08)
	HighSch teachers	1.47 (-7.55, 11.76)	0.98 (-7.21, 9.71)	1.3 (-192.31, 192.31)	0.86 (0, 0)
	Middle teachers	1.15 (-8.99, 11.36)	0.83 (-6.98, 9.42)	6.8 (-222.22, 224.87)	0.75 (0, 0)
	Elementary teachers	0.58 (-5.31, 6.96)	0.35 (-5.06, 5.24)	0.3 (-88.5, 88.5)	0.38 (0, 0)
	Students (all)	0.78 (-0.7, 2.35)	0.15 (-0.34, 0.66)	-0.01 (0, 0)	0 (0, 0)
	HighSch students	1.1 (-1.2, 3.68)	0.22 (-0.56, 1.15)	0.03 (0, 0)	0 (0, 0)
	Middle students	1.12 (-1.4, 4.37)	0.24 (-0.65, 1.43)	-0.05 (0, 0)	0 (0, 0)
	Elementary students	0.39 (-1.44, 2.28)	0.07 (-0.59, 0.75)	-0.02 (0, 0)	0 (0, 0)
	HH member	0.23 (-2.87, 3.61)	0.14 (-2.08, 2.58)	1.2 (-25.43, 29.16)	0.27 (-7.48, 7.52)
	Community member	0.14 (-2.77, 3.29)	0.09 (-1.98, 2.13)	0.52 (-14.57, 17.2)	0.06 (-4.52, 4.56)
All interventions: 2-day staggered grades, cohorts, masks, monthly testing	Teachers (all)	0.31 (-3.78, 5.56)	0.2 (-3.27, 4.24)	0.65 (-48.6, 90.78)	0.19 (0, 44.97)
	HighSch teachers	0.64 (-7.7, 9.33)	0.43 (-7.35, 7.71)	0.57 (-196.05, 152.6)	0.37 (0, 0)
	Middle teachers	0.51 (-9, 11.1)	0.37 (-6.99, 9.2)	3.02 (-222.56, 219.72)	0.33 (0, 0)
	Elementary teachers	0.09 (-5.26, 6.9)	0.05 (-5.01, 5.19)	0.05 (-87.68, 87.69)	0.06 (0, 0)
	Students (all)	0.27 (-0.76, 1.79)	0.05 (-0.37, 0.5)	0 (0, 0)	0 (0, 0)
	HighSch students	0.41 (-1.4, 2.28)	0.08 (-0.65, 0.71)	0.01 (0, 0)	0 (0, 0)
	Middle students	0.44 (-1.53, 3.37)	0.09 (-0.71, 1.1)	-0.02 (0, 0)	0 (0, 0)
	Elementary students	0.09 (-1.5, 1.69)	0.02 (-0.61, 0.55)	0 (0, 0)	0 (0, 0)
	HH member	0.12 (-2.95, 3.69)	0.07 (-2.14, 2.64)	0.6 (-26.16, 29.83)	0.14 (-7.69, 7.69)
	Community member	0.09 (-2.73, 3.55)	0.06 (-1.95, 2.3)	0.35 (-14.37, 18.59)	0.04 (-4.46, 4.93)

Table S8. Effect of reopening strategies: children equally as susceptible, moderate community transmission.

Excess proportion of infections, symptomatic infections, hospitalizations, or deaths attributable to school re-openings, that would be experienced by each sub-group over a four-month period if schools were allowed to open under certain circumstances compared to if schools remained closed.

Interv.	Subgroup	Excess percent affected, % (95% CI)		Excess rate per 10,000 sub-pop. (95% CI)	
		Infection	Symptomatic infection	Hospitalizations	Deaths
None	Teachers (all)	39.65 (14.35, 54.59)	27.34 (9.8, 39.18)	104.65 (0, 283.02)	8.02 (0, 47.85)
	HighSch teachers	58.94 (3.7, 82.69)	40.74 (1.85, 61.11)	156.86 (0, 576.92)	14.04 (0, 192.31)
	Middle teachers	54.63 (9.09, 77.08)	37.25 (4.65, 58.14)	141.94 (0, 652.17)	11.41 (0, 222.22)
	Elementary teachers	24.94 (6.84, 40.83)	17.3 (4.46, 30.01)	65.94 (0, 260.87)	3.98 (0, 86.21)
	Students (all)	54.39 (21, 70.68)	11.42 (4.47, 15.31)	0.34 (0, 4.85)	0.01 (0, 0.01)
	HighSch students	69.75 (6.16, 90.52)	14.65 (1.26, 20.65)	0.55 (0, 15.11)	0.02 (0, 0.02)
	Middle students	70.63 (10.94, 88.33)	14.83 (2.47, 20.48)	0.54 (0, 18.4)	0 (0, 0)
	Elementary students	36.09 (11.7, 54.17)	7.57 (2.19, 12.07)	0.09 (0, 0.09)	0.01 (0, 0.01)
	HH member	7.84 (2.03, 12.29)	5.22 (1.3, 8.2)	25.4 (0, 57.93)	2.44 (-3.74, 11.12)
	Community member	4.9 (1.11, 8.1)	3.3 (0.78, 5.48)	16.85 (0.9, 34.18)	2.18 (-1.81, 6.39)
Stable cohorts (weak)	Teachers (all)	13.5 (1.44, 26.32)	9.34 (0.94, 19.43)	36.83 (-46.73, 142.87)	2.9 (0, 47.17)
	HighSch teachers	24.14 (-1.85, 53.86)	16.75 (-1.85, 39.62)	67.08 (0, 384.62)	5.08 (0, 181.82)
	Middle teachers	18.19 (-2.22, 43.48)	12.48 (-2.22, 32.61)	50.17 (0, 434.78)	3.99 (0, 3.99)
	Elementary teachers	6.76 (-0.85, 17.24)	4.7 (-0.9, 12.82)	17.64 (-86.21, 170.94)	1.47 (0, 1.47)
	Students (all)	21.38 (3.39, 40.22)	4.49 (0.67, 8.48)	0.18 (0, 4.82)	0 (0, 0)
	HighSch students	32.42 (0.49, 65.59)	6.8 (0.14, 14.42)	0.32 (0, 0.32)	0 (0, 0)
	Middle students	27.16 (1, 58.96)	5.74 (0, 12.86)	0.31 (0, 0.31)	0 (0, 0)
	Elementary students	11.13 (2.01, 24.55)	2.32 (0.3, 5.44)	0.03 (0, 0.03)	0 (0, 0)
	HH member	2.69 (-0.04, 6.09)	1.79 (-0.08, 4.23)	9.44 (-11.09, 34.01)	0.98 (-3.77, 7.47)
	Community member	1.6 (-0.21, 4.04)	1.08 (-0.21, 2.73)	5.65 (-5.43, 18.07)	0.75 (-1.83, 4.5)
Stable cohorts (strong)	Teachers (all)	4.04 (-0.94, 10.8)	2.84 (-0.95, 7.87)	12.53 (-47.39, 94.34)	0.8 (0, 46.51)
	HighSch teachers	6.01 (-3.7, 21.15)	4.16 (-3.64, 15.09)	20.03 (0, 196.08)	1.52 (0, 1.52)
	Middle teachers	5.41 (-4.55, 18.6)	3.79 (-4.35, 13.96)	14.7 (-217.39, 227.27)	1.36 (0, 1.36)
	Elementary teachers	2.61 (-2.56, 8.77)	1.87 (-1.77, 6.96)	8.23 (-86.21, 88.5)	0.26 (0, 0.26)
	Students (all)	7.04 (1.05, 16.24)	1.48 (0.14, 3.55)	0.04 (0, 0.04)	0 (0, 0)
	HighSch students	9.27 (-0.15, 27.21)	1.95 (-0.16, 6.16)	0.04 (0, 0.04)	0 (0, 0)
	Middle students	8.69 (-0.22, 25.36)	1.84 (-0.21, 6.32)	0.06 (0, 0.06)	0 (0, 0)
	Elementary students	4.75 (0.42, 12.15)	1 (-0.1, 2.84)	0.02 (0, 0.02)	0 (0, 0)
	HH member	0.95 (-0.98, 3.19)	0.65 (-0.76, 2.12)	3.61 (-11.35, 21.97)	0.28 (-3.79, 7.24)
	Community member	0.6 (-0.97, 2.23)	0.4 (-0.66, 1.53)	1.89 (-8.13, 11.84)	0.17 (-2.73, 3.6)
2-day half class shifts	Teachers (all)	1.07 (-2.33, 4.74)	0.72 (-1.9, 3.74)	3.85 (-47.39, 48.08)	0 (0, 0)
	HighSch teachers	1.65 (-4.28, 10.22)	1.14 (-3.85, 7.69)	4.72 (0, 188.68)	-0.18 (-0.18, 0)
	Middle teachers	1.29 (-6.38, 9.09)	0.84 (-4.65, 7.14)	4.76 (-217.39, 222.22)	0.49 (0, 0.49)
	Elementary teachers	0.71 (-3.51, 5.26)	0.49 (-2.65, 4.35)	3.12 (-86.96, 87.72)	-0.08 (0, 0.08)
	Students (all)	1.06 (-0.19, 2.9)	0.21 (-0.2, 0.74)	0 (0, 0)	0 (0, 0)
	HighSch students	1.45 (-0.67, 5.06)	0.3 (-0.33, 1.39)	-0.02 (-0.02, 0)	0 (0, 0)
	Middle students	1.25 (-0.86, 4.58)	0.26 (-0.43, 1.3)	0.02 (0, 0.02)	0 (0, 0)
	Elementary students	0.71 (-0.73, 2.47)	0.13 (-0.41, 0.73)	0 (0, 0)	0 (0, 0)
	HH member	0.14 (-1.38, 1.78)	0.11 (-1.04, 1.27)	0.76 (-15.04, 18.19)	0.05 (-3.79, 3.82)
	Community member	0.09 (-1.24, 1.38)	0.06 (-0.89, 0.96)	0.37 (-9.02, 9.91)	0.07 (-2.74, 2.74)
2-day staggered grades	Teachers (all)	1.43 (-1.93, 5.21)	1 (-1.86, 4.27)	3.94 (-47.39, 47.85)	0.19 (0, 0.19)
	HighSch teachers	2.46 (-4, 11.77)	1.72 (-3.85, 9.43)	5.72 (-4.39, 188.68)	0.2 (0, 0.2)
	Middle teachers	2.03 (-4.55, 11.11)	1.42 (-4.44, 8.7)	5.15 (0, 222.22)	0.01 (0, 0.01)
	Elementary teachers	0.72 (-3.48, 5.22)	0.51 (-2.61, 4.35)	2.68 (-86.96, 87.72)	0.26 (0, 0.26)
	Students (all)	2.11 (-0.05, 5.65)	0.44 (-0.1, 1.33)	0.02 (0, 0.02)	0 (0, 0)
	HighSch students	3.29 (-0.61, 11.16)	0.71 (-0.31, 2.6)	0.05 (0, 0.05)	0 (0, 0)
	Middle students	3.03 (-0.83, 10.37)	0.62 (-0.43, 2.48)	0 (0, 0)	0 (0, 0)
	Elementary students	0.88 (-0.72, 3.11)	0.19 (-0.32, 0.92)	0.01 (0, 0)	0 (0, 0)
	HH member	0.34 (-1.2, 1.95)	0.22 (-0.89, 1.44)	1.4 (-14.86, 18.67)	0.12 (-3.78, 7.24)
	Community member	0.21 (-1.05, 1.53)	0.15 (-0.75, 1.1)	0.59 (-8.28, 10.01)	0.07 (-2.73, 2.74)

Masks	Teachers (all)	8.87 (0, 18.98)	6.16 (-0.47, 13.62)	23.66 (-47.18, 138.25)	1.98 (0, 46.95)
	HighSch teachers	21.34 (-1.82, 50)	14.83 (-1.85, 36.37)	59.05 (0, 377.36)	3.99 (0, 4.46)
	Middle teachers	10.75 (-2.27, 30.96)	7.47 (-2.27, 23.4)	25.59 (-217.39, 232.56)	2.74 (0, 2.74)
	Elementary teachers	2.41 (-2.59, 8.55)	1.66 (-2.56, 6.19)	6.61 (-86.96, 88.5)	0.8 (0, 0.08)
	Students (all)	18.8 (1.52, 34.78)	3.93 (0.29, 7.49)	0.12 (0, 0.12)	0 (0, 0)
	HighSch students	39.47 (0.62, 74.24)	8.25 (0, 16.56)	0.25 (0, 0.25)	0 (0, 0)
	Middle students	21.84 (0.83, 51.01)	4.58 (0, 11.39)	0.19 (0, 0.19)	0 (0, 0)
	Elementary students	3.63 (0, 8.59)	0.76 (-0.2, 2)	0 (0, 0)	0 (0, 0)
	HH member	2.34 (-0.6, 5.87)	1.58 (-0.51, 3.95)	8.04 (-11.26, 32.59)	0.74 (-3.75, 7.45)
	Community member	1.45 (-0.47, 3.91)	0.97 (-0.35, 2.62)	4.74 (-6.34, 18.24)	0.66 (-2.72, 4.52)
Monthly testing w/ case isolation & class quarantine	Teachers (all)	37.77 (10.64, 53.31)	26.04 (7.4, 38.14)	162.47 (0, 588.24)	8.12 (0, 47.85)
	HighSch teachers	57.51 (3.84, 83.33)	39.67 (1.92, 62.75)	138.26 (0, 666.67)	13.98 (0, 192.31)
	Middle teachers	52.36 (6.65, 76.09)	35.81 (2.32, 55.56)	60.91 (-84.75, 260.87)	12.34 (0, 222.22)
	Elementary teachers	23.02 (5.92, 39.32)	15.96 (3.54, 28.7)	0.33 (0, 4.85)	3.82 (0, 86.23)
	Students (all)	52.07 (16.82, 69.12)	10.94 (3.28, 14.98)	0.58 (0, 15.27)	0 (0, 0)
	HighSch students	68.11 (4.74, 90.11)	14.3 (0.92, 20.63)	0.54 (0, 19.2)	0.02 (0, 0.02)
	Middle students	68.47 (10.59, 87.71)	14.4 (2.07, 20.73)	0.05 (0, 0.05)	0 (0, 0)
	Elementary students	33.26 (9.67, 51.75)	6.99 (1.86, 11.39)	23.3 (-3.68, 54.99)	0 (0, 0)
	HH member	7.31 (1.36, 11.97)	4.87 (0.8, 7.89)	15.48 (0, 32.3)	2.42 (-3.74, 11.05)
	Community member	4.53 (0.5, 7.85)	3.05 (0.34, 5.32)	1.96 (-47.39, 47.62)	1.99 (-1.81, 6.37)
Stable cohorts (strong), masks, monthly testing	Teachers (all)	0.69 (-2.35, 4.23)	0.49 (-2.33, 3.3)	1.96 (-47.39, 47.62)	0.05 (0, 0.05)
	HighSch teachers	1.14 (-5.46, 7.69)	0.78 (-3.92, 5.88)	4.5 (0, 188.68)	0.59 (0, 0.59)
	Middle teachers	0.86 (-6.52, 8.89)	0.56 (-4.55, 6.67)	0.49 (-217.39, 222.22)	0 (0, 0)
	Elementary teachers	0.41 (-3.51, 4.52)	0.32 (-2.75, 3.51)	1.38 (-86.96, 86.96)	-0.16 (-0.16, 0)
	Students (all)	1.71 (-0.14, 4.86)	0.36 (-0.15, 1.09)	0 (0, 0)	0 (0, 0)
	HighSch students	2.86 (-0.65, 10.47)	0.59 (-0.32, 2.41)	0 (0, 0)	0 (0, 0)
	Middle students	1.99 (-0.91, 7.16)	0.43 (-0.43, 1.85)	0 (0, 0)	0 (0, 0)
	Elementary students	0.82 (-0.75, 2.87)	0.17 (-0.41, 0.79)	0 (0, 0)	0 (0, 0)
	HH member	0.27 (-1.34, 1.97)	0.19 (-0.92, 1.42)	1.07 (-15.03, 18.48)	0.01 (-3.77, 3.79)
	Community member	0.18 (-1.13, 1.63)	0.12 (-0.8, 1.17)	0.52 (-8.19, 9.9)	0.11 (-2.75, 3.61)
2-day half class shifts + stable cohorts	Teachers (all)	0.38 (-2.76, 3.29)	0.24 (-1.93, 2.82)	1.18 (-47.39, 47.62)	-0.33 (-0.33, 0)
	HighSch teachers	0.4 (-5.56, 5.88)	0.19 (-4.08, 5.56)	2.42 (-4.31, 185.19)	-0.02 (-0.02, 0)
	Middle teachers	0.45 (-6.38, 6.82)	0.34 (-4.76, 6.52)	-1.06 (-217.39, 212.88)	-0.42 (-0.42, 0)
	Elementary teachers	0.34 (-3.57, 4.39)	0.23 (-3.39, 4.13)	1.47 (-86.96, 86.96)	-0.43 (-0.43, 0)
	Students (all)	0.37 (-0.55, 1.45)	0.08 (-0.25, 0.43)	0 (0, 0)	0 (0, 0)
	HighSch students	0.39 (-1.1, 2.28)	0.08 (-0.47, 0.79)	-0.02 (-0.02, 0)	0 (0, 0)
	Middle students	0.4 (-1.23, 2.36)	0.09 (-0.44, 0.82)	0 (0, 0)	0 (0, 0)
	Elementary students	0.35 (-0.88, 1.75)	0.07 (-0.42, 0.61)	0.01 (0, 0.01)	0 (0, 0)
	HH member	0.07 (-1.44, 1.58)	0.06 (-1.11, 1.26)	0.41 (-15.52, 17.68)	-0.03 (-3.8, 3.8)
	Community member	0.05 (-1.24, 1.35)	0.03 (-0.87, 0.97)	0.08 (-9.08, 9.97)	0 (-2.74, 2.74)
2-day staggered grades + stable cohorts	Teachers (all)	0.49 (-2.34, 3.38)	0.36 (-1.9, 2.83)	2.45 (-47.39, 47.85)	0.28 (0, 0.28)
	HighSch teachers	0.82 (-5.56, 7.55)	0.61 (-3.92, 5.66)	3.76 (0, 188.68)	0.57 (0, 0.57)
	Middle teachers	0.78 (-6.38, 8.52)	0.57 (-4.55, 6.67)	2.68 (0, 217.39)	0.45 (0, 0.45)
	Elementary teachers	0.22 (-3.62, 4.31)	0.16 (-3.45, 3.51)	1.73 (-86.96, 87.72)	0.08 (0, 0.08)
	Students (all)	0.85 (-0.34, 2.43)	0.17 (-0.19, 0.65)	0 (0, 0)	0 (0, 0)
	HighSch students	1.17 (-0.83, 4.41)	0.25 (-0.34, 1.13)	0.02 (0, 0.02)	0 (0, 0)
	Middle students	1.21 (-1.03, 4.8)	0.24 (-0.6, 1.27)	-0.02 (-0.02, 0)	0 (0, 0)
	Elementary students	0.45 (-0.82, 2.11)	0.09 (-0.4, 0.62)	0 (0, 0)	0 (0, 0)
	HH member	0.14 (-1.41, 1.65)	0.09 (-1.06, 1.19)	0.92 (-14.79, 18.36)	0.05 (-3.8, 3.8)
	Community member	0.07 (-1.2, 1.32)	0.04 (-0.86, 0.9)	0.26 (-8.24, 9.1)	0.04 (-2.73, 2.73)
All interventions: 2-day staggered grades, cohorts, masks, monthly testing	Teachers (all)	0.32 (-3.7, 4.21)	0.07 (-2.28, 2.33)	-0.61 (-47.62, 47.62)	0.47 (0, 0.5)
	HighSch teachers	0.07 (-2.8, 2.79)	0.15 (-3.92, 4.28)	-0.6 (-0.6, 0)	0.19 (0, 0.2)
	Middle teachers	0.08 (-5.66, 5.77)	0.08 (-4.76, 4.65)	-1.36 (-217.39, 0)	0.22 (0, 0.3)
	Elementary teachers	0.13 (-6.52, 6.67)	0.04 (-2.65, 3.39)	-0.35 (-86.96, 86.96)	0.7 (0, 0.7)
	Students (all)	0.04 (-3.54, 3.54)	0.06 (-0.24, 0.43)	0.01 (0, 0.01)	0 (0, 0)
	HighSch students	0.27 (-0.62, 1.3)	0.08 (-0.48, 0.76)	0.03 (0, 0.02)	0 (0, 0)
	Middle students	0.41 (-1.11, 2.39)	0.09 (-0.45, 0.89)	0 (0, 0)	0 (0, 0)
	Elementary students	0.39 (-1.11, 2.42)	0.02 (-0.44, 0.53)	0 (0, 0)	0 (0, 0)
	HH member	0.12 (-0.99, 1.34)	0.03 (-1.04, 1.18)	0.15 (-18.26, 15.22)	0.05 (-3.79, 3.83)
	Community member	0.04 (-1.4, 1.55)	0 (-0.85, 0.97)	0.13 (-8.3, 9.09)	0.02 (-2.72, 2.72)

Table S9. Effect of reopening strategies: children equally as susceptible, high community transmission.

Excess proportion of infections, symptomatic infections, hospitalizations, or deaths attributable to school re-openings, that would be experienced by each sub-group over a four-month period if schools were allowed to open under certain circumstances compared to if schools remained closed.

Interv.	Subgroup	Excess percent affected, % (95% CI)		Excess rate per 10,000 sub-pop. (95% CI)	
		Infection	Symptomatic infection	Hospitalizations	Deaths
None	Teachers (all)	46.81 (26.94, 59.43)	32.16 (18.05, 42.59)	119.83 (0, 285.75)	9.39 (0, 47.85)
	HighSch teachers	63.32 (18, 83.93)	43.55 (12.96, 62.51)	159.84 (0, 576.92)	12.38 (0, 192.31)
	Middle teachers	59.55 (19.15, 80)	40.85 (12.76, 58.7)	159.64 (0, 666.67)	11.78 (0, 222.22)
	Elementary teachers	34.24 (17.94, 47.86)	23.53 (11.71, 34.49)	85.71 (-84.75, 333.47)	7.13 (0, 87.72)
	Students (all)	61.92 (36.75, 73.92)	13 (7.69, 16.16)	0.37 (0, 4.86)	0.01 (0, 0.01)
	HighSch students	74.17 (21.16, 90.05)	15.62 (4, 20.55)	0.58 (0, 15.36)	0 (0, 0)
	Middle students	75.05 (27.13, 88.76)	15.68 (5.31, 20.61)	0.59 (0, 19.27)	0.04 (0, 0.04)
	Elementary students	47.25 (27.47, 60.33)	9.92 (5.63, 13.59)	0.11 (0, 0.11)	0 (0, 0)
	HH member	12.02 (5.25, 16.85)	7.97 (3.48, 11.33)	38.49 (3.75, 74.46)	3.92 (-3.75, 14.82)
	Community member	9.47 (3.93, 13.35)	6.38 (2.78, 9.13)	32.53 (11.71, 52.83)	4.44 (-0.91, 10.04)
Stable cohorts (weak)	Teachers (all)	22.28 (6.1, 35.24)	15.31 (3.72, 24.89)	58.21 (-46.73, 189.57)	3.99 (0, 47.39)
	HighSch teachers	34.22 (1.92, 59.26)	23.73 (0, 44.23)	87.61 (0, 392.16)	5.5 (0, 185.19)
	Middle teachers	28.88 (2.27, 52.4)	19.88 (0, 38.31)	76.09 (0, 454.55)	6.11 (0, 217.39)
	Elementary teachers	14.2 (2.7, 26.72)	9.66 (0.87, 18.59)	37.64 (-86.96, 178.57)	2.51 (0, 85.47)
	Students (all)	33.75 (11.89, 49.03)	7.08 (2.31, 10.59)	0.25 (0, 4.83)	0 (0, 0)
	HighSch students	45.21 (4.89, 70.06)	9.43 (0.91, 15.86)	0.48 (0, 14.62)	0 (0, 0)
	Middle students	42.2 (6.93, 65.67)	8.88 (1.09, 14.76)	0.33 (0, 0.33)	0 (0, 0)
	Elementary students	21.89 (8.81, 35.15)	4.62 (1.5, 7.64)	0.05 (0, 0.05)	0 (0, 0)
	HH member	6.21 (1.18, 10.9)	4.12 (0.7, 7.39)	20.15 (-7.63, 55.17)	1.98 (-3.94, 11.14)
	Community member	4.91 (0.63, 8.84)	3.3 (0.5, 6.01)	16.72 (-0.92, 34.89)	2.09 (-2.73, 7.38)
Stable cohorts (strong)	Teachers (all)	7.71 (0.47, 16.36)	5.26 (-0.47, 11.59)	19.08 (-47.62, 96.15)	1.5 (0, 46.73)
	HighSch teachers	11.35 (-2, 29.42)	7.77 (-2, 21.16)	28.9 (-181.82, 200)	2.07 (0, 2.07)
	Middle teachers	9.52 (-2.27, 26.68)	6.53 (-2.33, 20.01)	28.26 (0, 232.56)	1.53 (0, 1.53)
	Elementary teachers	5.34 (-1.75, 13.56)	3.61 (-1.77, 10.26)	10.92 (-87.72, 90.09)	1.21 (0, 1.21)
	Students (all)	12.65 (2.99, 23.01)	2.64 (0.56, 5.17)	0.07 (0, 0.07)	0 (0, 0)
	HighSch students	16.38 (0.9, 37.02)	3.41 (0, 8.13)	0.19 (0, 0.19)	0 (0, 0)
	Middle students	15.36 (0.8, 34.32)	3.19 (0, 7.79)	0.02 (0, 0.02)	0 (0, 0)
	Elementary students	8.82 (1.53, 17.24)	1.85 (0.11, 4.09)	0.02 (0, 0.02)	0 (0, 0)
	HH member	2.23 (-0.66, 5.45)	1.47 (-0.52, 3.67)	7.4 (-14.8, 30.64)	0.81 (-7.21, 7.53)
	Community member	1.75 (-0.77, 4.35)	1.18 (-0.58, 3)	6.06 (-7.21, 20.14)	0.85 (-2.75, 5.42)
2-day half class shifts	Teachers (all)	2.35 (-1.88, 7.11)	1.56 (-1.9, 5.26)	5.61 (-47.62, 93.47)	1.03 (0, 46.08)
	HighSch teachers	3.57 (-5.36, 14.01)	2.35 (-5.09, 10.72)	7.3 (-181.82, 192.31)	0.57 (0, 0.57)
	Middle teachers	2.56 (-6.67, 12.77)	1.73 (-4.88, 9.3)	10.29 (-204.19, 227.27)	2.2 (0, 2.2)
	Elementary teachers	1.71 (-3.48, 7.08)	1.13 (-3.39, 5.26)	3.04 (-87.72, 88.5)	0.78 (0, 0.78)
	Students (all)	1.98 (-0.19, 4.65)	0.41 (-0.23, 1.18)	0 (0, 0)	0 (0, 0)
	HighSch students	2.63 (-0.82, 7.74)	0.55 (-0.49, 1.96)	0.02 (0, 0.02)	0 (0, 0)
	Middle students	2.32 (-1.05, 7.04)	0.49 (-0.65, 1.92)	0 (0, 0)	0 (0, 0)
	Elementary students	1.38 (-0.86, 3.76)	0.28 (-0.49, 1.15)	0 (0, 0)	0 (0, 0)
	HH member	0.41 (-1.87, 2.68)	0.29 (-1.31, 1.94)	1.47 (-18.65, 22.2)	0.09 (-7.37, 7.38)
	Community member	0.34 (-1.8, 2.45)	0.23 (-1.24, 1.74)	1.2 (-10.94, 12.67)	0.23 (-3.64, 3.66)
2-day staggered grades	Teachers (all)	3.35 (-1.88, 9.26)	2.34 (-1.88, 7.14)	7.8 (-48.08, 94.34)	0.85 (0, 46.73)
	HighSch teachers	5.92 (-5.26, 19.64)	4.18 (-3.85, 14.81)	15.3 (-188.68, 196.08)	0.81 (0, 0.81)
	Middle teachers	4.56 (-6.52, 17.78)	3.15 (-4.65, 13.64)	11.94 (-217.39, 232.56)	1.56 (0, 1.56)
	Elementary teachers	1.7 (-4.31, 8.62)	1.19 (-3.45, 6.2)	2.83 (-87.72, 88.5)	0.63 (0, 0.63)
	Students (all)	4.88 (0.59, 10.04)	1.02 (0, 2.3)	0.05 (0, 0.05)	0 (0, 0)
	HighSch students	7.33 (0, 18.82)	1.53 (-0.18, 4.4)	0.06 (0, 0.06)	0 (0, 0)
	Middle students	6.73 (-0.19, 15.63)	1.41 (-0.42, 3.68)	0.08 (0, 0.08)	0 (0, 0)
	Elementary students	2.33 (-0.65, 5.99)	0.48 (-0.42, 1.51)	0.02 (0, 0.02)	0 (0, 0)
	HH member	1.07 (-1.7, 3.79)	0.72 (-1.08, 2.73)	2.99 (-21.93, 29.39)	0.47 (-7.34, 7.5)
	Community member	0.86 (-1.62, 3.47)	0.58 (-1.18, 2.34)	2.49 (-12.63, 17.28)	0.38 (-3.67, 4.57)

Masks	Teachers (all)	13.58 (2.38, 24.64)	9.33 (1.42, 17.21)	33.77 (-47.17, 142.86)	2.63 (0, 46.95)
	HighSch teachers	28.25 (1.82, 51.92)	19.5 (0, 38.18)	70 (0, 384.62)	4.52 (0, 4.52)
	Middle teachers	16.76 (0, 39.13)	11.58 (-2.13, 28.27)	45.05 (0, 260.2)	1.8 (0, 1.8)
	Elementary teachers	5.6 (-0.88, 13.68)	3.78 (-1.71, 10.34)	12.64 (-87.72, 91.26)	2.08 (0, 84.75)
	Students (all)	25.95 (4.89, 40.5)	5.43 (1.01, 8.9)	0.17 (0, 4.75)	0 (0, 0)
	HighSch students	49.56 (3.23, 76.01)	10.39 (0.8, 16.91)	0.33 (0, 0.33)	0 (0, 0)
	Middle students	31.72 (3.4, 57.43)	6.6 (0.42, 12.84)	0.21 (0, 0.21)	0.02 (0, 0.02)
	Elementary students	7.47 (1.86, 13.7)	1.57 (0.21, 3.33)	0.05 (0, 0.05)	0 (0, 0)
	HH member	4.51 (0.25, 8.8)	3.01 (0.04, 5.93)	14.86 (-11.08, 44.85)	1.53 (-3.88, 11.04)
	Community member	3.55 (0, 7.28)	2.39 (-0.06, 4.85)	12.4 (-3.61, 29.2)	1.71 (-2.7, 7.17)
Monthly testing w/ case isolation & class quarantine	Teachers (all)	45.53 (24.07, 58.02)	31.37 (15.96, 41.4)	116.18 (0, 284.39)	9.67 (0, 47.85)
	HighSch teachers	62.92 (16.98, 82.69)	43.26 (10.9, 61.11)	158.2 (0, 576.92)	13.09 (0, 192.31)
	Middle teachers	58.63 (19.13, 80.43)	40.45 (10.64, 59.09)	153.32 (0, 666.67)	11.12 (0, 222.22)
	Elementary teachers	32.42 (13.63, 47.37)	22.35 (9.31, 34.19)	82.27 (-84.75, 265.49)	7.57 (0, 87.72)
	Students (all)	60.35 (31, 72.52)	12.67 (6.98, 15.89)	0.33 (0, 4.84)	0 (0, 0)
	HighSch students	73.13 (19.49, 89.55)	15.39 (3.76, 20.29)	0.53 (0, 15.2)	0 (0, 0)
	Middle students	74.16 (28.31, 87.63)	15.54 (5.97, 20.48)	0.55 (0, 0.55)	0.02 (0, 0.02)
	Elementary students	44.97 (22.03, 58.52)	9.43 (4.46, 13.1)	0.08 (0, 0.08)	0 (0, 0)
	HH member	11.5 (4.09, 16.22)	7.63 (2.68, 11.01)	37.35 (6.97, 72.22)	3.64 (-3.77, 14.79)
	Community member	9.02 (2.92, 12.92)	6.08 (1.94, 8.75)	31.08 (9.09, 50.67)	4.14 (-0.92, 9.98)
Stable cohorts (strong), masks, monthly testing	Teachers (all)	1.45 (-2.36, 5.69)	0.97 (-2.29, 4.29)	3.74 (-47.85, 92.6)	0.57 (0, 0.57)
	HighSch teachers	2.3 (-5.66, 11.11)	1.6 (-5.46, 9.26)	5.36 (-185.19, 192.31)	-0.19 (-0.19, 0)
	Middle teachers	1.65 (-6.82, 11.11)	1.19 (-6.52, 8.89)	6.75 (-217.39, 222.35)	0.89 (0, 0.89)
	Elementary teachers	0.98 (-3.54, 6.14)	0.59 (-3.45, 5.17)	1.84 (-87.72, 88.5)	0.78 (0, 0.78)
	Students (all)	3.18 (0.2, 7.16)	0.67 (-0.05, 1.66)	0.03 (0, 0.03)	0 (0, 0)
	HighSch students	5.13 (-0.33, 13.71)	1.07 (-0.32, 3.3)	0.06 (0, 0.06)	0 (0, 0)
	Middle students	3.76 (-0.79, 10.31)	0.8 (-0.45, 2.65)	0.02 (0, 0.02)	0 (0, 0)
	Elementary students	1.6 (-0.72, 4.27)	0.35 (-0.42, 1.22)	0.02 (0, 0.02)	0 (0, 0)
	HH member	0.6 (-1.73, 2.95)	0.39 (-1.32, 2.03)	1.99 (-18.52, 22.91)	0.28 (-7.34, 7.44)
	Community member	0.48 (-1.51, 2.82)	0.33 (-1.02, 1.95)	1.96 (-10.22, 14.52)	0.29 (-3.61, 3.65)
2-day half class shifts + stable cohorts	Teachers (all)	0.82 (-2.87, 4.74)	0.56 (-2.76, 3.85)	1.4 (-47.62, 48.08)	0.51 (0, 0.51)
	HighSch teachers	0.96 (-5.89, 9.1)	0.62 (-5.56, 7.55)	0.06 (-185.19, 185.19)	-0.01 (-0.01, 0)
	Middle teachers	0.84 (-6.98, 10.64)	0.62 (-6.67, 8.33)	4.7 (-217.39, 222.22)	0.67 (0, 0.67)
	Elementary teachers	0.75 (-4.27, 6.09)	0.5 (-3.48, 5.17)	0.72 (-87.72, 88.5)	0.69 (0, 0.69)
	Students (all)	0.69 (-0.83, 2.38)	0.14 (-0.34, 0.65)	0 (0, 0)	0 (0, 0)
	HighSch students	0.7 (-1.41, 3.33)	0.15 (-0.62, 1.04)	0.02 (0, 0.02)	0 (0, 0)
	Middle students	0.8 (-1.51, 3.76)	0.17 (-0.68, 1.24)	0 (0, 0)	0 (0, 0)
	Elementary students	0.63 (-1.16, 2.72)	0.12 (-0.53, 0.84)	0 (0, 0)	0 (0, 0)
	HH member	0.17 (-1.97, 2.3)	0.11 (-1.38, 1.61)	0.59 (-18.77, 21.36)	0.05 (-7.34, 7.33)
	Community member	0.15 (-1.76, 2.25)	0.1 (-1.28, 1.56)	0.71 (-11.86, 12.72)	0.14 (-3.64, 3.65)
2-day staggered grades + stable cohorts	Teachers (all)	1.48 (-3.21, 6.25)	1 (-2.75, 5.16)	4.17 (-47.85, 93.47)	0.15 (0, 0.15)
	HighSch teachers	2.56 (-6, 12.73)	1.73 (-4.11, 9.44)	7.63 (-188.68, 192.31)	-0.36 (-0.36, 0)
	Middle teachers	2.23 (-6.82, 13.05)	1.58 (-6.52, 10.87)	4.41 (-217.39, 227.27)	0.49 (0, 0.49)
	Elementary teachers	0.69 (-5.22, 6.9)	0.45 (-4.35, 5.31)	2.55 (-87.72, 88.5)	0.27 (0, 0.27)
	Students (all)	2.27 (-0.1, 5.17)	0.46 (-0.24, 1.28)	0 (0, 0)	0 (0, 0)
	HighSch students	3.07 (-0.76, 8.43)	0.62 (-0.5, 2.09)	0.03 (0, 0.03)	0 (0, 0)
	Middle students	3.23 (-0.67, 8.99)	0.67 (-0.61, 2.4)	-0.02 (-0.02, 0)	0 (0, 0)
	Elementary students	1.25 (-1.11, 3.98)	0.25 (-0.54, 1.16)	0 (0, 0)	0 (0, 0)
	HH member	0.53 (-2.13, 3.32)	0.35 (-1.58, 2.26)	1.32 (-21.93, 25.03)	0.06 (-7.36, 7.37)
	Community member	0.41 (-2.06, 2.92)	0.28 (-1.46, 2.02)	1.43 (-13.44, 15.47)	0.15 (-3.67, 4.5)
All interventions: 2-day staggered grades, cohorts, masks, monthly testing	Teachers (all)	0.32 (-3.7, 4.21)	0.24 (-2.89, 3.7)	0.33 (-48.08, 48.31)	0.23 (0, 0.3)
	HighSch teachers	0.56 (-6.12, 7.69)	0.53 (-5.77, 7.41)	0.67 (-188.68, 188.68)	0.55 (0, 0.6)
	Middle teachers	0.36 (-8.7, 9.09)	0.26 (-6.82, 6.98)	1.92 (-222.22, 222.22)	0.19 (0, 0.2)
	Elementary teachers	0.19 (-5.26, 5.31)	0.1 (-4.35, 4.39)	-0.47 (-86.96, 88.5)	0.1 (0, 0.1)
	Students (all)	0.72 (-0.92, 2.57)	0.15 (-0.39, 0.71)	0.01 (0, 0.01)	0 (0, 0)
	HighSch students	1.1 (-1.43, 4.57)	0.23 (-0.65, 1.38)	0.03 (0, 0.03)	0 (0, 0)
	Middle students	0.94 (-1.79, 4.18)	0.2 (-0.84, 1.41)	0 (0, 0)	0 (0, 0)
	Elementary students	0.36 (-1.69, 2.3)	0.08 (-0.65, 0.79)	0 (0, 0)	0 (0, 0)
	HH member	0.25 (-2.37, 2.99)	0.16 (-1.61, 1.98)	1.15 (-21.63, 25.38)	0.08 (-7.29, 7.35)
	Community member	0.18 (-2.11, 2.65)	0.11 (-1.54, 1.88)	0.47 (-11.89, 13.64)	0.09 (-3.7, 3.72)

Table S10. The number of excess student cases attributable to school transmission expected across a four-month semester, for 50% community vaccination coverage.

	Excess student cases attributable to within-school transmission within:			
	380-person **elementary schools** (*half susceptibility*)	380-person **elementary schools** (*equal susceptibility*)	420-person **middle schools**	620-person **high schools**
No precautions	45 cases per school	59 cases per school	55 cases per school	81 cases per school
Masks	22 cases per school	44 cases per school	31 cases per school	35 cases per school
Masks + testing	15 cases per school	38 cases per school	24 cases per school	24 cases per school
Masks + cohorts	4 cases per school	15 cases per school	6 cases per school	6 cases per school

Table S11. The number of excess student cases attributable to school transmission expected across a four-month semester, for 60% community vaccination coverage.

	Excess student cases attributable to within-school transmission within:			
	380-person **elementary schools** (*half susceptibility*)	380-person **elementary schools** (*equal susceptibility*)	420-person **middle schools**	620-person **high schools**
No precautions	40 cases per school	56 cases per school	49 cases per school	70 cases per school
Masks	17 cases per school	40 cases per school	24 cases per school	21 cases per school
Masks + testing	10 cases per school	33 cases per school	16 cases per school	13 cases per school
Masks + cohorts	3 cases per school	12 cases per school	4 cases per school	3 cases per school

Table S12. The minimum non-pharmaceutical intervention needed to reduce the risk of symptomatic infection to beneath a given threshold (e.g., 50 cases per 1,000 population), assuming that 50% of the vaccine-eligible community has received a vaccine. 'Not observed' indicates that no combination of interventions examined in this study reduced excess risk beneath the indicated threshold.

		Threshold - symptomatic cases per 1,000 population			< 2 cases per school*
		<50	<25	<10	
Students	Elementary school – *half susceptibility*	Masks + testing	Masks + cohorts	Not observed**	Not observed**
	Elementary school – *equal susceptibility*	Masks + cohorts	Not observed**	Not observed**	Not observed**
	Middle school	Masks + cohorts	Masks + cohorts	Not observed**	Not observed**
	High school	Masks + testing	Masks + cohorts	Masks + cohorts	Not observed**
Teachers	Elementary school – *half susceptibility*	Masks + testing	Masks + cohorts	Not observed**	
	Elementary school – *equal susceptibility*	Masks + cohorts	Not observed**	Not observed**	
	Middle school	Masks + cohorts	Masks + cohorts	Not observed**	
	High school	Masks + testing	Masks + cohorts	Masks + cohorts	

*Assuming a 380-person elementary school, 420-person middle school, and 680-person high school
**not observed under the specific combination of interventions simulated

Table S13. The minimum non-pharmaceutical intervention to reduce the risk of symptomatic infection to beneath a given threshold (e.g., 50 cases per 1,000 population), assuming that 60% of the vaccine-eligible community has received a vaccine. 'Not observed' indicates that no combination of interventions examined in this study reduced excess risk beneath the indicated threshold.

		Threshold - symptomatic cases per 1,000 population			< 2 cases per school*
		<50	<25	<10	
Students	Elementary school – *half susceptibility*	Masks	Masks + cohorts	Masks + cohorts	Not observed**
	Elementary school – *equal susceptibility*	Masks + cohorts	Not observed**	Not observed**	Not observed**
	Middle school	Masks + testing	Masks + cohorts	Masks + cohorts	Not observed**
	High school	Masks	Masks + testing	Masks + cohorts	Not observed**
Teachers	Elementary school – *half susceptibility*	Masks	Masks + cohorts	Masks + cohorts	
	Elementary school – *equal susceptibility*	Masks + cohorts	Masks + cohorts	Not observed**	
	Middle school	Masks	Masks + cohorts	Masks + cohorts	
	High school	Masks	Masks + testing	Masks + cohorts	

*Assuming a 380-person elementary school, 420-person middle school, and 680-person high school
**not observed under the specific combination of interventions simulated

Table S14. Excess *symptomatic* infections attributable to school transmission by population subgroup and scenario examined – Delta variant and vaccination available

Vaccine coverage (%)	NPI	Susceptibility of children <10 years	Population	Mean (89% HPDI) excess infections per 100 population	Median excess infections per 100 population	Infections per school*
50	None	Equal	Elem. student	15.5 (10.8, 20.3)	15.7	59
50	None	Half	Elem. student	11.9 (9, 15.6)	12	45
50	None		Middle sch student	13.1 (8.8, 17.9)	13.2	55
50	None		High sch student	12 (8.6, 15.7)	12	81
50	None	Equal	Elem. teacher	21.4 (6, 37.8)	20.4	
50	None	Half	Elem. teacher	18.1 (4.2, 31.9)	17	
50	None		Middle sch teacher	23.3 (2.2, 44.2)	22	
50	None		High sch teacher	24.1 (3.5, 47.3)	21.6	
50	Masks	Equal	Elem. student	11.7 (5.9, 17.7)	11.2	44
50	Masks	Half	Elem. student	5.7 (1.8, 9.4)	5.8	22
50	Masks		Middle sch student	7.5 (1.8, 13.3)	7.1	31
50	Masks		High sch student	5.1 (1.3, 8.6)	5.1	35
50	Masks	Equal	Elem. teacher	11.8 (1.7, 22.2)	10.7	
50	Masks	Half	Elem. teacher	6.4 (-0.9, 13.8)	6	
50	Masks		Middle sch teacher	10 (-2.3, 23.4)	8.7	
50	Masks		High sch teacher	7.6 (-2, 18.2)	6.6	
50	Masks+Testing	Equal	Elem. student	10 (4.2, 16.8)	9.2	38
50	Masks+Testing	Half	Elem. student	3.9 (0.2, 7.1)	3.9	15
50	Masks+Testing		Middle sch student	5.7 (0, 10.8)	5.1	24
50	Masks+Testing		High sch student	3.5 (0, 6.6)	3.4	24
50	Masks+Testing	Equal	Elem. teacher	9.4 (0, 19.3)	8.5	
50	Masks+Testing	Half	Elem. teacher	4.3 (-2.7, 10.3)	3.5	
50	Masks+Testing		Middle sch teacher	7.3 (-2.4, 20)	6.5	
50	Masks+Testing		High sch teacher	5.2 (-2, 15.1)	3.8	
50	Masks+Cohorts	Equal	Elem. student	4 (0, 8.3)	3.5	15
50	Masks+Cohorts	Half	Elem. student	1.1 (-0.3, 2.7)	1.1	4
50	Masks+Cohorts		Middle sch student	1.5 (-0.7, 3.8)	1.3	6
50	Masks+Cohorts		High sch student	0.8 (-0.5, 2.5)	0.7	6
50	Masks+Cohorts	Equal	Elem. teacher	2.9 (-1.8, 7.6)	2.6	
50	Masks+Cohorts	Half	Elem. teacher	1 (-2.7, 4.4)	0.9	
50	Masks+Cohorts		Middle sch teacher	1.4 (-4.7, 6.8)	2.1	
50	Masks+Cohorts		High sch teacher	1 (-4, 5.8)	0.9	
60	None	Equal	Elem. student	14.7 (9.3, 20.4)	14.4	56
60	None	Half	Elem. student	10.6 (6.8, 15.1)	10.5	40
60	None		Middle sch student	11.7 (6.5, 16.9)	11.6	49
60	None		High sch student	10.4 (6.8, 14.3)	10.5	70
60	None	Equal	Elem. teacher	18 (3.4, 32.8)	17.2	

60	None	Half	Elem. teacher	14.4 (2.6, 27.6)	12.2	
60	None		Middle sch teacher	20 (0, 40)	17.6	
60	None		High sch teacher	20 (0, 40)	15.4	
60	Masks	Equal	Elem. student	10.6 (4.7, 17.5)	9.2	40
60	Masks	Half	Elem. student	4.5 (0.5, 8.3)	4.3	17
60	Masks		Middle sch student	5.8 (0.4, 11.1)	5.2	24
60	Masks		High sch student	3.2 (0.1, 5.9)	3.1	21
60	Masks	Equal	Elem. teacher	9.2 (0.9, 18.4)	8	
60	Masks	Half	Elem. teacher	4.3 (-0.9, 11)	3.5	
60	Masks		Middle sch teacher	6.8 (-2.3, 17.8)	4.7	
60	Masks		High sch teacher	4.3 (-2, 11.5)	3.8	
60	Masks+Testing	Equal	Elem. student	8.6 (2.3, 16.6)	7.2	33
60	Masks+Testing	Half	Elem. student	2.8 (-0.4, 6.2)	2.4	10
60	Masks+Testing		Middle sch student	3.8 (-0.7, 9.1)	2.9	16
60	Masks+Testing		High sch student	1.9 (-0.4, 4.4)	1.6	13
60	Masks+Testing	Equal	Elem. teacher	6.9 (-0.9, 15.3)	5.4	
60	Masks+Testing	Half	Elem. teacher	2.6 (-1.8, 7.6)	1.8	
60	Masks+Testing		Middle sch teacher	4.5 (-2.4, 15.2)	2.3	
60	Masks+Testing		High sch teacher	2.4 (-2, 9.4)	1.9	
60	Masks+Cohorts	Equal	Elem. student	3.1 (-0.4, 6.9)	2.5	12
60	Masks+Cohorts	Half	Elem. student	0.8 (-0.4, 2)	0.6	3
60	Masks+Cohorts		Middle sch student	0.9 (-0.7, 2.6)	0.7	4
60	Masks+Cohorts		High sch student	0.4 (-0.6, 1.5)	0.3	3
60	Masks+Cohorts	Equal	Elem. teacher	1.9 (-0.9, 6.1)	1.7	
60	Masks+Cohorts	Half	Elem. teacher	0.7 (-1.8, 3.5)	0.8	
60	Masks+Cohorts		Middle sch teacher	0.8 (-2.4, 4.8)	0	
60	Masks+Cohorts		High sch teacher	0.5 (-3.7, 3.9)	0	
70	None	Equal	Elem. student	13.5 (7.3, 20.1)	11.6	51
70	None	Half	Elem. student	9.1 (4.3, 15)	8.7	35
70	None		Middle sch student	9.9 (4.7, 16.4)	9.7	41
70	None		High sch student	8.2 (3.9, 12.8)	8.2	56
70	None	Equal	Elem. teacher	14.6 (2.5, 28.8)	10.8	
70	None	Half	Elem. teacher	10.8 (0, 21.8)	8	
70	None		Middle sch teacher	16 (0, 35)	11.1	
70	None		High sch teacher	15.2 (0, 35.3)	9.6	
70	Masks	Equal	Elem. student	9.5 (3, 17.8)	7.3	36
70	Masks	Half	Elem. student	3.4 (-0.2, 7.2)	3	13
70	Masks		Middle sch student	4.3 (0, 9.7)	3.4	18
70	Masks		High sch student	1.8 (-0.2, 3.9)	1.6	12
70	Masks	Equal	Elem. teacher	6.9 (0, 15.7)	5.9	
70	Masks	Half	Elem. teacher	2.8 (-0.9, 7.8)	1.8	
70	Masks		Middle sch teacher	4.5 (-2.3, 13.3)	2.3	
70	Masks		High sch teacher	2.1 (-2.1, 5.9)	1.9	

70	Masks+Testing	Equal	Elem. student	7.4 (0.6, 15.7)	5.6	28
70	Masks+Testing	Half	Elem. student	1.8 (-0.5, 4.4)	1.2	7
70	Masks+Testing		Middle sch student	2.4 (-0.5, 7.2)	1.4	10
70	Masks+Testing		High sch student	0.9 (-0.5, 2.7)	0.6	6
70	Masks+Testing	Equal	Elem. teacher	4.9 (-0.9, 12.9)	3.5	
70	Masks+Testing	Half	Elem. teacher	1.4 (-1.8, 4.3)	0.9	
70	Masks+Testing		Middle sch teacher	2.4 (-2.4, 9.1)	2.1	
70	Masks+Testing		High sch teacher	1 (-2, 5.6)	0	
70	Masks+Cohorts	Equal	Elem. student	2.5 (-0.4, 5.7)	1.8	9
70	Masks+Cohorts	Half	Elem. student	0.5 (-0.3, 1.6)	0.4	2
70	Masks+Cohorts		Middle sch student	0.6 (-0.7, 2)	0.4	3
70	Masks+Cohorts		High sch student	0.2 (-0.3, 1.1)	0.2	2
70	Masks+Cohorts	Equal	Elem. teacher	1.2 (-0.9, 4.4)	0.9	
70	Masks+Cohorts	Half	Elem. teacher	0.4 (-1.8, 2.6)	0	
70	Masks+Cohorts		Middle sch teacher	0.6 (-2.3, 4.4)	0	
70	Masks+Cohorts		High sch teacher	0.2 (-2, 3.7)	0	

*Assuming a 380-person elementary school, 420-person middle school, and 680-person high school

Table S15. Excess infections (including asymptomatic infection) attributable to school transmission by population subgroup and scenario examined – Delta variant and vaccine available

Vaccine coverage (%)	NPI	Susceptibility of children <10 years	Population	Mean (89% HPDI) excess infections per 100 population	Median excess infections per 100 population	Infections per school*
50	None	Equal	Elem. student	64.5 (37.7, 90.7)	64	245
50	None	Half	Elem. student	48.6 (30.8, 70)	41.1	185
50	None		Middle sch student	56.9 (30.8, 85)	46.4	239
50	None		High sch student	54.2 (31.8, 82.3)	43.6	368
50	None	Equal	Elem. teacher	37.2 (16.4, 58.6)	36.8	
50	None	Half	Elem. teacher	31.1 (12.2, 50.4)	28.7	
50	None		Middle sch teacher	40.6 (8.7, 69.6)	37.4	
50	None		High sch teacher	41.8 (10.9, 74.5)	35.8	
50	Masks	Equal	Elem. student	49.8 (21.9, 80.2)	35.1	189
50	Masks	Half	Elem. student	23.5 (7.2, 42.2)	21.5	89
50	Masks		Middle sch student	32.9 (6.9, 62.6)	24.2	138
50	Masks		High sch student	21.9 (6.3, 40.9)	20.3	149
50	Masks	Equal	Elem. teacher	20.4 (6.8, 33.9)	19.3	
50	Masks	Half	Elem. teacher	11 (0.8, 21.5)	10.5	
50	Masks		Middle sch teacher	16.8 (0, 37)	14.3	
50	Masks		High sch teacher	13.2 (-2, 26.9)	11.5	
50	Masks+Testing	Equal	Elem. student	43.1 (16.2, 76.3)	31.5	164
50	Masks+Testing	Half	Elem. student	16.4 (2.5, 33.3)	14.7	62
50	Masks+Testing		Middle sch student	25.3 (1.8, 54.1)	18	106
50	Masks+Testing		High sch student	15.5 (1.1, 30.4)	14.1	106
50	Masks+Testing	Equal	Elem. teacher	16.3 (2.6, 29.3)	15.5	
50	Masks+Testing	Half	Elem. teacher	7.6 (-1.8, 16.4)	6.9	
50	Masks+Testing		Middle sch teacher	12.3 (-2.4, 30.4)	10.9	
50	Masks+Testing		High sch teacher	8.9 (-2, 22.6)	7.5	
50	Masks+Cohorts	Equal	Elem. student	17.1 (0.5, 36.6)	12.4	65
50	Masks+Cohorts	Half	Elem. student	4.5 (-1.2, 9)	3.9	17
50	Masks+Cohorts		Middle sch student	6 (-1.8, 13.8)	4.3	25
50	Masks+Cohorts		High sch student	3.3 (-0.6, 7.2)	2.9	22
50	Masks+Cohorts	Equal	Elem. teacher	5.1 (-0.9, 11.8)	5.1	
50	Masks+Cohorts	Half	Elem. teacher	1.8 (-3.6, 6.8)	1.7	
50	Masks+Cohorts		Middle sch teacher	2.3 (-4.9, 11.1)	2.2	
50	Masks+Cohorts		High sch teacher	1.7 (-4, 9.3)	1.9	
60	None	Equal	Elem. student	63 (35.5, 92.5)	45.3	240
60	None	Half	Elem. student	44.8 (25.9, 69.9)	37.2	170
60	None		Middle sch student	53.1 (26.9, 83.8)	40.2	223
60	None		High sch student	48.9 (26.9, 77.6)	38	333

60	None	Equal	Elem. teacher	32.8 (12.8, 54.3)	29.5	
60	None	Half	Elem. teacher	25.9 (10.1, 46.6)	22.7	
60	None		Middle sch teacher	35.9 (6.4, 65.2)	29.5	
60	None		High sch teacher	36.2 (9.1, 70.4)	28.2	
60	Masks	Equal	Elem. student	46.6 (18.4, 80.7)	30.7	177
60	Masks	Half	Elem. student	19.2 (4.1, 40.3)	16.3	73
60	Masks		Middle sch student	26.2 (4.2, 56.4)	18	110
60	Masks		High sch student	13.7 (-0.2, 23.4)	13.1	93
60	Masks	Equal	Elem. teacher	16.5 (3.4, 29.2)	15.7	
60	Masks	Half	Elem. teacher	7.9 (0, 16.7)	7	
60	Masks		Middle sch teacher	12.2 (0, 30.4)	9.3	
60	Masks		High sch teacher	7.7 (-2, 17.6)	7.3	
60	Masks+Testing	Equal	Elem. student	38 (9.3, 76.3)	25.5	145
60	Masks+Testing	Half	Elem. student	12 (-0.9, 27.8)	9.5	46
60	Masks+Testing		Middle sch student	17.7 (-1.7, 45)	11.2	74
60	Masks+Testing		High sch student	8.3 (-1.4, 17.3)	7.2	57
60	Masks+Testing	Equal	Elem. teacher	12.5 (0.9, 25.2)	11.2	
60	Masks+Testing	Half	Elem. teacher	4.6 (-1.8, 12.2)	3.5	
60	Masks+Testing		Middle sch teacher	7.8 (-4.5, 21.4)	4.7	
60	Masks+Testing		High sch teacher	4.6 (-2, 14.3)	3.8	
60	Masks+Cohorts	Equal	Elem. student	13.8 (-0.1, 33.1)	9.4	53
60	Masks+Cohorts	Half	Elem. student	3.2 (-0.7, 7.7)	2.4	12
60	Masks+Cohorts		Middle sch student	4.1 (-2, 10)	2.7	17
60	Masks+Cohorts		High sch student	1.8 (-1, 4.3)	1.5	12
60	Masks+Cohorts	Equal	Elem. teacher	3.4 (-1.8, 8.4)	3.4	
60	Masks+Cohorts	Half	Elem. teacher	1.2 (-1.8, 5.3)	0.9	
60	Masks+Cohorts		Middle sch teacher	1.5 (-4.5, 6.7)	2.1	
60	Masks+Cohorts		High sch teacher	1 (-3.9, 5.9)	1.8	
70	None	Equal	Elem. student	60.1 (30.7, 92.2)	39.8	228
70	None	Half	Elem. student	39.9 (19.1, 69.1)	32.4	152
70	None		Middle sch student	47.3 (20.9, 80.7)	34.3	199
70	None		High sch student	28 (7.9, 48.3)	22.2	
70	None	Equal	Elem. teacher	20.6 (5.2, 40.4)	16.9	
70	None	Half	Elem. teacher	30 (2.1, 59.5)	22.2	
70	None		Middle sch teacher	40.5 (18.4, 72.2)	31.9	
70	None		High sch teacher	28.9 (3.6, 62.3)	20	
70	Masks	Equal	Elem. student	42.7 (13.5, 81.3)	25.4	162
70	Masks	Half	Elem. student	15 (0, 34)	12	57
70	Masks		Middle sch student	20.3 (-0.5, 47.1)	13.5	85
70	Masks		High sch student	7.9 (-0.3, 15.3)	7.6	54
70	Masks	Equal	Elem. teacher	13.1 (0.8, 24.8)	11.9	
70	Masks	Half	Elem. teacher	5.5 (-0.9, 12.8)	4.5	
70	Masks		Middle sch teacher	8.3 (-2.3, 20.8)	6.5	

70	Masks		High sch teacher	4.2 (-2, 11.1)	3.7	
70	Masks+Testing	Equal	Elem. student	33.3 (5, 75.5)	21.2	126
70	Masks+Testing	Half	Elem. student	7.7 (-1.3, 19.5)	5.2	29
70	Masks+Testing		Middle sch student	11.1 (-1.3, 33.7)	5.9	47
70	Masks+Testing		High sch student	4 (-0.9, 9.7)	2.8	27
70	Masks+Testing	Equal	Elem. teacher	9.4 (-0.9, 20.7)	7.8	
70	Masks+Testing	Half	Elem. teacher	2.7 (-1.8, 7.6)	1.8	
70	Masks+Testing		Middle sch teacher	4.3 (-4.5, 13.6)	2.3	
70	Masks+Testing		High sch teacher	2 (-2, 7.7)	1.9	
70	Masks+Cohorts	Equal	Elem. student	11.4 (-0.8, 27.5)	7.3	43
70	Masks+Cohorts	Half	Elem. student	2.3 (-0.7, 6.1)	1.5	9
70	Masks+Cohorts		Middle sch student	2.9 (-1.1, 7.9)	1.5	12
70	Masks+Cohorts		High sch student	1 (-0.6, 3.2)	0.8	7
70	Masks+Cohorts	Equal	Elem. teacher	2.4 (-0.9, 7)	1.8	
70	Masks+Cohorts	Half	Elem. teacher	0.8 (-1.8, 3.5)	0.9	
70	Masks+Cohorts		Middle sch teacher	1.1 (-2.4, 6.7)	0	
70	Masks+Cohorts		High sch teacher	0.4 (-3.8, 3.8)	0	

*Assuming a 380-person elementary school, 420-person middle school, and 680-person high school

Table S16. Excess infections (all, symptomatic, and severe) among household members of students and general community members, stratified by vaccination coverage and within-school intervention (Delta variant).

Vaccination coverage (%)	Population	Within –school intervention	Excess infections (any) per 100 population		Excess symptomatic infections per 100 population		Excess hospitalizations per 100,000 population	
			Mean (89% HPDI)	Median	Mean (89% HPDI)	Median	Mean (89% HPDI)	Median
50	Community member	No NPIs	14.7 (10.2, 20.3)	14.7	8.6 (6.1, 11.5)	8.8	34.9 (19.2, 50.8)	35.3
50	Household member	No NPIs	26.4 (15.2, 39.2)	23.5	10.9 (8.5, 13.6)	11.1	34.7 (8.2, 65.4)	33.8
50	Community member	Masks	6 (0.9, 10.9)	6	3.5 (0.5, 6.4)	3.5	14.1 (0.9, 29.8)	14.1
50	Household member	Masks	12.1 (2.3, 21.9)	10.9	4.9 (1.4, 8.1)	5.1	15.7 (-4.5, 38.5)	12.9
50	Community member	Masks + testing	3.8 (-0.7, 8)	3.8	2.2 (-0.7, 4.5)	2.1	8.9 (-4.5, 22)	8.8
50	Household member	Masks + testing	8.3 (0, 17.5)	7.4	3.3 (0.1, 6.3)	3.3	9.4 (-8.9, 30.1)	8.5
50	Community member	Masks +cohorts	1.1 (-1.7, 3.6)	1.1	0.7 (-1.2, 2.2)	0.6	2.6 (-7.1, 13.2)	2.6
50	Household member	Masks +cohorts	2.2 (-1, 5.3)	2.1	0.9 (-0.7, 2.5)	0.9	2.6 (-13.1, 20.7)	4.1
60	Community member	No NPIs	11 (6.6, 16.4)	11	6.1 (3.5, 8.6)	6.3	23.8 (11.3, 38)	23.9
60	Household member	No NPIs	22.3 (11.2, 36.6)	18.9	8.6 (5.9, 11.5)	8.9	25.1 (0, 47)	24.9
60	Community member	Masks	3.7 (0.4, 7.4)	3.6	2.1 (0.2, 4.1)	2	7.9 (-0.9, 18.4)	7.9
60	Household member	Masks	8.4 (0.3, 16.3)	7.4	3.2 (0.8, 6)	3.2	9.1 (-4.4, 25.7)	8.5
60	Community member	Masks +testing	2.1 (-0.8, 5.2)	1.9	1.1 (-0.5, 2.9)	1	4.4 (-4.4, 14)	3.6
60	Household member	Masks +testing	5.1 (-1, 12)	4.2	1.9 (-0.3, 4.5)	1.8	5.4 (-8.7, 21.3)	4.3
60	Community member	Masks +cohorts	0.7 (-1, 2.4)	0.6	0.4 (-0.7, 1.4)	0.3	1.5 (-5.3, 8.9)	1.7
60	Household member	Masks +cohorts	1.4 (-0.9, 3.9)	1.3	0.5 (-0.6, 1.7)	0.5	1.3 (-8.7, 13)	0
70	Community member	No NPIs	7.3 (3.2, 12.6)	7.4	3.8 (1.6, 6.2)	4	14.1 (2.6, 23.9)	14.2
70	Household member	No NPIs	17.7 (5.7, 32.3)	13.4	6.2 (3.2, 9.5)	6.5	16 (0, 34.1)	13
70	Community member	Masks	2.2 (-0.4, 4.6)	2	1.1 (-0.2, 2.4)	1	4.1 (-1.8, 11.5)	3.5
70	Household member	Masks	5.8 (-0.7, 13)	4.6	2 (0, 4)	1.9	4.7 (-4.4, 17.1)	4.2
70	Community member	Masks +testing	1 (-0.6, 3)	0.7	0.5 (-0.5, 1.5)	0.4	1.9 (-3.6, 7.9)	1.8
70	Household member	Masks +testing	2.9 (-1, 8.5)	1.7	1 (-0.5, 2.7)	0.8	2.1 (-4.4, 16.7)	0
70	Community member	Masks +cohorts	0.3 (-0.7, 1.3)	0.3	0.2 (-0.4, 0.8)	0.2	0.7 (-3.5, 5.3)	0.9
70	Household member	Masks +cohorts	0.9 (-0.6, 2.6)	0.7	0.3 (-0.5, 1)	0.3	0.7 (-8.6, 8.6)	0

Supplemental Information for Chapter 4

Supplementary Figures

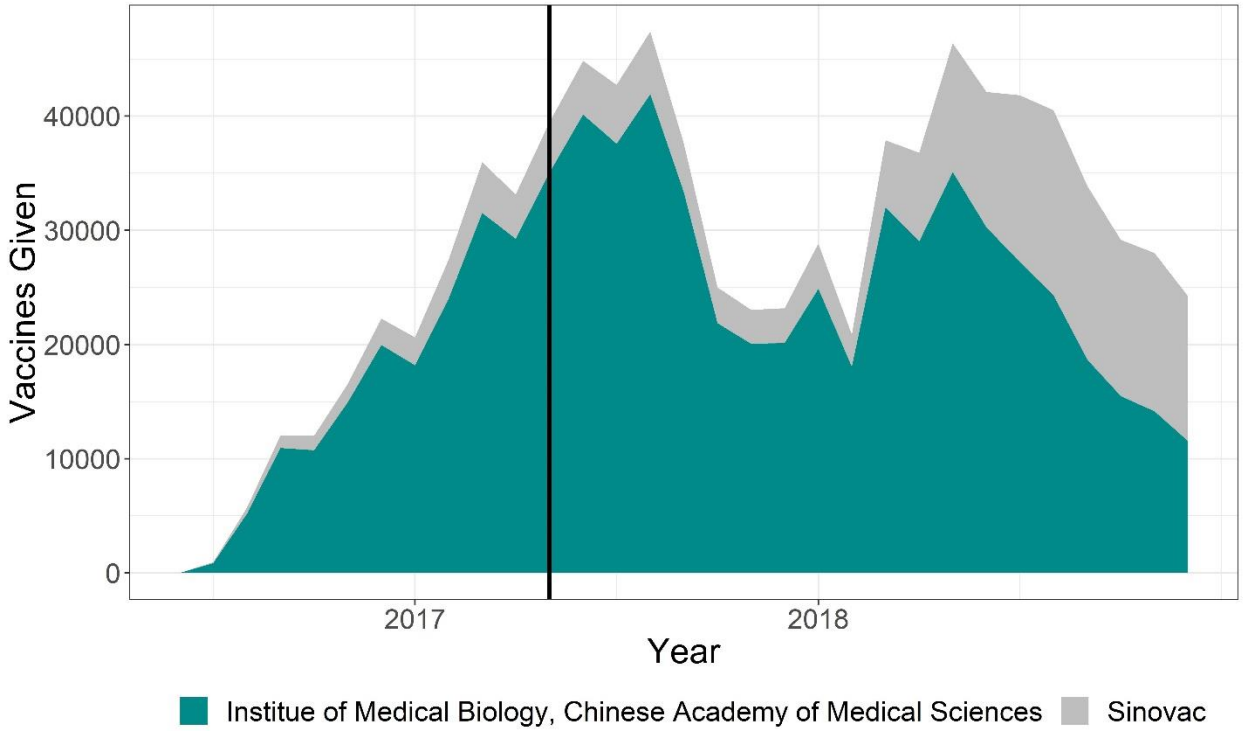


Figure S1. Number of doses of the EV71 vaccine given per month in Chengdu Province. Black line indicates change point identified from change point analysis.

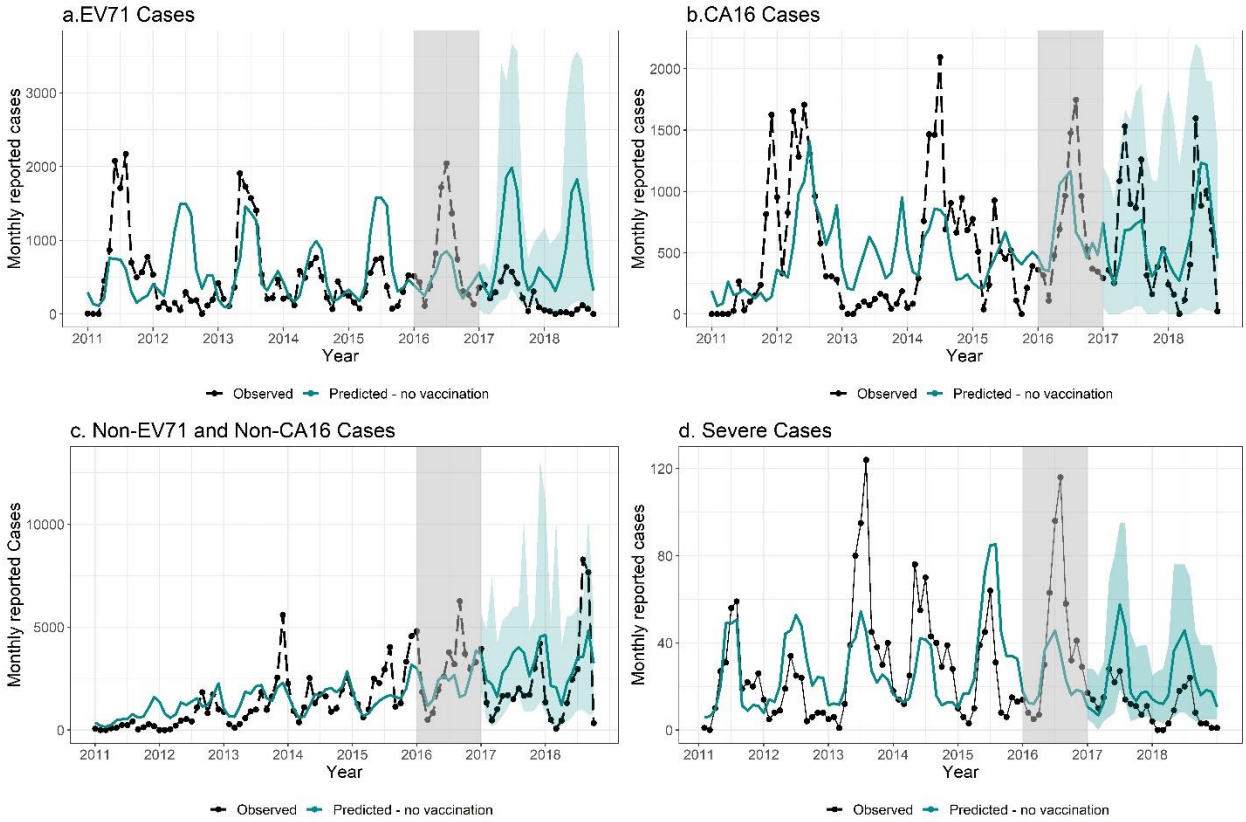


Figure S2. Each analysis examines trends from 2011-2015 for a) EV71-specific HFMD, b) CA16-specific HFMD, c) HFMD caused by another enterovirus, and d) severe HFMD, including comparison of observed events (black dotted lines and dots) to predicted events (teal solid line) before and after vaccination. Expected number of events and shown 95% prediction intervals are based on predictions from random forest models fitted to historic data. The gray box indicates the year vaccination first began and is removed from calculation of averted cases.

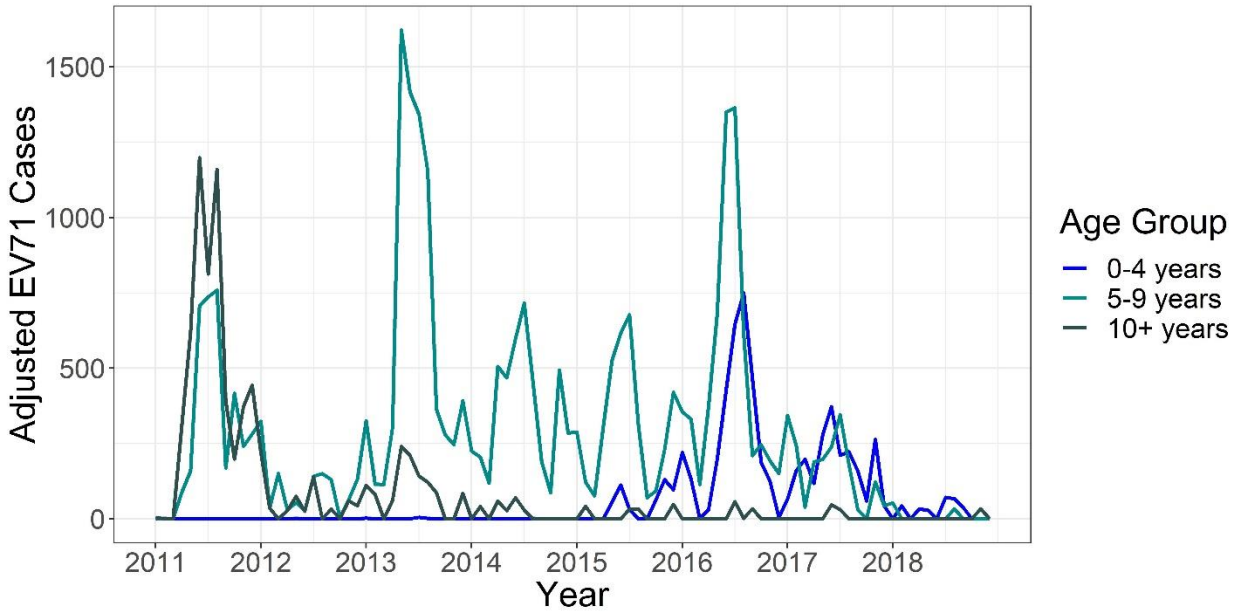


Figure S3. Adjusted time series of EV71 cases for age groups 0-4 years, 5-9 years, and 10+ years.

Supplementary Text 1. Inverse probability weighting

Because a subset of reporting cases underwent laboratory confirmation, we constructed adjusted time series for EV71, CA16, and other (defined as non-EV71 and non-CA16) enteroviruses. We defined the outcome *Y* as a Bernoulli random variable with possible values of 1 for tested for serotype, and 0 for not tested. We estimated the probability of test using logistic regression with fixed effects for month of diagnosis, case severity (binary, defined as any case suffering any cardiopulmonary or neurological complications), age group (0-4 years, 5-9 years, 10+ years), sex, and a linear trend by year stratified by case severity (see model specification below). We then calculated each case's individual probability of being tested according to their covariate pattern, and applied inverse probability weights to each case.

$$Y \sim \mathrm{Bern}(\mu)$$

$$\mathrm{logit}(\mu) = \beta_0 + \beta_1 \mathrm{sex} + \beta_2 \mathrm{severity} + \beta_3 \mathrm{year} + \beta_4 \mathrm{year} \times \mathrm{severity} + \sum_{i=1}^{11} [\alpha_i \times \mathbb{I}(\mathrm{month}_i)] + \epsilon$$

where

$$\epsilon \sim \mathrm{N}(0, \sigma^2).$$

Table ST1.1 Coefficients and standard errors from main effects in logistic regression model predicting the probability of being tested for serotype determination.

Coefficients:	**Estimate**	**Std. Error**
(Intercept)	-30.058625	11.09219
Sex (ref = female)	0.118798	0.022339
Severity	52.273042	44.793442
year	0.013066	0.005504
Month (ref = Jan.)		
February	0.096431	0.107513
March	0.315463	0.080369
April	0.436013	0.071652
May	0.577821	0.069562
June	0.342666	0.070141
July	0.22021	0.069785
August	0.117488	0.072792
September	0.640513	0.07469
October	0.422484	0.072794
November	-0.10741	0.074536
December	-0.24299	0.080367
Severity*year	-0.024157	0.022238

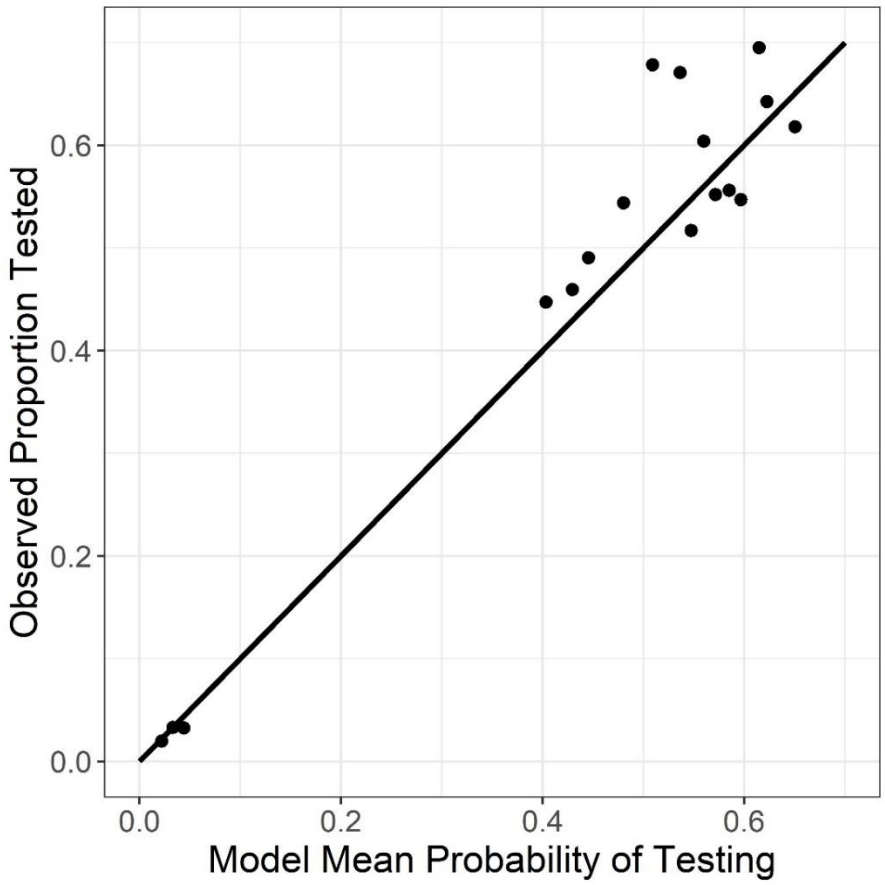


Figure ST1.1. The observed proportion of cases tested within bins of predicted probability of the cases being tested, where predicted probabilities are determined through logistic regression. Bins are determined based on percentiles of the data, ensuring each bin contains at least 30 observations. Points should lie along a straight line, shown, for ideal fit.

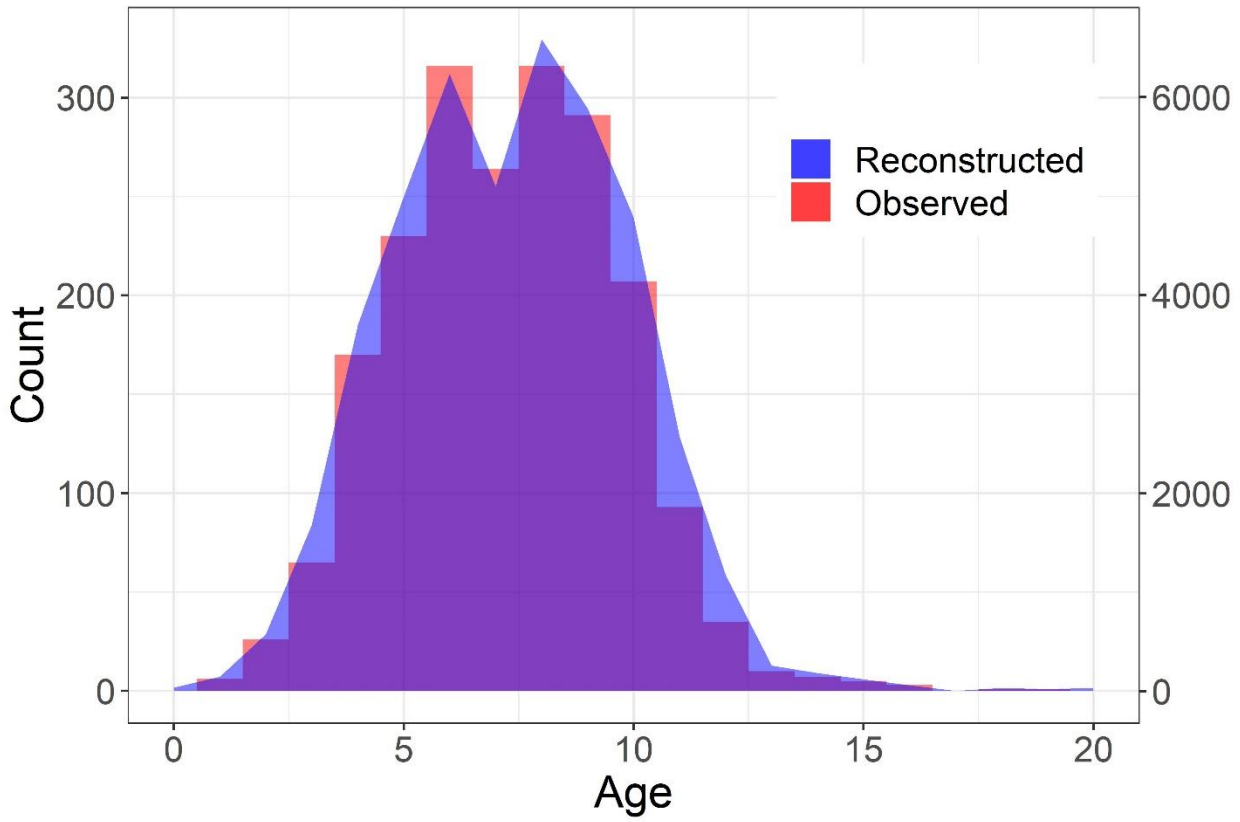


Figure ST1.2. Comparison of the observed (red, left axis) and reconstructed (blue, right axis) distribution of ages among EV71 cases.

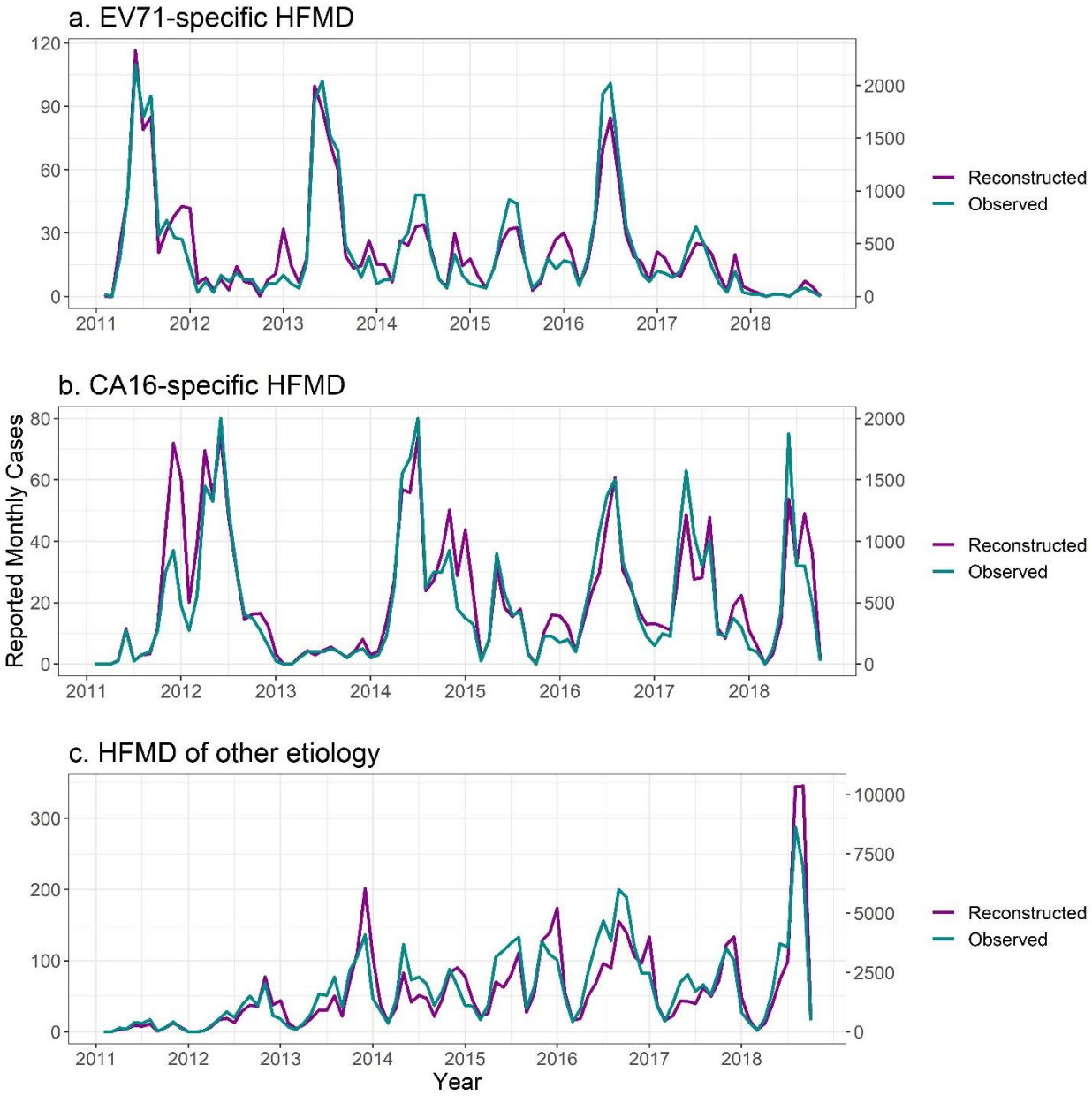


Figure ST1.3. Comparison of the observed (green, left axis) and reconstructed (purple, right axis) time series of HFMD cases caused by a) EV71, b) CA16, and c) other, non-EV71 or –CA16 etiology in Chengdu, 2011 - 2018

Supplementary Text 2. Comparing quasi-Poisson to Negative Binomial distributions

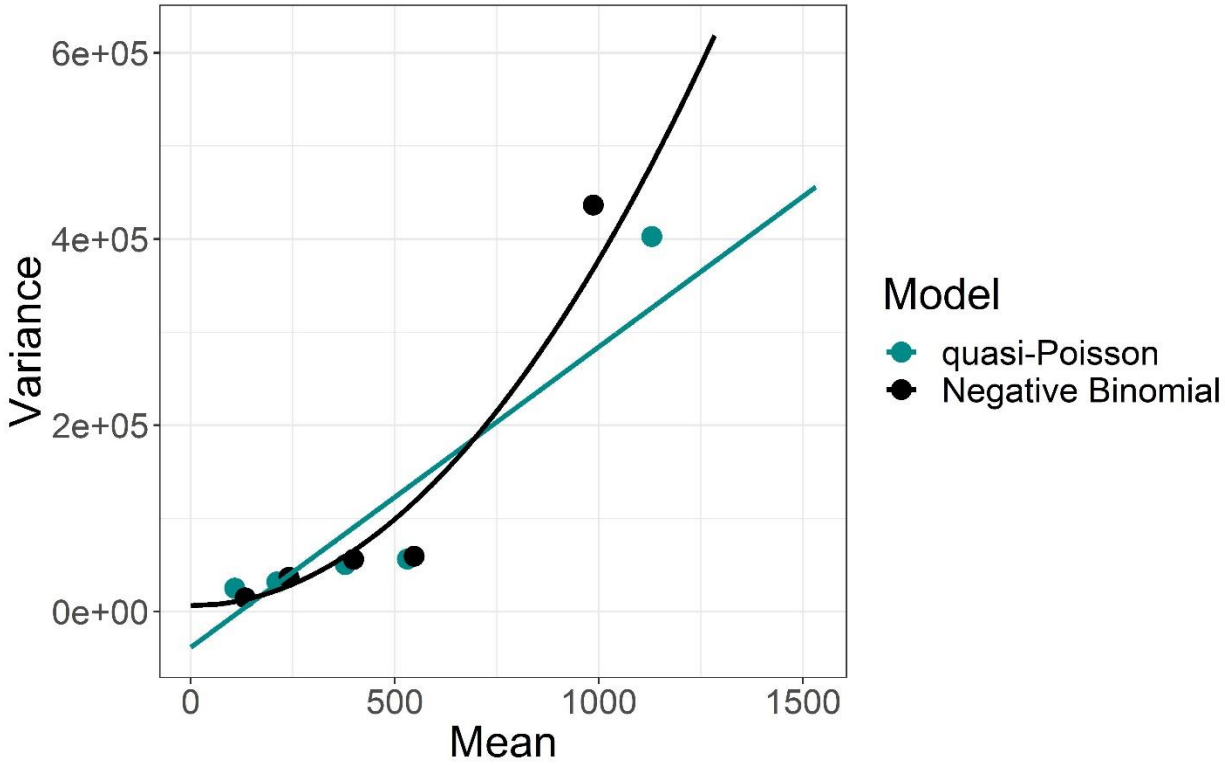


Figure ST2.1. When deciding between quasi-Poisson and negative binomial models, it is helpful to plot the mean as a function of the variance. If the variance varies as a function of the mean, as our does, a negative binomial model is a better fit.

Both the quasi-Poisson model and negative binomial model account for overdispersion by inclusion of a mean and variance parameter. An important difference is that for the quasi-Poisson model, the variance is linearly related to the mean, whereas for negative binomial, the variance is a quadratic function of the mean.[273] Ver Hoef and Boveng recommend to plot $(Y_i - \mu_i)^2$ versus μ_i as a tool in deciding whether to account for overdispersion using quasi-Poisson or Negative Binomial models [273]. Because this can be noisy, it is common to bin the predicted values and calculate means within each bin. We do so in Figure ST3.1. We then regressed the predicted values and the squared predicted values against the model residuals. For the quasi-Poisson model, a linear functional form better described the relationship between model predictions and squared residuals. The regression with the quadratic term on the negative binomial predictions had a lower AIC than did the regression with the linear term on the quasi-Poisson predictions. Therefore, we used the Negative Binomial model.

Supplementary Text 3. Model Formulas

Models below were selected based on their predictive ability in leave-out-one-year blocked cross validation. All models assume a negative binomial distribution of the outcome (see Supplemental Text File 3).

The model selected as best predicting the number of cases of EV71 per month was specified as:

$$\log(E[Y_t|\boldsymbol{x_t}, \beta_1, \boldsymbol{\alpha_i}]) = \log(\text{pop}_t) + \beta_0 + \beta_1 \times \text{EV}_{t-12} + \sum_{i=1}^{11}[\alpha_i \times \mathbb{I}(\text{month}_i)] + \epsilon_t$$

where:

$Y_t =$ number of cases of EV71 reported in month t, for $t \in (1,2,\dots,96)$

$\text{pop}_t =$ population in month t (yearly quantity)

$\text{EV}_{t-12} =$ last year's lagged incidence of EV71

$\mathbb{I}(\text{month}_i) =$ Indicator of calendar month

$\epsilon_t \sim \text{N}(0, \sigma^2)$

The model selected as best predicting the number of cases of CA16 per month was age adjusted and was specified as:

$$\begin{aligned}\log\big(E\big[Y_{\text{ag},t}\big|\boldsymbol{x_{\text{ag},t}}, \boldsymbol{\beta_j}, \boldsymbol{\gamma_i}, \boldsymbol{\delta_i}, \boldsymbol{\alpha_k}\big]\big) \\ = \log\big(pop_{\text{ag},t}\big) + \beta_0 + \beta_1 \times year + \beta_2 \times N_{\text{ag},t-12} + \beta_3 \times \sin\left(\frac{2\pi t}{12}\right) \\ + \sum_{i=1}^{2}[\gamma_i \times \mathbb{I}(\text{ag}_i)] + \sum_{i=1}^{2}[\delta_i \times \mathbb{I}(\text{ag}_i) \times \mathbb{I}(t > 50)] + \sum_{k=1}^{11}[\alpha_k \times \mathbb{I}(\text{month}_k)] \\ + \epsilon_{\text{ag},t}\end{aligned}$$

where:

$Y_{\text{ag},t} =$ number of cases of CA16 in month t in age group ag, for $ag \in (< 5\text{ yr}, 5-9\text{ yr}, \geq 10\text{ yr})$

$pop_{\text{ag},t} =$ population of age group ag in month t, for $t \in (1,2,\dots,96)$

$year =$ continuous term on year

$\mathbb{I}(\text{ag}_i) =$ Indicator for age group

$\mathbb{I}(t > 50) =$ Indicator for time past month 50

$\mathbb{I}(\text{month}_k) =$ Indicator of calendar month

$N_{\text{ag},t-12} =$ last year's lagged total incidence in age group ag

$\epsilon_{\text{ag},t} \sim \text{N}(0, \sigma^2)$

The model selected as best predicting the number of cases of other enteroviruses per month was age adjusted and was specified as:

$$\log\left(E\left[Y_{\text{ag},t} \middle| \boldsymbol{x}_{\mathbf{ag},t}, \boldsymbol{\beta_j}, \boldsymbol{\alpha_k}\right]\right) = \log(pop_{\text{ag},t}) + \beta_0 + \beta_1 \times year + \beta_2 \sin\left(\frac{2\pi t}{12}\right) + \beta_3 \cos\left(\frac{2\pi t}{12}\right) + \sum_{i=1}^{2} \gamma_i \times \mathbb{I}(\text{ag}_i) + \sum_{k=1}^{11} \alpha_k \times \mathbb{I}(month_k) + \epsilon_{\text{ag},t}$$

Where:

$Y_{\text{ag},t} =$ number of cases of non $-$ EV71 and non $-$ CA16 enteroviruses in month t in age group ag, for $ag \in (< 5 \text{ yr}, 5-9 \text{ yr}, \geq 10 \text{ yr})$

$pop_{\text{ag},t} =$ population of age group ag in month t, for $t \in (1,2,\dots,96)$

$year =$ continuous term on year

$\mathbb{I}(\text{ag}_i) =$ Indicator for age group

$\mathbb{I}(\text{month}_k) =$ Indicator of calendar month

$\epsilon_{\text{ag},t} \sim \text{N}(0, \sigma^2)$

The model selected as best predicting the number of severe cases per month was specified as:

$$\log\left(E\left[Y_t \middle| \boldsymbol{x_t}, \boldsymbol{\beta_j}, \boldsymbol{\alpha_i}\right]\right) = \log(pop_t) + \beta_0 + \beta_1 \times \mathbb{I}(year < 2013) + \beta_2 \times N_{t-12} + \beta_3 \sin\left(\frac{2\pi t}{6}\right) + \beta_4 \cos\left(\frac{2\pi t}{6}\right) + \beta_5 \sin\left(\frac{2\pi t}{12}\right) + \beta_6 \cos\left(\frac{2\pi t}{12}\right) + \beta_7 \sin\left(\frac{2\pi t}{24}\right) + \beta_8 \cos\left(\frac{2\pi t}{24}\right) + \epsilon_t$$

Where all terms are defined previously and:

$Y_t =$ number of severe cases reported in month t, t ϵ 1, … ,96.

$\epsilon_t \sim \mathrm{N}(0, \sigma^2)$

Change Point Model:

For all values of θ between 2 and 95 (number of months between 2011 and 2018 excluding the first and last month), we fit the model described and evaluated the goodness of fit of that model using according to its log likelihood.

The model specification for the change point analysis was as follows:

$$\log\left(E\left[Y_t \middle| \boldsymbol{x_t}, \boldsymbol{\beta_i}, \boldsymbol{\delta_j}, \boldsymbol{\gamma_j}\right]\right) =$$

$$\log(pop_t) + \beta_0 + \beta_1 year + \sum_{j=1}^{3}\left[\delta_j \sin\left(\frac{2\pi t}{12\tau_j}\right) + \gamma_j \cos\left(\frac{2\pi t}{12\tau_j}\right)\right] + \beta_2 \times year \times \mathbb{I}(t \geq \theta) + \epsilon_t$$

Where all terms are as defined previously and:

$Y_t =$ Number of EV71 cases in month t, t ϵ 1, … ,96

θ is a possible month of a change point, between 2 and 95

$\mathbb{I}(t \geq \theta) =$ Indicator for month at or past change point, θ

$\tau_1 = 0.5, \tau_2 = 1, \tau_3 = 2$

$$\epsilon_t \sim \mathrm{N}(0, \sigma^2)$$

Supplementary Tables

Table S1. Distribution of HFMD incidence by sex, age, severity, and serotype in Chengdu, China for each year 2011-2018. Yearly totals are used as denominators for percentage male, severe, deaths, and serotyped. Yearly serotyped totals are used as denominators for percentage EV71, CA16, and other enteroviruses.

	2011	2012	2013	2014	2015	2016	2017	2018	Total
Total, N	19,413	23,559	22,356	40,544	28,888	45,895	28,623	70,074	279,352
Male, N (%)	11,709 (60.3)	14,177 (60.2)	13,255 (59.3)	23,479 (57.9)	16,786 (58.1)	26,630 (58.0)	16,420 (57.4)	40301 (57.5)	162,757 (58.3)
Age, Median (Range)	10 (1 - 65)	9 (2 - 84)	8 (1 - 68)	7 (2 - 73)	6 (3 - 62	5 (1 - 74)	4 (1 - 71)	2 (0 - 65)	5 (0 - 84)
Severe, N (%)	285 (1.5)	155 (0.7)	528 (2.4)	441 (1.1)	254 (0.9)	502 (1.1)	175 (0.6)	90 (0.1)	2,430 (0.9)
Deaths, N	3	2	6	2	1	4	1	0	19
Serotyped, N (%)	685 (3.5)	743 (3.2)	1,115 (5.0)	1,342 (3.3)	1,362 (4.7)	1,989 (4.3)	1,184 (4.1)	1,459 (2.1)	9,879 (3.5)
EV71	491 (71.7)	79 (10.6)	441 (39.6)	235 (17.5)	220 (16.2)	425 (21.4)	151 (12.8)	11 (0.8)	2,053 (20.8)
CA16	117 (17.1)	354 (47.6)	36 (3.2)	401 (29.9)	142 (10.4)	303 (15.2)	287 (24.2)	219 (15.0)	1,859 (18.8)
Other	77 (11.2)	310 (41.7)	638 (57.2)	706 (52.6)	1,000 (73.4)	1,261 (63.4)	746 (63.0)	1,229 (84.2)	5,967 (60.4)

Table S2. Observed EV71, CA16, and other etiology HFMD case counts, expected cases under random forest and change point models, and incidence rate ratios for the post-vaccine period 2017-2018 in Chengdu, China

	Cases in 2017		Cases in 2018		Combined Cases, 2017 - 2018		
	Observed	**Averted (95% PI)****	**Observed**	**Averted (95% PI)**	**Observed**	**Averted (95% PI)**	**IRR (95% PI)**
EV71*	4,198	3,692 (-2,124, 12,981)	400	4,622 (182, 12,266)	4,598	8,314 (-1,942, 25,247)	0.35 (0.15, 1.73)
Change point*	3,150†	7,427 (-14, 22,438) †	400	4,992 (1,570, 11,154)	3,550	12,326 (3,520, 28,114)	0.22 (0.11, 0.50)
CA16*	8,465	-252 (-7,430, 5,619)	6,702	-96 (-4,737, 16,854)	15,167	-348 (-12,167, 22,473)	1.02 (0.40, 5.06)
Other enterovirus*	26,647	-16,563 (-17,938, 17,099)	41,658	-20,374 (-24,705, 7,555)	68,304	-36,937 (-42,643, 24,654)	2.18 (0.73, 2.66)
Severe cases	175	115 (-92, 439)	90	195 (-26, 493)	265	310 (-118, 932)	0.46 (0.22, 1.80)

IRR = Incidence Rate Ratio; PI = Prediction Intervals

*Adjusted by inverse probability of testing. See Supplemental Text File 1

**Prediction intervals calculated via 1,000,000 Monte Carlo simulation draws using fitted model parameters

†Values in 2017 reflect those occurring after the change point