**1. Introduction**

Emerging respiratory infections impose substantial burdens on human health. During the first two decades of the 21st century, novel respiratory viruses – influenza A/H1N1 (H1N1) and SARS-CoV-2 – have produced two global pandemics that caused over 15 million excess deaths worldwide [1, 2]. In densely populated metropolitan areas, respiratory outbreaks often start early and propagate rapidly. During the COVID-19 pandemic, New York City (NYC) was the epicenter for the US during the spring of 2020 and has reported over 2.2 million cases and 40,000 deaths as of May 1st 2022 [3]. Studying novel respiratory infections in cities is also particularly interesting from a behavioral perspective [4], given that individual decision-making under infection risk varies across local communities with different socioeconomic status [5]. These differences impose a scale of variability—the neighborhood level—that is important to observe and understand. To improve the control and mitigation of future pandemics, understanding and forecasting transmission of SARS-CoV-2 at the neighborhood level in urban settings is critical.

Mathematical models have been increasingly used to support public health decision-making in response to epidemic outbreaks [6, 7]. Over the last decades, real-time forecasting for seasonal influenza has been significantly advanced [8–21]. Some of these forecasting systems, with validated predictive accuracy, have been operationalized and tested in the real world [22–25]. However, most **predictive** models for infectious disease transmission have not explicitly and adequately represented human behavior, which can greatly modulate the transmission of respiratory pathogens [26, 27]. In particular, reactive behavior change under infection risk during a pandemic can produce feedback between disease outcomes and mobility patterns [26]. Theoretical epidemic models with increasing complexity were proposed to improve behavioral realism [28–36], but most of these models relied on strong assumptions on human behavior and were rarely validated by real-world behavioral data and real-time disease forecasting. The absence of validated data-driven epidemic models including behavior drivers and feedbacks has impeded accurate prediction of respiratory diseases.

During the COVID-19 pandemic, studies including human behavior in mathematical models predominantly focused on mobility features such as travel distance and the time staying at home [37–40]. Other online behavior features such as media attention and Wikipedia page visits have also been used as indicators of behavior change in earlier modeling studies on H1N1 and SARS [35, 36]. However, these features are not directly relevant to virus transmission and some were found to provide limited information [41]. To date, behavioral factors that **directly** modulate the transmission of respiratory pathogens are still not adequately represented in most models. These factors include **contact patterns indoors** (e.g., crowdedness and dwell time) where most transmission occurs, **reactive behavior changes** to infection risk (e.g., avoiding high-risk places), and the adoption of **protective measures** such as mask-wearing. In this project, we will leverage behavior theory and detailed data to develop behavior-driven epidemic models, study the transmission dynamics of COVID-19 in NYC, and generate improved forecasting systems.

Our project aims to address three key questions in developing next-generation mathematical models for respiratory diseases with better incorporation of human behavior:

**1)**. *Can we model behavior changes in response to infection risk and validate the behavior model?*

**2)**. *Can we accurately represent the impact of indoor crowdedness, dwell time, and masking in epidemic models?*

**3)**. *Can we develop and validate behavior-driven predictive models for respiratory pathogens in cities?*

To address these questions, we have assembled a multidisciplinary team and multiple datasets tailored to support the project. Our team brings a breadth of disciplinary expertise spanning mathematical modeling, statistical inference, behavior science, data science, and infectious disease epidemiology. Team members have been working on infectious disease modeling and forecasting for over a decade and have developed multiple novel mathematical and computational methods for a range of infectious diseases including influenza [8–15], COVID-19 [42–47], measles [48–50], dengue [51, 52], Ebola [53, 54], and antimicrobial-resistant pathogens [55, 56]. Our team has also worked on core questions in behavioral science, such as loss-aversion, temporal discounting, positive deviance, rational choice models, nudging and boosting, and the impact of social/systemic barriers on inequalities (health, social, economic) [57–65].

The proposed project will be supported by rich and diverse datasets include neighborhood level COVID-19 data in NYC, detailed foot traffic data from 84,717 points of interest (POIs, e.g., restaurants, bars, etc.) in NYC [66], masking data derived from the Delphi COVID-19 Trends and Impact Surveys [67, 68], and data on socio-economic indicators that may impact human behavior. During the project, we will additionally conduct surveys to collect complementary behavioral data at the neighborhood level. Guided by *well-established behavior theories*, we will develop models to test different modes of feedback between observed behaviors and disease transmission. We will then use state-of-the-art Bayesian filtering techniques to calibrate high-dimensional epidemic models to these data and generate predictions.

Leveraging our existing collaborations with federal and local public health agencies, particularly the US Centers for Disease Control and Prevention (CDC) and the New York City Department of Health and Mental Hygiene (DOHMH) (see letter of collaboration), we will disseminate project findings and products to translate research into strategies for disease prevention and mitigation. The project will train next-generation scientists working at the interface between infectious disease modeling and behavior science, and the transdisciplinary research conducted in this project will lead to significant advances in mathematical modeling of respiratory diseases with deep integration of human behavior and epidemiological models.

Graphical user interface, diagram

Description automatically generated

**Fig. 1.** Overview of the proposed project.

**2. Intellectual Merit**

The project will promote the integration of mathematical, epidemiological, and behavior sciences. Efforts in this project will fundamentally improve disease model realism, incorporate behavior models, and advance mathematical methods for respiratory disease forecasts. Proposed studies are organized around three synergistic research cores.

**1)**. *Research core I* addresses behavior change and decision-making under infection risk. We will use core behavioral science theory (*temporal discounting, loss aversion, agency,* and *normative/non-normative decisions*) to inform models for risk-driven behavior change (e.g., avoiding restaurants and public transportation) and validate the models using real-world behavioral data, including foot traffic records and surveys on mask wearing. We will further validate the predictive model for reactive behavior change.

**2)**. *Research core II* focuses on the mathematical representation of human behaviors in epidemic models. We will explore how *dwell time* and *crowdedness* in different indoor settings and *population-level* *masking* impact respiratory disease transmission. We will develop a data-driven functional form to link behavior to transmission rates in epidemic models and incorporate this form into a behavior-driven metapopulation model (see Section 4.4.1), which will be validated through model fitting to multiple datasets collected during multiple COVID-19 waves in NYC.

**3)**. *Research core III* incorporates the risk-driven behavior model developed in Research core I into the metapopulation model developed in Research core II. This metapopulation model will be coupled with statistical filtering algorithms to update model states in real time using recent observations and will be used to generate probabilistic predictions of COVID-19 at the neighborhood level. Predictive skill will be thoroughly evaluated through retrospective forecasting that mimics predictions in real time.

The structure of the proposed project is shown in **Fig. 1**. The model framework will allow the real-time forecasting of COVID-19 in NYC at the neighborhood scale. Findings from this project can potentially inform questions and approaches for modeling other emergent respiratory diseases.

**3. Formation of an Interdisciplinary Team**

To successfully implement the proposed project, we have formed a leadership team comprising experts from a multitude of scientific fields broadly spanning mathematics, statistics, behavior science, data science, and infectious disease epidemiology. The PI and co-PIs are affiliated in three departments (Environmental Health Sciences, Health Policy and Management, and Epidemiology) within the same institution (Mailman School of Public Health at Columbia University) and are close collaborators. Specifically, our interdisciplinary team includes:

**1)**. The PI, Dr. Sen Pei, is an **applied mathematician** with a strong background in infectious disease modeling, statistical inference, nonlinear dynamical systems, and network science. His research centers on the transmission dynamics of infectious disease. He has developed multiple novel mathematical and computational methods for infectious disease modeling, surveillance, and forecasting. Over the last 7 year, he has led a line of data-driven modeling research on endemic and emerging infectious diseases such as COVID-19 [42–44, 69–77], influenza [10, 11, 78–81] and antibiotic-resistant pathogens [55, 56]. He actively participated in the CDC-led real-time influenza forecasting challenge since 2016 and developed one of the first COVID-19 forecast models at the US county level [70, 82].

**2)**. Co-PI, Dr. Kai Ruggeri, studies **population behavior** and **decision-making,** focusing on integrating behavioral evidence into policies to improve health equity [57–65]. He has completed dozens of published studies looking at the links between behavior, evidence, and policy, exploring everything from neighborhood-specific factors in urban areas (mainly NYC) to global comparative studies involving over 60 countries. He works with organizations across industrial, scientific, government, and non-profit sectors in these studies, including recent work with UN Women, the National Academies (in partnership with the CDC), NASA, and a network of federally qualified urban health centers. He also contributed to the New York State COVID-19 response through State Active Duty military service from 2020 to 2022.

**3)**. Co-PI, Dr. Jeffrey Shaman, is an expert in **infectious disease epidemiology** and **modeling** who has made significant contributions to the understanding of environmental drivers of respiratory disease spread [83, 84] and real-time disease forecasting [8, 9]. He has designed, coded, and experimented with a spectrum of disease, climate, hydrology, and ecology models. These models range from simple oscillators to models of intermediate complexity, up to large-scale global climate models [85–93]. Dr. Shaman also has extensive experience developing and working with the model-inference methods proposed for use in this project [12, 94].

**4)**. Co-PI, Dr. Wan Yang has an interdisciplinary background in **civil engineering**, **infectious disease modeling** and **epidemiology**. She has studied the airborne and droplet transmission of influenza virus in indoor environment [95–98] and extensively modeled influenza and COVID-19 [13, 14, 45–47, 99, 100]. During the COVID-19 pandemic, she has worked closely with the **NYC DOHMH** to support public health decision-making.

Our team has been at the vanguard of infectious disease modeling and prediction for the last decade. We have developed and used novel approaches for a variety of disease systems, including COVID-19 [42–47], influenza [8–15], dengue [51, 52], RSV [101], WNV [102, 103], Ebola [53, 54], and antimicrobial-resistant pathogens [55, 56]. In the context of COVID-19, our team estimated that SARS-CoV-2 transmission was primarily driven by undocumented infections in China [42] and generated one of the first US COVID-19 projections in March 2020 featured in the New York Times [82]. Routine predictions have informed federal and local public health authorities and supported the clinical trials for the Pfizer-BioNTech mRNA vaccine and Regeneron monoclonal antibody therapy. We have also evaluated the seasonality of SARS-CoV-2 [72] and developed a framework for estimating the epidemiological characteristics of variants of concern (VOC) [46]. Three investigators (Drs. Pei, Shaman, and Yang) are members of the newly established NYC Pandemic Response Institute (PRI) [104] and engage in frequent communications with NYC DOHMH in monthly meetings. We have also been involved in both theoretical and applied work looking at testing/vaccine compliance during the COVID-19 pandemic, including one multi-country study looking at the link between vaccine choices and long-term decision-making [64, 65] and another applied report on how military operations were successful at supporting the public health response in the State of New York. Given the interdisciplinary expertise, prior work, and relevant experience, our team is well positioned to successfully carry out the proposed research project.

**4. Research Agenda**

**4.1. Unmet gaps and study design**

We propose to address the grand challenge of advancing predictive intelligence for respiratory diseases in urban settings using human behavior data and theories. Currently, a few critical gaps exist limiting our understanding of the interplay between human behavior and disease transmission as well as capabilities to generate accurate operational forecasts. Our study design targets these significant unmet gaps and will produce more precise predictive models for novel respiratory pathogens such as SARS-CoV-2 at the neighborhood scale. Specifically, our approaches to address these gaps are shown in **Table 1**.

**Table 1.** Significant unmet gaps and study design of the proposed project.

|  |  |  |
| --- | --- | --- |
| **Key Research Question** | **Unmet Gaps** | **Proposed Approach** |
| *How to model reactive behavior changes under infection risk?* | • Limited availability of behavior data  • Rely on assumptions and predefined functions on behavior change  • Rarely validated using real-world data | • Employ validated behavior theories  • Surveys to collect behavior data  • Validate the behavior model using foot traffic and mask wearing data |
| *How to quantify a broad range of behaviors during public health emergencies?* | • Collect data on specific behaviors (e.g., riding subway, avoiding gathering)  • Use deterministic and discrete measures (e.g., vaccinated/unvaccinated) | • Quantify core behavioral and decision characteristics that apply more broadly  • Conduct surveys in neighborhoods to measure these characteristics |
| *What behavior features should be represented in epidemic models and how to quantify their impact on disease transmission?* | • Use behavior indicators such as online data streams, travel distance, staying-at-home time, etc.  • Indicators do not causally impact respiratory virus transmission  • Unsatisfactory performance in predictive models | • Focus on indoor contact patterns (crowdedness and dwell time) and mask wearing  • Directly modulate virus transmission rate in a nonlinear fashion  • Validate the behavior-driven model using short-term predictions |
| *How to calibrate complex behavior-driven epidemic models to generate real-time predictions?* | • Increasing model complexity to improve behavior realism  • Hard to calibrate complex models to real-world data  • Cumbersome to use in real time to support policymaking | • Develop a more realistic yet relatively parsimonious metapopulation model  • Couple with advanced Bayesian filtering techniques to produce forecasts  • Evaluate the predictive skills through retrospective forecasting |

Process-based epidemic models need to better represent the factors shaping the transmission dynamics of infectious diseases. To develop next-generation behavior-driven epidemic models, it is critical to identify behavior features that are most relevant to the transmission process. Respiratory viruses such as SARS-CoV-2 are primarily transmitted via airborne particles and droplets [105, 106]. Indoor contact patterns can significantly modulate the transmission of respiratory viruses. In this project, we will focus on three key behavior features in indoor settings where most transmission events occurred: **crowdedness**, **dwell time**, and **mask wearing**. Some of these features have been used to model COVID-19 spread [107, 108] but a representation of their reactive changes remains lacking in existing models. To capture changes of these indoor conditions, we will focus on population *mobility patterns* to different types of places and *mask wearing* during the COVID-19 pandemic.

**4.2. Data** We assembled a list of datasets tailored for the studies to be undertaken in this project.

**Disease surveillance data.** We will use daily COVID-19 surveillance data from all 42 UHF (United Hospital Fund) neighborhoods in NYC to track the transmission of SARS-CoV-2 in the city. The geographical units are shown in **Fig. 2A**. Data include neighborhood-level case, death, test, and vaccination. The surveillance data cover four COVID-19 waves in NYC – the first wave (Jan 2020 – Jun 2020); the second wave (Sep 2020 – Jun 2021); the third/Delta wave (Jul 2021 – Oct 2022); and the fourth/Omicron wave (Nov 2021 – Mar 2022). In total, over 2.3 million confirmed and probable COVID-19 cases and over 40 thousand deaths were documented as of May 1st, 2022 [3].

**Chart, histogram

Description automatically generated**

**Fig. 2.** (A) 42 UHF neighborhoods in NYC. (B) Daily numbers of visitors to four different place categories in the foot traffic data. (C) Mask-wearing survey data in NYC boroughs from CTIS surveys.

**Foot traffic and place data.** We will use fine-grained foot traffic data derived from mobile devices from January 2018 to present. This dataset was made available by SafeGraph [109], a company that aggregates anonymized location data from mobile applications to provide insights about physical places. SafeGraph data captures the movement of people between census block groups (CBGs, a geographical unit that typically contains a population of between 600 and 3,000 people), and millions of POIs across the US [110]. These data record the number of visitors in each POI during each hour. The category of POIs, square footage, and average dwell time are also available [111, 112]. We will use this dataset to measure population mobility and indoor contact patterns for 84,717 POIs in NYC. The daily numbers of visitors in four different place categories (Restaurant & Bars, Grocery & Pharmacy, Workplace, and Educational Service) since January 1st 2018 are shown in **Fig. 2B**. We will cross-validate the SafeGraph foot traffic data using another independent mobility data source, the Google COVID-19 Community Mobility Reports [113] at the US county level for different place categories. The Google Mobility reports chart movement trends over time across six categories of places including retail and recreation, groceries and pharmacies, parks, transit stations, workplaces, and residential.

**Survey on mask wearing.** Masking behavior in NYC will be derived from the Delphi/Facebook COVID-19 Trends and Impact Surveys (CTIS) (**Fig. 2C**) [67]. CTIS aims to monitor the spread and impact of the COVID-19 pandemic in the United States. The survey has run continuously since early April 2020, and about 50,000 people in the United States participate every day. The survey question on mask use is “In the past 7 days, did you wear a mask most or all of the time in public?” Results were aggregated to the county level. In this project, we will use the survey results on mask wearing in different NYC boroughs to present to capture the population-level mask wearing. Data cover the second to the fourth COVID-19 waves in NYC.

**Socioeconomic and health indicators at zip-code level.** Socioeconomic data for each neighborhood in NYC will be obtained from the 5-year American Community Survey (ACS) [114]. We will compile variables for each neighborhood that can impact human behavior, including age structure, race/ethnicity, household income, household size, education level, employment, political alignment, healthcare access and related data from NYC-specific health databases (Community Health Profiles, Environmental Health Data Portal, and EpiQuery) [115, 116]. We will additionally conduct surveys at the neighborhood level to collect complimentary behavior data (see details in Research core I).

**4.3. Research Core I: Model behavior change under infection risk**

We will conduct surveys in NYC neighborhoods to collect new data to support behavior modeling. While some of the survey and secondary data available described in Section 4.2 include valuable behavioral measures, these data may not necessarily provide indication about the specific behaviors needed for this project, such as using public transportation, wearing a mask, or dining in restaurants and bars.

To predict changes of multiple behaviors under infection risk, it is desirable to use characteristics that can describe broader behavioral patterns. To that end, we will conduct surveys in NYC neighborhoods to measure *four behavioral characteristics that* *have been validated as predictors of human behaviors across domains.* These characteristics include **temporal discounting** [57], **loss aversion** [117], **agency** [118], and **normative/non-normative decisions** [119] (see details in Section 4.3.1). As these characteristics may inform a broad range of behaviors, we will use the obtained neighborhood-level metrics for all four characteristics as variables to predict human mobility patterns and mask wearing.

**4.3.1. Conduct surveys to quantify key behavior characteristics**

Surveys will be conducted in NYC neighborhoods to quantify temporal discounting, loss aversion, agency, and normative/non-normative decisions. These surveys will result in indices of these behavior characteristics for each participant, which will then be aggregated to obtain neighborhood-level mean and variation. We will use these indices as predictors of behavior change under infection risk. In addition, to understand *how people perceive infection risk*, we will use the surveys to identify disease indictors (e.g., incidence or death) that mostly affect individual decision-making. We will additionally survey participants’ mask-wearing behavior during different COVID-19 waves to cross-validate the CTIS survey results.

**Temporal discounting.** Individuals tend to place more weight on immediate gains and losses than on future gains and losses [57]. Short-form instruments have shown considerable predictive value for explaining broad behavioral patterns [57, 120]. For example, individuals who indicate extremely high discounting rates for the future tend to also be less likely to get vaccinated, wear masks, or avoid travel options that increase risk of infection. Some work has already associated discounting with preference of transportation modes [121].

**Loss aversion.** When gains are present, individuals tend to be more risk-averse; inversely, individuals tend to be more open to risk when it may increase chances of minimizing or avoiding losses [63]. For example, most people will choose a 100% chance of gaining $90 over a 90% chance at $100, but this inverts to preferring a 90% chance of losing $100 over a 100% chance of losing $90. These patterns map onto many behaviors [64, 122]: how individuals may choose to travel for work (when staying home is an option) or when individuals might choose not to mask if they believe rates of COVID are not high in their immediate area. Having indications of these at a community level can directly inform models that predict behavior change broadly, as opposed to limiting to deterministic behavior measures, which typically only predict the behavior being measured.

**Agency.** Critically, individuals do not always have or perceive to have the same options and alternatives in their daily choices [123]. Many individuals in NYC have a diverse set of choices about behaviors such as working from home, travel options, and public transportation during the COVID-19 pandemic [124, 125]. The availability of choices as well as how realistic each may be is critical to understand and model likely behaviors. Lower agency will typically indicate minimal capacity to change behaviors (potentially galvanizing certain risk factors for at-risk communities, even where risk tolerance is optimal); greater agency allows individuals to optimize behaviors, but those choices are more dependent on risk preferences, hence the value of having all in the model [118].

**Norms and deviation.** Behaviors are often influenced by many factors unrelated to choices and values [126]. We will therefore look at aggregate measures of behaviors for communities and look at deviations within the sample at an aggregate level. For example, when we ask about mask-wearing, we will assess not only individual utilization, but how often certain individuals report behaviors different from the norms in the community. These “deviants” are critical for understanding where policy may have both major advantages as well as challenges.

**Data collection.** *A brief introduction of survey questions and resulting indices* for the four behavior characteristics is provided in **Table 2**. We aim to collect data from 1000 total participants across the five boroughs. We will randomly select 15 neighborhoods, balanced roughly by population (i.e., Brooklyn – 4 neighborhoods, Queens-4, Manhattan-3, The Bronx-2, Staten Island-2). Surveying will take place during a two-month period. All participants will receive approximately $10-$20 for completing the 30-minute survey. Neighborhood selection will be mostly at-random, but we will check general representativeness of the 15 selected before proceeding to ensure the final target sample would in theory reflect the city in terms of income, race/ethnicity, age, and proximity to POIs. Total sample distribution will aim to represent the city (e.g., ~310 participants from Brooklyn, ~165 from The Bronx), split evenly between the neighborhoods. We will aim to augment these numbers through voluntary participation in the survey after funds are expended. The co-I Dr. Ruggeri has successfully led several large-scale international survey studies [57, 63], so we anticipate the proposed surveys will proceed smoothly.

**Table 2.** Survey questions and outcomes for each behavior characteristic.

|  |  |  |
| --- | --- | --- |
| **Behavior framework** | **Survey Questions** | **Index** |
| *Temporal discounting* | A short, adaptive instrument that involves between six and nine questions (depending on participant choices) offering choices between a small immediate gain or payment versus a larger, later gain or payment [57]. | An index of discounting ranging from 1 to 15 within each neighborhood. Lower scores indicate lower discounting (e.g., better long-term decision-making, greater patience) [127]. |
| *Loss aversion* | A validated set of five simple measures presenting risky choices with gains and losses (e.g. a 90% chance of losing $100 vs a 100% chance of losing $90) [117]. | An index indicating risk tolerance in each neighborhood [128]. |
| *Agency* | An extensive list of questions for options that participants a) have available to them, and b) measure whether these are realistic. This list will include approximately 20 options, each which will be answered on a 1-5 Likert-type scale (strongly disagree to strongly agree) at two levels (a & b): available to them regularly and perceived as realistic options to them [123]. | A standardized regression-based factor score that combines both available options and the extent to which each is feasible. Scores are standardized at zero, so negative values indicate lower agency; higher values indicate greater agency. |
| *Norms and deviation* | Individual behavior and norms in the community about mobility patterns (working from home, dining out in restaurants, using public transportations, etc.) and mask-wearing (individual and community usage. | An average score for each neighborhood quantifying individual deviance from neighborhood norms for relevant behaviors. |

**Imputation for neighborhoods without surveys.** As the behavior surveys only cover part of NYC neighborhoods, we will develop a generalized linear model (GLM) to impute the indices of the four behavior characteristics in neighborhoods without surveys. Neighborhood-level socioeconomic and demographic indictors (e.g., age profile, race/ethnicity, income, education, employment, etc.) will be used as variables in the GLM. Accuracy of the imputation will be evaluated by out-of-sample cross-validation using collected data.

**4.3.2. Predict mobility pattern and mask wearing under infection risk**

We will develop *random forest models* to predict mobility pattern and mask wearing respectively under infection risk using existing and newly collected data. Random forests are machine learning models that can flexibly capture nonlinear relationships between variables and outcomes [129, 130].

To quantify human mobility patterns using the foot traffic data, we will categorize all POIs in NYC into several broader categories with distinct indoor contact patterns – *residential places, workplaces, schools, restaurants and bars, groceries and pharmacies, etc*. We will focus on the probability of individuals living in each neighborhood visiting different place categories in the same and other neighborhoods. We will additionally develop a model to predict the percentage of population in each borough wearing masks in public spaces. Denote as the probability of an individual living in location visiting place category in location at time . Define as the percentage of population in location wearing masks in public places. We will use historical foot traffic data on and CTIS survey data on to train the random forest models.

**Model variables.** The random forest models will generate short-term predictions of and under varying levels of infection risk. Input variables include **1)** indices of temporal discounting, loss aversion, agency, and normative/non-normative decisions at the neighborhood level obtained from surveys; **2)** SES and healthcare factors in each neighborhood that can affect human behaviors (such as age profile, race/ethnicity, household income, employment, health insurance, etc.); and **3)** time-varying indictors of infection risk (such as the COVID-19 incidence or death). As behavior characteristics, SES and healthcare factors can broadly impact decision-making under infection risk, we will include these neighborhood-specific *time-invariant* variables as indictors of propensity for changing behaviors. COVID-19 incidence or death will serve as *time-dependent* predictors to produce longitudinal forecasts of behaviors.

**Model validation.** We will use the foot traffic and mask wearing data in the second, third and fourth COVID-19 waves to validate the predictive models for behaviors. In particular, we will examine the forecast accuracy for one-week ahead prediction and check whether the random forest models trained for one COVID-19 wave can be transferred to predict behavior change in another wave. The predictive skills of the random forests will be systematically evaluated.

**Potential challenges and solutions.** Our team has rich experience working with behavior surveys and data analysis, so we do not anticipate issues with this work. If the SafeGraph data have strong selection bias among different place categories, we will adjust the prediction target to the relative change of visiting patterns compared to the pre-pandemic level. We will evaluate whether the survey data can support robust imputation of behavior characteristics. If necessary, additional surveys will be conducted. If the random forecast model cannot support reliable short-term predictions, other machine learning approaches (e.g., neural networks) will be tested. We will also adjust variables in the predictive model if the forecasting performance is not satisfactory.

**4.4. Research Core II: Develop epidemic models incorporating indoor contact patterns and mask-wearing**

In this research core, we aim to incorporate indoor contact patterns and mask-wearing into an epidemic model for SARS-CoV-2. Previous studies indicate that models representing crowdedness and dwell time can explain the variations of disease transmission across cities [107] but have not systematically examined how these features modulate transmission rates. We hypothesize that three features (i.e., crowdedness, dwell time, and mask-wearing) modulate the virus transmission rate in a nonlinear fashion. We will develop a metapopulation model with these behavior features fit to neighborhood-level COVID-19 data in multiple waves.

**4.4.1. Incorporate behavior-driven transmission rate into a metapopulation model**

We will develop a *data-driven metapopulation SEIRVS* (susceptible-exposed-infectious-removed-vaccinated-susceptible) model for NYC at the UHF neighborhood scale that parameterizes transmission rates using behavioral data. The model has two major components that are interconnected with each other: 1) *behavior-driven force of infection*, 2) *incidence-driven behavior model*.

**Model framework.** The metapopulation model will simulate SARS-CoV-2 transmission for different place categories in all neighborhoods, informed by foot traffic and mask-wearing data. A diagram of the model is shown in **Fig. 3.** The transmission dynamics at the neighborhood level are described by a set of differential equations:

Here , , , , are the total, susceptible, exposed, infectious, and removed population in neighborhood ; and are the effectively vaccinated susceptible population for the first and second doses in neighborhood on day , computed using neighborhood-level vaccination data and published vaccine efficacy; represents the fraction of population living in location visiting place category in location on day ; is the force of infection in place category in location , parameterized using indoor contact patterns and mask-wearing data (see details below); is the mean immunity duration, is the mean latency period, and is the mean infectious period. The model will represent underreporting such that infected individuals are observed following an ascertainment rate . We will additionally consider the reporting delay from infection acquisition to case confirmation, using information provided by NYC DOHMH [42, 45, 70]. The metapopulation model will be integrated daily and stochastically to reflect the stochastic transmission dynamics.

**Fig. 3.** A diagram of the behavior-driven transmission model. Research Core I will develop random forest models to predict behavior change using neighborhood survey data and current infection risk generated from the disease transmission model. Research Core II will study how behaviors modulate transmission rates in the metapopulation model. Research Core III will combine both the behavior and disease transmission models to generate forecasts with behavior-disease feedbacks.

**Diagram

Description automatically generated**

**Behavior-driven force of infection.** We will use the foot traffic and place data to quantify indoor crowdedness and dwell time in different POI categories in each location. Specifically, the crowdedness for place category in location is quantified using the average population density in relevant POIs, informed by the number of visitors and POI square footage. The average dwell time for place category in location , , is available from the foot traffic data. We will further use the CTIS survey to inform the population-level masking (percentage of survey participants who wore masks in public), , in location . We will additionally use a parameter to represent the aggregated effect of other unmeasured interventions (e.g., improved cleaning and ventilation) on disease transmission. Recall that represents the probability of an individual living in location visiting place category in location on day . We define the force of infection in place category in location as

where and are the infectious and total population visiting place category in location , which can be computed using the mobility pattern . The number of new infected individuals in location on day is , where is the susceptible population in location on day .

The function shares the same functional form across all locations and depicts how crowdedness , dwell time , mask wearing , and other unmeasured interventions modulate the transmission rate in place category . For model parsimony, we assume the effects of these factors are independent and multiplicative – . Using the functional form of , we can translate indoor contact patterns and mask-wearing to transmission rates.

We will test a base model form , where , , and are free parameters that will be determined by model fitting. In addition to this base form, a diverse set of functional forms will be tested. We will test the dependence of the transmission rate on crowdedness and dwell time as a *linear*, *power-law*, *piecewise linear,* or *sigmoid function* (S-shaped curves with threshold and saturation effects). Combinations of these functional forms will be used to define and tested through model fitting.

**4.4.2. Model selection and validation using COVID-19 data**

We will select the best functional form by comparing model fitting to the neighborhood-level COVID-19 data. During the COVID-19 pandemic, model parameters (e.g., the transmission rate and ascertainment rate) varied over time. As such, we are unlikely to identify a set of fixed parameters that can fit the entire pandemic curve. To account for time-varying parameters, we will calibrate the model sequentially to the most recent COVID-19 data and compare the model performance using short-term predictions.

**Model fitting.** For each functional form of , we will apply a Bayesian filtering algorithm that is applicable to high-dimensional models – the ensemble adjustment Kalman filter (EAKF) (xxx) -- to estimate model parameters and variables. In our prior work, we have successfully used similar frameworks to estimate the fraction of undocumented SARS-CoV-2 infections in China using a metapopulation model for 375 cities [42] and key epidemiological parameters in an agent-based model for healthcare-associated infections [55, 56]. The EAKF is a sequential data assimilation method that is widely used in numerical weather prediction [131]. The EAKF maintains an ensemble of model parameters and variables, acting as samples from their distributions. The framework allows the representation of the uncertainty in model state.

Given the high dimensionality of the metapopulation model, alternative filtering techniques such as the particle filtering (or sequential Monte Carlo) [132] and iterated filtering [133, 134] require a prohibitive number of particles to fully explore parameter space [135]. Previous studies indicate that a few hundred ensemble members in the EAKF can support accurate parameter inference in high-dimensional disease transmission models [10, 43]. The EAKF is optimal for linear systems; however, it also has satisfactory performance for disease transmission models with weak nonlinear components [12], such as the model to be developed in this project, and high-dimension, strongly nonlinear models of the atmosphere [131].

In model fitting, the population density and dwell time for each place category and mask-wearing in the model will be informed by the real-world foot traffic data and the CTIS survey data. We will fit the model to the neighborhood-level COVID-19 data. Model parameters and variables will be updated daily. An illustration of the model-data assimilation framework is shown in **Fig. 4**. This sequential update procedure can flexibly capture time-varying parameters [131, 136, 137]. Before applying the inference framework to the real-world COVID-19 data, we will use model-simulated outbreaks to systematically evaluate its ability to identify parameters. Specifically, we will generate outbreaks using a given set of parameters in the function and examine if the inference framework can identify these parameters using error-laden observations.

Diagram

Description automatically generated

**Fig. 4.** A diagram of the model-data assimilation system to generate real-time forecasts. The dotted red line is the unobserved “truth” and red crosses are noisy observations. We use a model (blue line) to represent disease transmission. An ensemble of models with different model states are used to represent the uncertainty (black circle). Once new observations become available, prior model states are updated to posterior states (closer to observation). The updated models are used to generate forecasts.

**Model selection.** To compare the performance of different functional forms, we will fit models with different functions to the neighborhood level COVID-19 data for the second, third and fourth waves in NYC separately. We will calibrate the model to observations until a given day and integrate the model into the next week to generate short-term forecasts. The indoor contact patterns and mask-wearing in the next week after day will be set by the real-world data. In real-time forecasting, these future behavior data after day are not available; however, for model comparison here, we will use the real-world human behavior data to test the validity of the nonlinear functions . The best model with the lowest short-term forecast errors will be selected to generate longer-term forecasts in Research Core III.

**Potential challenges and solutions.** We have used the proposed model-inference framework extensively, so we do not anticipate issues implementing these methods. However, a few challenges may arise in the study. If the inference algorithm cannot fully identify model parameters, we will assign key epidemiological parameters (e.g., latency period and infectious period) to reduce the number of unknown parameters. We can also test other more computationally demanding methods such as the approximate Bayesian computation [138, 139] and Markov Chain Monte Carlo approaches [140, 141].

**4.5. Research Core III: Develop and validate a behavior-driven predictive model**

In Research Core III, we propose to combine the risk-driven behavior model developed in Research Core I and the behavior-driven metapopulation model validated in Research Core II into a predictive model for COVID-19. This model will represent behavior-disease feedbacks and be used to generate real-time predictions. To develop a reliable forecasting model for operational use, the predictive skill of the system needs to be systematically vetted. Skill measures include both *forecast accuracy* and *forecast uncertainty*. We will couple the predictive model with a data assimilation algorithm and generate retrospective longer-term forecasts for the second, third and fourth COVID-19 waves in NYC. Predictive skill at the neighborhood level will be thoroughly evaluated using historical data.

**4.5.1. Develop a predictive metapopulation model with behavior-disease feedback**

To generate retrospective forecasts for COVID-19, we will perform a two-step procedure. First, we will calibrate the behavior-driven metapopulation model using the most recent COVID-19 data, foot traffic data, and mask-wearing survey data. This process will ensure the model state agrees with current observations. Second, we will integrate the calibrated model into the future to generate forecasts. We will also alternately run the disease transmission model validated in Research Core II and the risk-driven behavior model developed in Research Core I to capture the behavior-disease feedbacks. A diagram of the forecasting process is shown in **Fig. 5**.

Diagram

Description automatically generated

**Fig. 5.** Workflow of the forecasting system with behavior-disease feedbacks.

**Model calibration.** In order to calibrate the model to real-world observations, model parameters and variables need to be frequently updated using recent data. Here, we will use the EAKF for model calibration (see details in Section 4.4.2). Similar to numerical weather prediction, we will generate an ensemble (e.g., 300) of model trajectories with different initial conditions, which can support probabilistic forecasts with associated uncertainty. We will use the neighborhood level COVID-19 data, behavior data on indoor contact patterns, and mask-wearing data to constrain the model. Model parameters and variables will be updated daily once new observations become available. This procedure can capture time-varying parameters and alleviate the potential adverse effect of model misspecification.

**Forecasting with behavior-disease feedbacks.** We will integrate the calibrated model into the future and alternately run both the calibrated model and the behavior change model to generate forecasts. For the latter, for each day in the future, we will integrate the model as follows: **1)** integrate the metapopulation transmission model using current mobility pattern and mask-wearing data ; **2)** use the random forecast model to generate mobility pattern and mask-wearing estimates on day given current disease situation, and update population density in different place categories using . This process will be repeated daily to create the behavior-disease feedback, potentially resulting in a *self-limiting* effect on disease transmission. Note that we will use the disease indictors that most affect human behaviors identified in surveys to represent individuals’ perception of risk. We will generate an ensemble of model predictions to produce probabilistic forecasts.

**4.5.2. Retrospective forecasting for SARS-CoV-2 in NYC**

**Retrospective forecasting.** Retrospective forecasts will be generated weekly for the second, third and fourth COVID-19 waves in NYC. For the third (Delta) and fourth (Omicron) waves, substantial immunity evasion led to increased population susceptibility [142, 143]. We will use estimated population susceptibility in NYC [46, 47] to initiate the metapopulation model. For each wave, we will focus on one- to four-week ahead predictions for confirmed case, hospitalization, and deaths. We will additionally predict longer-term targets including peak timing and peak intensity of cases, hospitalizations, and deaths. For predicting hospitalizations and deaths, we will map the estimated infection to those targets using most recent infection-hospitalization and infection-fatality rates as well as the delays from case confirmation to hospitalization and death. To generate forecasts, we will integrate the calibrated model into the future with the behavior-disease feedback. For comparison, a baseline forecast without the behavior-disease feedback will also be generated, assuming transmission rates and human mobility remains unchanged in the future. Such “status quo” model was widely used in existing predictive models for COVID-19 [144].

**Predictive skill evaluation.** Forecasts will be evaluated for both prediction accuracy and uncertainty. **1).** We will use mean absolute errors (MAE) to quantify the accuracy of point prediction. To measure the quality of probabilistic prediction, we will use the log-score, which has been widely employed to evaluate real-time influenza forecasts [22–25], as well as the weighted interval score (WIS) [145], which is currently used to evaluate COVID-19 forecasts [146]. The log-score is calculated as the logarithm (base e) of the forecast probability assigned to an interval around the observed target. The WIS is a proper forecast evaluation metric designed for predictions in an interval format when the full prediction distribution is not available [145]. **2).** The uncertainty of forecasts will be evaluated using confidence interval (CI) coverage and reliability [147]. For an ideal probabilistic forecast, the uncertainty of predictions should be well calibrated to reflect the uncertainty in real-world outcomes. For instance, the 90% (50%) predicted CI should cover 90% (50%) observed targets in a reliable forecast.

The performance of the behavior-driven forecasts and baseline forecasts will be compared using the above metrics. Forecasts will be evaluated at both neighborhood and city levels. We will additionally examine the forecast horizon for which reliable predictions are feasible and the performance at different regimes of the epidemic (onset, exponential growth, peak, and decline).

**Potential challenges and solutions.** We expect a smooth implementation of the proposed projects given our prior work on real-time disease forecasting. If the performance of the predictive model is not satisfactory, alternate techniques will be applied. For instance, if the prediction CI is too narrow, an inflation process can be used to expand the ensemble spread during data assimilation. We can also update the number of place categories to capture more detailed human activities, or add additional parameters to reflect heterogeneous transmission rate not explained by other factors.

**4.6. Project Management and Timeline**

The PI, Dr. Sen Pei, will lead and oversee all aspects of this project. His responsibilities include: 1) lead the research agenda and coordinate monthly meetings among all investigators to facilitate communications across disciplines; 2) participate in all research cores and lead the studies in cores II and III; and 3) supervise a postdoctoral scientist working on model development, data analysis, and manuscript preparation. Dr. Kai Ruggeri will lead the studies in research core I and contribute to the integration of the behavior model into the disease transmission model. Dr. Jeffrey Shaman will assist with the design of the model-data assimilation system, help with the evaluation of real-time forecasting system, and liaise with the NYC PRI and other stakeholders. Dr. Wan Yang will provide expertise on indoor virus transmission, participate in the development of the behavior-driven epidemic model, and connect with NYC DOHMH. Investigators will meet regularly to the coordinate research approach, share research findings, and discuss results. Subgroups working on specific research cores will also meeting regularly to discuss progress and technical problems.

Four years of funding are requested for this project in order to address the research objectives thoroughly and to tackle any unforeseen issues. A timeline of work for all research cores is summarized below.

**Table 3.** Project Timeline

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Year 1 | Year 2 | Year 3 | Year 4 |
| Research Core I: Conduct surveys |  |  |  |  |
| Research Core I: Model reactive behavior |  |  |  |  |
| Research Core II: Behavior-disease modeling |  |  |  |  |
| Research Core III: Retrospective forecasting |  |  |  |  |

**5. Broader Impacts**

**5.1. Benefits to Public Health.** The project will have long-term benefits for capacity building in outbreak preparedness and response. Findings in Research Core I will deepen our understanding on the nonlinear modulation of indoor contact patterns on respiratory virus transmission, which can inform mitigation strategies in indoor settings. Incorporation of human behavior into metapopulation models can produce more realistic yet parsimonious epidemiological models that can be readily deployed in real time for rapid response. The forecasting system can be used to evaluate the impact of government mandates such as restrictions on business operating capacity and masking mandates in real time tailored to local disease situations. In addition, the behavior models can support the assessment of the roles of risk communication and situation awareness in voluntary behavior change. Together, these efforts will enable reliable predictions of emerging respiratory diseases in densely populated urban settings, which can reduce disease risk, save lives, and contribute to economic stability. Our focus on NYC can additionally address the issue of health disparities and highlight the differential impacts of disease outbreaks on vulnerable communities.

**5.2. Education and Training.** The research activities will provide in-depth training of a postdoctoral scientist working at the interface between infectious disease modeling and behavior science (see *Postdoctoral Mentoring Plan*). It is imperative that a new generation of scientists be trained with exposure to the methods and systems developed through this project. Such training will ensure that disease forecast continues to evolve and improve, and that a new cohort of scientists will be in place to lead the disease modeling field in the future. The postdoctoral scientist will be responsible for building the core transmission model and forecasting system. As a result, the postdoctoral scientist will be trained in the state-of-the-art techniques in behavior-driven real-time disease forecasting and gain a highly interdisciplinary skillset needed to tackle pressing questions in outbreak preparedness and response. The project will support the postdoctoral scientist to attend multiple scientific conferences to disseminate findings, which will cultivate public speaking skills and provide opportunities for networking and career advancement. Based within a public health school, the PI and co-PIs will present research findings in department and school seminars. The seminars will effectively promote the framework of considering human behavior in infectious disease models and serve as a vehicle to disseminate results from this project to next-generation modelers. In addition, they will educate a broader audience of public health students who do not directly work on infectious disease modeling.

**5.3. Community outreach.** Throughout the COVID-19 pandemic, we have worked closely with public health officials, policy makers, and private sectors to support COVID-19 response. At the national and state level, we have publicly disseminated our real-time weekly COVID-19 forecasts at the county scale [148] – we have shared these forecasts with the White House Task Force, CDC, CSTE, and State of New York, among others, and supported the clinical trials for the Pfizer-BioNTech mRNA vaccine and Regeneron monoclonal antibody therapeutic. We have also contributed to the CDC COVID-19 forecast hub, the scenario hub, and the CSTE forecasting group. Locally, we have worked closely with NYC DOHMH and its contact tracing team. Drs. Pei, Shaman, and Yang are currently the core members of the newly established NYC Pandemic Response Institute (PRI). Leveraging these existing connections, we will disseminate advances in behavior-driven modeling and obtain feedbacks to improve the predictive models developed in this project. We will also present in seminars within the NYC PRI to promote the use of modeling in supporting public health decision-making. Communications with end-users of mathematical disease models is essential to make these models useful in real-world applications.

**6. Results from Prior NSF Support**

Co-PIs Drs. Jeffrey Shaman and Wan Yang. DMS-2027369 “RAPID: Inference, Forecasting, and Intervention Modeling of COVID-19” 04/15/2020-06/30/2021, $198,643. Shaman was PI for this project. Yang was co-PI for this project.

**Intellectual Merit**: This project leveraged state-of-the-art dynamic modeling and model-inference approaches to study and forecast the SARS-CoV-2 outbreak. We used a data-driven model-inference approach to simulate the pandemic at county-scale in the US during 2020 and estimate critical, time-varying epidemiological properties underpinning the dynamics of the virus [43]. We also developed a meta-population network model-inference system, resolved at neighborhood spatial scales, to estimate the underlying SARS-CoV-2 infection rate in NYC during the 2020 spring pandemic wave using available case, mortality, and mobility data and applied this network model-inference system to estimate the infection-fatality risk for five age groups and all ages overall, from Mar 1 to Jun 6, 2020 [45]. Additionally, we evaluated the effects of socioeconomic factors on COVID-19 infection. We examined 1) the extent to which the variability in ZIP code-level case positivity can be explained by aggregate markers of socioeconomic status and daily change in mobility; and 2) the extent to which daily changes in mobility independently predicts case positivity [149]. We also evaluated the extent to which the geographical variability of mortality in the US can be explained by population characteristics that predate the epidemic [150]; and we tested the hypothesis that lower air temperature and lower humidity are associated with increased SARS-CoV-2 transmission by estimating the reproduction number (Rt) in 913 counties from Mar 15 to Aug 31, 2020 and calculating the fraction of Rt attributable to these meteorological conditions [72].

**Broader Impact**: Work from project DMS-2027369 provided modeling, analysis, projections, and public health support to a variety of governmental agencies, including New York City Department of Health and Mental Hygiene, the New York City Mayor’s Office, New York City Health+Hospital, the states of New York, Maryland, Vermont, California and Hawaii, the White House Task Force, and the Centers for Disease Control and Prevention. This support included presentation of findings, advice and consultation with public health officials, delivery of regular (weekly or twice weekly) projections of future cases, hospitalization and deaths, and counterfactual simulations of various control measures and policies. Two Associate Research Scientists and one postdoc were supported under this grant. Further, although not funded under this grant, two first-year PhD students did rotations with Dr. Shaman during the 2020-21 academic year. One assessed the association between early country-level testing capacity and later COVID-19 mortality outcomes [151]; the second investigated socioeconomic disparities in SARS-CoV-2 serological testing and positivity in New York City [152].

Drs. Sen Pei and Kai Ruggeri have no prior NSF support.

**7. Applicability to the RAISE Guidelines**

***Scientific advances lie in great part outside the scope of a single program or discipline, such that substantial funding support from more than one program or discipline is necessary.***

Our project requires transdisciplinary expertise in mathematical, behavioral, and epidemiological sciences. We will draw from well-established behavior theories on decision-making and represent these theories mathematically in epidemic models. The focus on respiratory diseases will additionally necessitate expertise in infectious disease epidemiology. Such cross-cutting research lies outside the traditional scope of a single program or discipline.

***Lines of research promise transformational advances.***

Our project will lead to transformational advances by producing more realistic yet parsimonious epidemiological models that can be more functionally deployed to inform public health interventions. Specifically, we will 1) incorporate key behavior features that directly impact disease transmission, 2) bring behavior theories on decision-making into epidemiological models, and 3) systematically evaluate the predictive skill of these developed behavior-driven models for influenza and COVID-19.

***Prospective discoveries reside at the interfaces of disciplinary boundaries that may not be recognized through traditional review or co-review.***

Our project will promote the deep integration of mathematical, epidemiological, and behavioral sciences. Prospective transdisciplinary discoveries include: 1) a quantitative understanding of how human behavior directly modulates disease transmission; 2) a validated behavioral model for decision-making under infection risk; and 3) a behavior-driven model with demonstrated predictive skill for simulating and forecasting respiratory disease transmission. To date, mathematical models incorporating behavior have favored complex model forms with detailed human behavior (e.g., agent-based models with assumed individual behavior dynamics). However, complex models can be over-parameterized and are often cumbersome and inferior in operational use. Our focus on the tradeoff between model realism and parsimony may not be recognized through traditional review or co-review.

**Reference**

1. WHO 14.9 million excess deaths associated with the COVID-19 pandemic in 2020 and 2021. https://www.who.int/news/item/05-05-2022-14.9-million-excess-deaths-were-associated-with-the-covid-19-pandemic-in-2020-and-2021

2. WHO Swine flu | WHO emergency situation overview. https://www.who.int/emergencies/situations/influenza-a-(h1n1)-outbreak

3. COVID-19: Data Trends and Totals - NYC Health. https://www1.nyc.gov/site/doh/covid/covid-19-data-totals.page

4. Bavel JJV, Baicker K, Boggio PS, Capraro V, Cichocka A, Cikara M, Crockett MJ, Crum AJ, Douglas KM, Druckman JN, Drury J, Dube O, Ellemers N, Finkel EJ, Fowler JH, Gelfand M, Han S, Haslam SA, Jetten J, Kitayama S, Mobbs D, Napper LE, Packer DJ, Pennycook G, Peters E, Petty RE, Rand DG, Reicher SD, Schnall S, Shariff A, Skitka LJ, Smith SS, Sunstein CR, Tabri N, Tucker JA, Linden S van der, Lange P van, Weeden KA, Wohl MJA, Zaki J, Zion SR, Willer R (2020) Using social and behavioural science to support COVID-19 pandemic response. *Nature Human Behaviour*, 4(5):460–471. https://doi.org/10.1038/s41562-020-0884-z

5. Reed-Thryselius S, Fuss L, Rausch D (2022) The Relationships Between Socioeconomic Status, COVID-19 Risk Perceptions, and the Adoption of Protective Measures in a Mid-Western City in the United States. *Journal of Community Health*, https://doi.org/10.1007/s10900-022-01070-y

6. Metcalf CJE, Morris DH, Park SW (2020) Mathematical models to guide pandemic response. *Science*, 369(6502):368–369. https://doi.org/10.1126/science.abd1668

7. Lofgren ET, Halloran ME, Rivers CM, Drake JM, Porco TC, Lewis B, Yang W, Vespignani A, Shaman J, Eisenberg JNS, Eisenberg MC, Marathe M, Scarpino SV, Alexander KA, Meza R, Ferrari MJ, Hyman JM, Meyers LA, Eubank S (2014) Mathematical models: A key tool for outbreak response. *Proceedings of the National Academy of Sciences*, 111(51):18095–18096. https://doi.org/10.1073/pnas.1421551111

8. Shaman J, Karspeck A (2012) Forecasting seasonal outbreaks of influenza. *Proceedings of the National Academy of Sciences*, 109(50):20425–20430. https://doi.org/10.1073/pnas.1208772109

9. Shaman J, Karspeck A, Yang W, Tamerius J, Lipsitch M (2013) Real-time influenza forecasts during the 2012–2013 season. *Nature Communications*, 4(1):2837. https://doi.org/10.1038/ncomms3837

10. Pei S, Kandula S, Yang W, Shaman J (2018) Forecasting the spatial transmission of influenza in the United States. *Proceedings of the National Academy of Sciences*, 115(11):2752–2757. https://doi.org/10.1073/pnas.1708856115

11. Pei S, Shaman J (2017) Counteracting structural errors in ensemble forecast of influenza outbreaks. *Nature Communications*, 8(1):925. https://doi.org/10.1038/s41467-017-01033-1

12. Yang W, Karspeck A, Shaman J (2014) Comparison of Filtering Methods for the Modeling and Retrospective Forecasting of Influenza Epidemics. *PLOS Computational Biology*, 10(4):e1003583. https://doi.org/10.1371/journal.pcbi.1003583

13. Yang W, Olson DR, Shaman J (2016) Forecasting Influenza Outbreaks in Boroughs and Neighborhoods of New York City. *PLOS Computational Biology*, 12(11):e1005201. https://doi.org/10.1371/journal.pcbi.1005201

14. Yang W, Cowling BJ, Lau EHY, Shaman J (2015) Forecasting Influenza Epidemics in Hong Kong. *PLOS Computational Biology*, 11(7):e1004383. https://doi.org/10.1371/journal.pcbi.1004383

15. Kandula S, Yamana T, Pei S, Yang W, Morita H, Shaman J (2018) Evaluation of mechanistic and statistical methods in forecasting influenza-like illness. *Journal of The Royal Society Interface*, 15(144):20180174. https://doi.org/10.1098/rsif.2018.0174

16. Osthus D, Moran KR (2021) Multiscale influenza forecasting. *Nature Communications*, 12(1):2991. https://doi.org/10.1038/s41467-021-23234-5

17. Osthus D, Gattiker J, Priedhorsky R, Valle SYD (2019) Dynamic Bayesian Influenza Forecasting in the United States with Hierarchical Discrepancy (with Discussion). *Bayesian Analysis*, 14(1):261–312. https://doi.org/10.1214/18-BA1117

18. Farrow DC, Brooks LC, Hyun S, Tibshirani RJ, Burke DS, Rosenfeld R (2017) A human judgment approach to epidemiological forecasting. *PLOS Computational Biology*, 13(3):e1005248. https://doi.org/10.1371/journal.pcbi.1005248

19. Brooks LC, Farrow DC, Hyun S, Tibshirani RJ, Rosenfeld R (2018) Nonmechanistic forecasts of seasonal influenza with iterative one-week-ahead distributions. *PLOS Computational Biology*, 14(6):e1006134. https://doi.org/10.1371/journal.pcbi.1006134

20. Ray EL, Reich NG (2018) Prediction of infectious disease epidemics via weighted density ensembles. *PLOS Computational Biology*, 14(2):e1005910. https://doi.org/10.1371/journal.pcbi.1005910

21. Reich NG, McGowan CJ, Yamana TK, Tushar A, Ray EL, Osthus D, Kandula S, Brooks LC, Crawford-Crudell W, Gibson GC, Moore E, Silva R, Biggerstaff M, Johansson MA, Rosenfeld R, Shaman J (2019) Accuracy of real-time multi-model ensemble forecasts for seasonal influenza in the U.S. *PLOS Computational Biology*, 15(11):e1007486. https://doi.org/10.1371/journal.pcbi.1007486

22. Biggerstaff M, Alper D, Dredze M, Fox S, Fung IC-H, Hickmann KS, Lewis B, Rosenfeld R, Shaman J, Tsou M-H, Velardi P, Vespignani A, Finelli L, for the Influenza Forecasting Contest Working Group (2016) Results from the centers for disease control and prevention’s predict the 2013–2014 Influenza Season Challenge. *BMC Infectious Diseases*, 16(1):357. https://doi.org/10.1186/s12879-016-1669-x

23. Biggerstaff M, Johansson M, Alper D, Brooks LC, Chakraborty P, Farrow DC, Hyun S, Kandula S, McGowan C, Ramakrishnan N, Rosenfeld R, Shaman J, Tibshirani R, Tibshirani RJ, Vespignani A, Yang W, Zhang Q, Reed C (2018) Results from the second year of a collaborative effort to forecast influenza seasons in the United States. *Epidemics*, 24:26–33. https://doi.org/10.1016/j.epidem.2018.02.003

24. Cramer EY, Ray EL, Lopez VK, Bracher J, Brennen A, Rivadeneira AJC, Gerding A, Gneiting T, House KH, Huang Y, Jayawardena D, Kanji AH, Khandelwal A, Le K, Mühlemann A, Niemi J, Shah A, Stark A, Wang Y, Wattanachit N, Zorn MW, Gu Y, Jain S, Bannur N, Deva A, Kulkarni M, Merugu S, Raval A, Shingi S, Tiwari A, White J, Abernethy NF, Woody S, Dahan M, Fox S, Gaither K, Lachmann M, Meyers LA, Scott JG, Tec M, Srivastava A, George GE, Cegan JC, Dettwiller ID, England WP, Farthing MW, Hunter RH, Lafferty B, Linkov I, Mayo ML, Parno MD, Rowland MA, Trump BD, Zhang-James Y, Chen S, Faraone SV, Hess J, Morley CP, Salekin A, Wang D, Corsetti SM, Baer TM, Eisenberg MC, Falb K, Huang Y, Martin ET, McCauley E, Myers RL, Schwarz T, Sheldon D, Gibson GC, Yu R, Gao L, Ma Y, Wu D, Yan X, Jin X, Wang Y-X, Chen Y, Guo L, Zhao Y, Gu Q, Chen J, Wang L, Xu P, Zhang W, Zou D, Biegel H, Lega J, McConnell S, Nagraj VP, Guertin SL, Hulme-Lowe C, Turner SD, Shi Y, Ban X, Walraven R, Hong Q-J, Kong S, Walle A van de, Turtle JA, Ben-Nun M, Riley S, Riley P, Koyluoglu U, DesRoches D, Forli P, Hamory B, Kyriakides C, Leis H, Milliken J, Moloney M, Morgan J, Nirgudkar N, Ozcan G, Piwonka N, Ravi M, Schrader C, Shakhnovich E, Siegel D, Spatz R, Stiefeling C, Wilkinson B, Wong A, Cavany S, España G, Moore S, Oidtman R, Perkins A, Kraus D, Kraus A, Gao Z, Bian J, Cao W, Ferres JL, Li C, Liu T-Y, Xie X, Zhang S, Zheng S, Vespignani A, Chinazzi M, Davis JT, Mu K, Piontti AP y, Xiong X, Zheng A, Baek J, Farias V, Georgescu A, Levi R, Sinha D, Wilde J, Perakis G, Bennouna MA, Nze-Ndong D, Singhvi D, Spantidakis I, Thayaparan L, Tsiourvas A, Sarker A, Jadbabaie A, Shah D, Penna ND, Celi LA, Sundar S, Wolfinger R, Osthus D, Castro L, Fairchild G, Michaud I, Karlen D, Kinsey M, Mullany LC, Rainwater-Lovett K, Shin L, Tallaksen K, Wilson S, Lee EC, Dent J, Grantz KH, Hill AL, Kaminsky J, Kaminsky K, Keegan LT, Lauer SA, Lemaitre JC, Lessler J, Meredith HR, Perez-Saez J, Shah S, Smith CP, Truelove SA, Wills J, Marshall M, Gardner L, Nixon K, Burant JC, Wang L, Gao L, Gu Z, Kim M, Li X, Wang G, Wang Y, Yu S, Reiner RC, Barber R, Gakidou E, Hay SI, Lim S, Murray CJL, Pigott D, Gurung HL, Baccam P, Stage SA, Suchoski BT, Prakash BA, Adhikari B, Cui J, Rodríguez A, Tabassum A, Xie J, Keskinocak P, Asplund J, Baxter A, Oruc BE, Serban N, Arik SO, Dusenberry M, Epshteyn A, Kanal E, Le LT, Li C-L, Pfister T, Sava D, Sinha R, Tsai T, Yoder N, Yoon J, Zhang L, Abbott S, Bosse NI, Funk S, Hellewell J, Meakin SR, Sherratt K, Zhou M, Kalantari R, Yamana TK, Pei S, Shaman J, Li ML, Bertsimas D, Lami OS, Soni S, Bouardi HT, Ayer T, Adee M, Chhatwal J, Dalgic OO, Ladd MA, Linas BP, Mueller P, Xiao J, Wang Y, Wang Q, Xie S, Zeng D, Green A, Bien J, Brooks L, Hu AJ, Jahja M, McDonald D, Narasimhan B, Politsch C, Rajanala S, Rumack A, Simon N, Tibshirani RJ, Tibshirani R, Ventura V, Wasserman L, O’Dea EB, Drake JM, Pagano R, Tran QT, Ho LST, Huynh H, Walker JW, Slayton RB, Johansson MA, Biggerstaff M, Reich NG (2021) Evaluation of individual and ensemble probabilistic forecasts of COVID-19 mortality in the US. :2021.02.03.21250974. https://doi.org/10.1101/2021.02.03.21250974

25. McGowan CJ, Biggerstaff M, Johansson M, Apfeldorf KM, Ben-Nun M, Brooks L, Convertino M, Erraguntla M, Farrow DC, Freeze J, Ghosh S, Hyun S, Kandula S, Lega J, Liu Y, Michaud N, Morita H, Niemi J, Ramakrishnan N, Ray EL, Reich NG, Riley P, Shaman J, Tibshirani R, Vespignani A, Zhang Q, Reed C (2019) Collaborative efforts to forecast seasonal influenza in the United States, 2015–2016. *Scientific Reports*, 9(1):683. https://doi.org/10.1038/s41598-018-36361-9

26. Bedson J, Skrip LA, Pedi D, Abramowitz S, Carter S, Jalloh MF, Funk S, Gobat N, Giles-Vernick T, Chowell G, Almeida JR de, Elessawi R, Scarpino SV, Hammond RA, Briand S, Epstein JM, Hébert-Dufresne L, Althouse BM (2021) A review and agenda for integrated disease models including social and behavioural factors. *Nature Human Behaviour*, 5(7):834–846. https://doi.org/10.1038/s41562-021-01136-2

27. Salon D, Conway MW, Capasso da Silva D, Chauhan RS, Derrible S, Mohammadian A (Kouros), Khoeini S, Parker N, Mirtich L, Shamshiripour A, Rahimi E, Pendyala RM (2021) The potential stickiness of pandemic-induced behavior changes in the United States. *Proceedings of the National Academy of Sciences*, 118(27):e2106499118. https://doi.org/10.1073/pnas.2106499118

28. Wang Z, Andrews MA, Wu Z-X, Wang L, Bauch CT (2015) Coupled disease–behavior dynamics on complex networks: A review. *Physics of Life Reviews*, 15:1–29. https://doi.org/10.1016/j.plrev.2015.07.006

29. Fenichel EP, Castillo-Chavez C, Ceddia MG, Chowell G, Parra PAG, Hickling GJ, Holloway G, Horan R, Morin B, Perrings C, Springborn M, Velazquez L, Villalobos C (2011) Adaptive human behavior in epidemiological models. *Proceedings of the National Academy of Sciences*, 108(15):6306–6311. https://doi.org/10.1073/pnas.1011250108

30. Weitz JS, Park SW, Eksin C, Dushoff J (2020) Awareness-driven behavior changes can shift the shape of epidemics away from peaks and toward plateaus, shoulders, and oscillations. *Proceedings of the National Academy of Sciences*, https://doi.org/10.1073/pnas.2009911117

31. Granell C, Gómez S, Arenas A (2013) Dynamical Interplay between Awareness and Epidemic Spreading in Multiplex Networks. *Physical Review Letters*, 111(12):128701. https://doi.org/10.1103/PhysRevLett.111.128701

32. Ferguson NM, Cummings DAT, Fraser C, Cajka JC, Cooley PC, Burke DS (2006) Strategies for mitigating an influenza pandemic. *Nature*, 442(7101):448–452. https://doi.org/10.1038/nature04795

33. Epstein JM, Parker J, Cummings D, Hammond RA (2008) Coupled Contagion Dynamics of Fear and Disease: Mathematical and Computational Explorations. *PLOS ONE*, 3(12):e3955. https://doi.org/10.1371/journal.pone.0003955

34. Epstein JM, Goedecke DM, Yu F, Morris RJ, Wagener DK, Bobashev GV (2007) Controlling Pandemic Flu: The Value of International Air Travel Restrictions. *PLOS ONE*, 2(5):e401. https://doi.org/10.1371/journal.pone.0000401

35. Fast SM, González MC, Wilson JM, Markuzon N (2015) Modelling the propagation of social response during a disease outbreak. *Journal of The Royal Society Interface*, 12(104):20141105. https://doi.org/10.1098/rsif.2014.1105

36. Bosetti P (2019) Social dynamics and behavioral response during health threats.

37. Nouvellet P, Bhatia S, Cori A, Ainslie KEC, Baguelin M, Bhatt S, Boonyasiri A, Brazeau NF, Cattarino L, Cooper LV, Coupland H, Cucunuba ZM, Cuomo-Dannenburg G, Dighe A, Djaafara BA, Dorigatti I, Eales OD, Elsland SL van, Nascimento FF, FitzJohn RG, Gaythorpe KAM, Geidelberg L, Green WD, Hamlet A, Hauck K, Hinsley W, Imai N, Jeffrey B, Knock E, Laydon DJ, Lees JA, Mangal T, Mellan TA, Nedjati-Gilani G, Parag KV, Pons-Salort M, Ragonnet-Cronin M, Riley S, Unwin HJT, Verity R, Vollmer MAC, Volz E, Walker PGT, Walters CE, Wang H, Watson OJ, Whittaker C, Whittles LK, Xi X, Ferguson NM, Donnelly CA (2021) Reduction in mobility and COVID-19 transmission. *Nature Communications*, 12(1):1090. https://doi.org/10.1038/s41467-021-21358-2

38. Kissler SM, Kishore N, Prabhu M, Goffman D, Beilin Y, Landau R, Gyamfi-Bannerman C, Bateman BT, Snyder J, Razavi AS, Katz D, Gal J, Bianco A, Stone J, Larremore D, Buckee CO, Grad YH (2020) Reductions in commuting mobility correlate with geographic differences in SARS-CoV-2 prevalence in New York City. *Nature Communications*, 11(1):4674. https://doi.org/10.1038/s41467-020-18271-5

39. Engle S, Stromme J, Zhou A (2020) Staying at Home: Mobility Effects of COVID-19. https://doi.org/10.2139/ssrn.3565703

40. Badr HS, Du H, Marshall M, Dong E, Squire MM, Gardner LM (2020) Association between mobility patterns and COVID-19 transmission in the USA: a mathematical modelling study. *The Lancet Infectious Diseases*, 20(11):1247–1254. https://doi.org/10.1016/S1473-3099(20)30553-3

41. Badr HS, Gardner LM (2020) Limitations of using mobile phone data to model COVID-19 transmission in the USA. *The Lancet Infectious Diseases*, 0(0)https://doi.org/10.1016/S1473-3099(20)30861-6

42. Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, Shaman J (2020) Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science*, 368(6490):489–493. https://doi.org/10.1126/science.abb3221

43. Pei S, Yamana TK, Kandula S, Galanti M, Shaman J (2021) Burden and characteristics of COVID-19 in the United States during 2020. *Nature*, 598(7880):338–341. https://doi.org/10.1038/s41586-021-03914-4

44. Pei S, Kandula S, Shaman J (2020) Differential effects of intervention timing on COVID-19 spread in the United States. *Science Advances*, 6(49):eabd6370. https://doi.org/10.1126/sciadv.abd6370

45. Yang W, Kandula S, Huynh M, Greene SK, Wye GV, Li W, Chan HT, McGibbon E, Yeung A, Olson D, Fine A, Shaman J (2021) Estimating the infection-fatality risk of SARS-CoV-2 in New York City during the spring 2020 pandemic wave: a model-based analysis. *The Lancet Infectious Diseases*, 21(2):203–212. https://doi.org/10.1016/S1473-3099(20)30769-6

46. Yang W, Shaman J (2021) Development of a model-inference system for estimating epidemiological characteristics of SARS-CoV-2 variants of concern. *Nature Communications*, 12(1):5573. https://doi.org/10.1038/s41467-021-25913-9

47. Yang W, Greene SK, Peterson ER, Li W, Mathes R, Graf L, Lall R, Hughes S, Wang J, Fine A Epidemiological characteristics of the B.1.526 SARS-CoV-2 variant. *Science Advances*, 8(4):eabm0300. https://doi.org/10.1126/sciadv.abm0300

48. Yang W Transmission dynamics of and insights from the 2018–2019 measles outbreak in New York City: A modeling study. *Science Advances*, 6(22):eaaz4037. https://doi.org/10.1126/sciadv.aaz4037

49. Yang W, Wen L, Li S-L, Chen K, Zhang W-Y, Shaman J (2017) Geospatial characteristics of measles transmission in China during 2005−2014. *PLOS Computational Biology*, 13(4):e1005474. https://doi.org/10.1371/journal.pcbi.1005474

50. Yang W, Li J, Shaman J (2019) Characteristics of measles epidemics in China (1951-2004) and implications for elimination: A case study of three key locations. *PLOS Computational Biology*, 15(2):e1006806. https://doi.org/10.1371/journal.pcbi.1006806

51. Bomfim R, Pei S, Shaman J, Yamana T, Makse HA, Andrade JS, Lima Neto AS, Furtado V (2020) Predicting dengue outbreaks at neighbourhood level using human mobility in urban areas. *Journal of The Royal Society Interface*, 17(171):20200691. https://doi.org/10.1098/rsif.2020.0691

52. Yamana TK, Kandula S, Shaman J (2016) Superensemble forecasts of dengue outbreaks. *Journal of The Royal Society Interface*, 13(123):20160410. https://doi.org/10.1098/rsif.2016.0410

53. Shaman J, Yang W, Kandula S (2014) Inference and Forecast of the Current West African Ebola Outbreak in Guinea, Sierra Leone and Liberia. *PLOS Currents Outbreaks*, https://doi.org/10.1371/currents.outbreaks.3408774290b1a0f2dd7cae877c8b8ff6

54. Yang W, Zhang W, Kargbo D, Yang R, Chen Y, Chen Z, Kamara A, Kargbo B, Kandula S, Karspeck A, Liu C, Shaman J (2015) Transmission network of the 2014–2015 Ebola epidemic in Sierra Leone. *Journal of The Royal Society Interface*, 12(112):20150536. https://doi.org/10.1098/rsif.2015.0536

55. Pei S, Morone F, Liljeros F, Makse H, Shaman JL (2018) Inference and control of the nosocomial transmission of methicillin-resistant Staphylococcus aureus. *eLife*, 7:e40977. https://doi.org/10.7554/eLife.40977

56. Pei S, Liljeros F, Shaman J (2021) Identifying asymptomatic spreaders of antimicrobial-resistant pathogens in hospital settings. *Proceedings of the National Academy of Sciences*, 118(37):e2111190118. https://doi.org/10.1073/pnas.2111190118

57. Ruggeri K The globalizability of temporal discounting. *Nature Human Behaviour*, Forthcoming:42.

58. Ruggeri K (2021) Psychology and Behavioral Economics: Applications for Public Policy.

59. Ruggeri K (2018) Behavioral Insights for Public Policy: Concepts and Cases.

60. Ruggeri K, Folke T (2021) Unstandard Deviation: The Untapped Value of Positive Deviance for Reducing Inequalities. *Perspectives on Psychological Science*, :17456916211017864. https://doi.org/10.1177/17456916211017865

61. Ruggeri K, Benzerga A, Verra S, Folke T (2020) A behavioral approach to personalizing public health. *Behavioural Public Policy*, :1–13. https://doi.org/10.1017/bpp.2020.31

62. Ruggeri K, Folke T, Benzerga A, Verra S, Büttner C, Steinbeck V, Yee S, Chaiyachati K (2020) Nudging New York: adaptive models and the limits of behavioral interventions to reduce no-shows and health inequalities. *BMC Health Services Research*, 20(1):363. https://doi.org/10.1186/s12913-020-05097-6

63. Ruggeri K, Alí S, Berge ML, Bertoldo G, Bjørndal LD, Cortijos-Bernabeu A, Davison C, Demić E, Esteban-Serna C, Friedemann M, Gibson SP, Jarke H, Karakasheva R, Khorrami PR, Kveder J, Andersen TL, Lofthus IS, McGill L, Nieto AE, Pérez J, Quail SK, Rutherford C, Tavera FL, Tomat N, Reyn CV, Većkalov B, Wang K, Yosifova A, Papa F, Rubaltelli E, Linden S van der, Folke T (2020) Replicating patterns of prospect theory for decision under risk. *Nature Human Behaviour*, 4(6):622–633. https://doi.org/10.1038/s41562-020-0886-x

64. Halilova JG, Fynes-Clinton S, Green L, Myerson J, Wu J, Ruggeri K, Addis DR, Rosenbaum RS (2022) Short-sighted decision-making by those not vaccinated against COVID-19. https://doi.org/10.31234/osf.io/6uqky

65. Schumpe BM, Van Lissa CJ, Bélanger JJ, Ruggeri K, Mierau J, Nisa CF, Molinario E, Gelfand MJ, Stroebe W, Agostini M, Gützkow B, Jeronimus BF, Kreienkamp J, Kutlaca M, Lemay EP, Reitsema AM, vanDellen MR, Abakoumkin G, Abdul Khaiyom JH, Ahmedi V, Akkas H, Almenara CA, Atta M, Bagci SC, Basel S, Berisha Kida E, Bernardo ABI, Buttrick NR, Chobthamkit P, Choi H-S, Cristea M, Csaba S, Damnjanović K, Danyliuk I, Dash A, Di Santo D, Douglas KM, Enea V, Faller D, Fitzsimons GJ, Gheorghiu A, Gómez Á, Hamaidia A, Han Q, Helmy M, Hudiyana J, Jiang D-Y, Jovanović V, Kamenov Z, Kende A, Keng S-L, Kieu TTT, Koc Y, Kovyazina K, Kozytska I, Krause J, Kruglanski AW, Kurapov A, Lantos NA, Lesmana CBJ, Louis WR, Lueders A, Malik NI, Martinez AP, McCabe KO, Mehulić J, Milla MN, Mohammed I, Moyano M, Muhammad H, Mula S, Muluk H, Myroniuk S, Najafi R, Nyúl B, O’Keefe PA, Olivas Osuna JJ, Osin EN, Park J, Pica G, Pierro A, Rees JH, Resta E, Rullo M, Ryan MK, Samekin A, Santtila P, Sasin E, Selim HA, Stanton MV, Sultana S, Sutton RM, Tseliou E, Utsugi A, Breen JA van, Van Veen K, Vázquez A, Wollast R, Yeung VW-L, Zand S, Žeželj IL, Zheng B, Zick A, Zúñiga C, Leander NP (2022) Predictors of adherence to public health behaviors for fighting COVID-19 derived from longitudinal data. *Scientific Reports*, 12(1):3824. https://doi.org/10.1038/s41598-021-04703-9

66. SafeGraph POI Data, Business Listings, & Foot-Traffic Data. https://www.safegraph.com/

67. Delphi Group Delphi’s COVID-19 Trends and Impact Surveys (CTIS). https://delphi.cmu.edu/covid19/ctis/

68. Salomon JA, Reinhart A, Bilinski A, Chua EJ, La Motte-Kerr W, Rönn MM, Reitsma MB, Morris KA, LaRocca S, Farag TH, Kreuter F, Rosenfeld R, Tibshirani RJ (2021) The US COVID-19 Trends and Impact Survey: Continuous real-time measurement of COVID-19 symptoms, risks, protective behaviors, testing, and vaccination. *Proceedings of the National Academy of Sciences*, 118(51):e2111454118. https://doi.org/10.1073/pnas.2111454118

69. Qi Y, Shaman J, Pei S (2021) Quantifying the Impact of COVID-19 Nonpharmaceutical Interventions on Influenza Transmission in the United States. *The Journal of Infectious Diseases*, 224(9):1500–1508. https://doi.org/10.1093/infdis/jiab485

70. Pei S, Shaman J (2020) Initial Simulation of SARS-CoV2 Spread and Intervention Effects in the Continental US. *medRxiv*, :2020.03.21.20040303. https://doi.org/10.1101/2020.03.21.20040303

71. Pei S, Dahl KA, Yamana TK, Licker R, Shaman J (2020) Compound Risks of Hurricane Evacuation Amid the COVID-19 Pandemic in the United States. *GeoHealth*, 4(12):e2020GH000319. https://doi.org/10.1029/2020GH000319

72. Ma Y, Pei S, Shaman J, Dubrow R, Chen K (2021) Role of meteorological factors in the transmission of SARS-CoV-2 in the United States. *Nature Communications*, 12(1):3602. https://doi.org/10.1038/s41467-021-23866-7

73. Galanti M, Pei S, Yamana TK, Angulo FJ, Charos A, Swerdlow DL, Shaman J (2021) Social distancing remains key during vaccinations. *Science*, 371(6528):473–474. https://doi.org/10.1126/science.abg2326

74. Zebrowski A, Rundle A, Pei S, Yaman T, Yang W, Carr BG, Sims S, Doorley R, Schluger N, Quinn JW, Shaman J, Branas CC (2021) A Spatiotemporal Tool to Project Hospital Critical Care Capacity and Mortality From COVID-19 in US Counties. *American Journal of Public Health*, 111(6):1113–1122. https://doi.org/10.2105/AJPH.2021.306220

75. Xu X-K, Liu XF, Wang L, Wu Y, Lu X, Wang X, Pei S (2022) Assessing the spread risk of COVID-19 associated with multi-mode transportation networks in China. *Fundamental Research*, https://doi.org/10.1016/j.fmre.2022.04.006

76. Galanti M, Pei S, Yamana TK, Angulo FJ, Charos A, Khan F, Shea KM, Swerdlow DL, Shaman J (2021) Non-pharmaceutical interventions and inoculation rate shape SARS-CoV-2 vaccination campaign success. *Epidemiology & Infection*, 149https://doi.org/10.1017/S095026882100217X

77. Xu X-K, Wang L, Pei S (2021) Multiscale mobility explains differential associations between the gross domestic product and COVID-19 transmission in Chinese cities. *Journal of Travel Medicine*, 28(2):taaa236. https://doi.org/10.1093/jtm/taaa236

78. Pei S, Teng X, Lewis P, Shaman J (2021) Optimizing respiratory virus surveillance networks using uncertainty propagation. *Nature Communications*, 12(1):222. https://doi.org/10.1038/s41467-020-20399-3

79. Pei S, Cane MA, Shaman J (2019) Predictability in process-based ensemble forecast of influenza. *PLOS Computational Biology*, 15(2):e1006783. https://doi.org/10.1371/journal.pcbi.1006783

80. Pei S, Shaman J (2020) Aggregating forecasts of multiple respiratory pathogens supports more accurate forecasting of influenza-like illness. *PLOS Computational Biology*, 16(10):e1008301. https://doi.org/10.1371/journal.pcbi.1008301

81. Kramer SC, Pei S, Shaman J (2020) Forecasting influenza in Europe using a metapopulation model incorporating cross-border commuting and air travel. *PLOS Computational Biology*, 16(10):e1008233. https://doi.org/10.1371/journal.pcbi.1008233

82. Glanz J, Leatherby L, Bloch M, Smith M, Buchanan L, Wu J, Bogel-Burroughs N (2020) Coronavirus Could Overwhelm U.S. Without Urgent Action, Estimates Say. *The New York Times*, https://www.nytimes.com/interactive/2020/03/20/us/coronavirus-model-us-outbreak.html

83. Shaman J, Kohn M (2009) Absolute humidity modulates influenza survival, transmission, and seasonality. *Proceedings of the National Academy of Sciences*, 106(9):3243–3248. https://doi.org/10.1073/pnas.0806852106

84. Shaman J, Pitzer VE, Viboud C, Grenfell BT, Lipsitch M (2010) Absolute Humidity and the Seasonal Onset of Influenza in the Continental United States. *PLOS Biology*, 8(2):e1000316. https://doi.org/10.1371/journal.pbio.1000316

85. Shaman J, Day JF (2007) Reproductive Phase Locking of Mosquito Populations in Response to Rainfall Frequency. *PLOS ONE*, 2(3):e331. https://doi.org/10.1371/journal.pone.0000331

86. Shaman J, Spiegelman M, Cane M, Stieglitz M (2006) A hydrologically driven model of swamp water mosquito population dynamics. *Ecological Modelling*, 194(4):395–404. https://doi.org/10.1016/j.ecolmodel.2005.10.037

87. Shaman J, Tziperman E (2016) The Superposition of Eastward and Westward Rossby Waves in Response to Localized Forcing. *Journal of Climate*, 29(20):7547–7557. https://doi.org/10.1175/JCLI-D-16-0119.1

88. Shaman J, Samelson RM, Tziperman E (2012) Complex Wavenumber Rossby Wave Ray Tracing. *Journal of the Atmospheric Sciences*, 69(7):2112–2133. https://doi.org/10.1175/JAS-D-11-0193.1

89. Shaman J (2007) Amplification due to spatial clustering in an individual-based model of mosquito–avian arbovirus transmission. *Transactions of The Royal Society of Tropical Medicine and Hygiene*, 101(5):469–483. https://doi.org/10.1016/j.trstmh.2006.11.007

90. Shaman J (2011) Strategies for Controlling the Epizootic Amplification of Arboviruses. *Journal of Medical Entomology*, 48(6):1189–1196. https://doi.org/10.1603/ME11085

91. Shaman J, Tziperman E (2007) Summertime ENSO–North African–Asian Jet teleconnection and implications for the Indian monsoons. *Geophysical Research Letters*, 34(11)https://doi.org/10.1029/2006GL029143

92. Rydbeck AV, Maloney ED, Xie S-P, Hafner J, Shaman J (2013) Remote Forcing versus Local Feedback of East Pacific Intraseasonal Variability during Boreal Summer. *Journal of Climate*, 26(11):3575–3596. https://doi.org/10.1175/JCLI-D-12-00499.1

93. Shaman J (2014) The Seasonal Effects of ENSO on Atmospheric Conditions Associated with European Precipitation: Model Simulations of Seasonal Teleconnections. *Journal of Climate*, 27(3):1010–1028. https://doi.org/10.1175/JCLI-D-12-00734.1

94. Yang W, Lipsitch M, Shaman J (2015) Inference of seasonal and pandemic influenza transmission dynamics. *Proceedings of the National Academy of Sciences*, 112(9):2723–2728. https://doi.org/10.1073/pnas.1415012112

95. Yang W, Elankumaran S, Marr LC (2011) Concentrations and size distributions of airborne influenza A viruses measured indoors at a health centre, a day-care centre and on aeroplanes. *Journal of The Royal Society Interface*, 8(61):1176–1184. https://doi.org/10.1098/rsif.2010.0686

96. Yang W, Marr LC (2011) Dynamics of Airborne Influenza A Viruses Indoors and Dependence on Humidity. *PLOS ONE*, 6(6):e21481. https://doi.org/10.1371/journal.pone.0021481

97. Yang W, Elankumaran S, Marr LC (2012) Relationship between Humidity and Influenza A Viability in Droplets and Implications for Influenza’s Seasonality. *PLOS ONE*, 7(10):e46789. https://doi.org/10.1371/journal.pone.0046789

98. Yang W, Marr LC (2012) Mechanisms by Which Ambient Humidity May Affect Viruses in Aerosols. *Applied and Environmental Microbiology*, 78(19):6781–6788. https://doi.org/10.1128/AEM.01658-12

99. Yuan H, Kramer SC, Lau EHY, Cowling BJ, Yang W (2021) Modeling influenza seasonality in the tropics and subtropics. *PLOS Computational Biology*, 17(6):e1009050. https://doi.org/10.1371/journal.pcbi.1009050

100. Yang W, Cummings MJ, Bakamutumaho B, Kayiwa J, Owor N, Namagambo B, Byaruhanga T, Lutwama JJ, O’Donnell MR, Shaman J (2018) Transmission dynamics of influenza in two major cities of Uganda. *Epidemics*, 24:43–48. https://doi.org/10.1016/j.epidem.2018.03.002

101. Reis J, Shaman J (2016) Retrospective Parameter Estimation and Forecast of Respiratory Syncytial Virus in the United States. *PLOS Computational Biology*, 12(10):e1005133. https://doi.org/10.1371/journal.pcbi.1005133

102. DeFelice NB, Little E, Campbell SR, Shaman J (2017) Ensemble forecast of human West Nile virus cases and mosquito infection rates. *Nature Communications*, 8(1):14592. https://doi.org/10.1038/ncomms14592

103. DeFelice NB, Schneider ZD, Little E, Barker C, Caillouet KA, Campbell SR, Damian D, Irwin P, Jones HMP, Townsend J, Shaman J (2018) Use of temperature to improve West Nile virus forecasts. *PLOS Computational Biology*, 14(3):e1006047. https://doi.org/10.1371/journal.pcbi.1006047

104. (2021) Mayor de Blasio Announces Columbia University Will Operate and Oversee Pandemic Response Institute. *The official website of the City of New York*, http://www1.nyc.gov/office-of-the-mayor/news/660-21/mayor-de-blasio-columbia-university-will-operate-oversee-pandemic-response-institute

105. Wang CC, Prather KA, Sznitman J, Jimenez JL, Lakdawala SS, Tufekci Z, Marr LC (2021) Airborne transmission of respiratory viruses. *Science*, 373(6558)https://doi.org/10.1126/science.abd9149

106. Harrison AG, Lin T, Wang P (2020) Mechanisms of SARS-CoV-2 Transmission and Pathogenesis. *Trends in Immunology*, 41(12):1100–1115. https://doi.org/10.1016/j.it.2020.10.004

107. Chang S, Pierson E, Koh PW, Gerardin J, Redbird B, Grusky D, Leskovec J (2021) Mobility network models of COVID-19 explain inequities and inform reopening. *Nature*, 589(7840):82–87. https://doi.org/10.1038/s41586-020-2923-3

108. Hou X, Gao S, Li Q, Kang Y, Chen N, Chen K, Rao J, Ellenberg JS, Patz JA (2021) Intracounty modeling of COVID-19 infection with human mobility: Assessing spatial heterogeneity with business traffic, age, and race. *Proceedings of the National Academy of Sciences*, 118(24)https://doi.org/10.1073/pnas.2020524118

109. SafeGraph | Academics. https://www.safegraph.com/academics

110. Weekly Patterns | SafeGraph Docs. *SafeGraph*, https://docs.safegraph.com/docs/weekly-patterns

111. Geometry | SafeGraph Docs. *SafeGraph*, https://docs.safegraph.com/docs/weekly-patterns

112. Core Places | SafeGraph Docs. *SafeGraph*, https://docs.safegraph.com/docs/weekly-patterns

113. COVID-19 Community Mobility Report. *COVID-19 Community Mobility Report*, https://www.google.com/covid19/mobility?hl=en

114. Bureau UC American Community Survey (ACS). *Census.gov*, https://www.census.gov/programs-surveys/acs

115. Data C of NY NYC Open NYC Open Data. http://nycod-wpengine.com/

116. EpiQuery | Search for data, surveys and records on the health of New Yorkers. https://a816-health.nyc.gov/hdi/epiquery/visualizations

117. Kahneman D, Tversky A (2012) Prospect Theory: An Analysis of Decision Under Risk. *Handbook of the Fundamentals of Financial Decision Making*, Volume 4:99–127. https://doi.org/10.1142/9789814417358\_0006

118. Emirbayer M, Mische A (2015) What Is Agency? *American Journal of Sociology*, https://doi.org/10.1086/231294

119. Legros S, Cislaghi B (2020) Mapping the Social-Norms Literature: An Overview of Reviews. *Perspectives on Psychological Science*, 15(1):62–80. https://doi.org/10.1177/1745691619866455

120. Falk A, Hermle J (2018) Relationship of gender differences in preferences to economic development and gender equality. *Science*, 362(6412):eaas9899. https://doi.org/10.1126/science.aas9899

121. Szimba E, Hartmann M (2020) Assessing travel time savings and user benefits of automated driving – A case study for a commuting relation. *Transport Policy*, 98:229–237. https://doi.org/10.1016/j.tranpol.2020.03.007

122. Camerer C (2005) Three Cheers—Psychological, Theoretical, Empirical—for Loss Aversion. *Journal of Marketing Research*, 42(2):129–133. https://doi.org/10.1509/jmkr.42.2.129.62286

123. Barlas Z, Obhi S (2013) Freedom, choice, and the sense of agency. *Frontiers in Human Neuroscience*, 7https://www.frontiersin.org/article/10.3389/fnhum.2013.00514

124. Loo BPY, Chen C, Chan ETH (2010) Rail-based transit-oriented development: Lessons from New York City and Hong Kong. *Landscape and Urban Planning*, 97(3):202–212. https://doi.org/10.1016/j.landurbplan.2010.06.002

125. Sanchez TW (2008) Poverty, policy, and public transportation. *Transportation Research Part A: Policy and Practice*, 42(5):833–841. https://doi.org/10.1016/j.tra.2008.01.011

126. Thaler RH, Sunstein CR (2008) Nudge: Improving decisions about health, wealth, and happiness. :x, 293.

127. DEEP Software - CDS Public Wiki. https://sites.google.com/a/decisionsciences.columbia.edu/cds-wiki/deep-software

128. Rieger MO, Wang M, Hens T (2011) Prospect Theory Around the World. *SSRN Electronic Journal*, https://doi.org/10.2139/ssrn.1957606

129. Ho TK (1995) Random decision forests. *Proceedings of 3rd International Conference on Document Analysis and Recognition*, 1:278–282 vol.1. https://doi.org/10.1109/ICDAR.1995.598994

130. Segal MR (2004) Machine Learning Benchmarks and Random Forest Regression. https://escholarship.org/uc/item/35x3v9t4

131. Anderson JL (2001) An Ensemble Adjustment Kalman Filter for Data Assimilation. *Monthly Weather Review*, 129(12):2884–2903. https://doi.org/10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2

132. Arulampalam MS, Maskell S, Gordon N, Clapp T (2002) A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, 50(2):174–188. https://doi.org/10.1109/78.978374

133. Ionides EL, Bretó C, King AA (2006) Inference for nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 103(49):18438–18443. https://doi.org/10.1073/pnas.0603181103

134. Ionides EL, Bhadra A, Atchadé Y, King A (2011) Iterated filtering. *Annals of Statistics*, 39(3):1776–1802. https://doi.org/10.1214/11-AOS886

135. Snyder C, Bengtsson T, Bickel P, Anderson J (2008) Obstacles to High-Dimensional Particle Filtering. *Monthly Weather Review*, 136(12):4629–4640. https://doi.org/10.1175/2008MWR2529.1

136. Evensen G (2009) Data assimilation: the ensemble Kalman filter.

137. Katzfuss M, Stroud JR, Wikle CK (2016) Understanding the Ensemble Kalman Filter. *The American Statistician*, 70(4):350–357. https://doi.org/10.1080/00031305.2016.1141709

138. Csilléry K, Blum MGB, Gaggiotti OE, François O (2010) Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology & Evolution*, 25(7):410–418. https://doi.org/10.1016/j.tree.2010.04.001

139. Sunnåker M, Busetto AG, Numminen E, Corander J, Foll M, Dessimoz C (2013) Approximate Bayesian Computation. *PLOS Computational Biology*, 9(1):e1002803. https://doi.org/10.1371/journal.pcbi.1002803

140. Brooks S, Gelman A, Jones G, Meng X-L (2011) Handbook of Markov Chain Monte Carlo.

141. Gilks WR, Richardson S, Spiegelhalter D (1995) Markov Chain Monte Carlo in Practice.

142. Mlcochova P, Kemp SA, Dhar MS, Papa G, Meng B, Ferreira IATM, Datir R, Collier DA, Albecka A, Singh S, Pandey R, Brown J, Zhou J, Goonawardane N, Mishra S, Whittaker C, Mellan T, Marwal R, Datta M, Sengupta S, Ponnusamy K, Radhakrishnan VS, Abdullahi A, Charles O, Chattopadhyay P, Devi P, Caputo D, Peacock T, Wattal C, Goel N, Satwik A, Vaishya R, Agarwal M, Mavousian A, Lee JH, Bassi J, Silacci-Fegni C, Saliba C, Pinto D, Irie T, Yoshida I, Hamilton WL, Sato K, Bhatt S, Flaxman S, James LC, Corti D, Piccoli L, Barclay WS, Rakshit P, Agrawal A, Gupta RK (2021) SARS-CoV-2 B.1.617.2 Delta variant replication and immune evasion. *Nature*, 599(7883):114–119. https://doi.org/10.1038/s41586-021-03944-y

143. Cao Y, Wang J, Jian F, Xiao T, Song W, Yisimayi A, Huang W, Li Q, Wang P, An R, Wang J, Wang Y, Niu X, Yang S, Liang H, Sun H, Li T, Yu Y, Cui Q, Liu S, Yang X, Du S, Zhang Z, Hao X, Shao F, Jin R, Wang X, Xiao J, Wang Y, Xie XS (2022) Omicron escapes the majority of existing SARS-CoV-2 neutralizing antibodies. *Nature*, 602(7898):657–663. https://doi.org/10.1038/s41586-021-04385-3

144. Ray EL, Wattanachit N, Niemi J, Kanji AH, House K, Cramer EY, Bracher J, Zheng A, Yamana TK, Xiong X, Woody S, Wang Y, Wang L, Walraven RL, Tomar V, Sherratt K, Sheldon D, Reiner RC, Prakash BA, Osthus D, Li ML, Lee EC, Koyluoglu U, Keskinocak P, Gu Y, Gu Q, George GE, España G, Corsetti S, Chhatwal J, Cavany S, Biegel H, Ben-Nun M, Walker J, Slayton R, Lopez V, Biggerstaff M, Johansson MA, Reich NG, Consortium on behalf of the C-19 FH (2020) Ensemble Forecasts of Coronavirus Disease 2019 (COVID-19) in the U.S. :2020.08.19.20177493. https://doi.org/10.1101/2020.08.19.20177493

145. Bracher J, Ray EL, Gneiting T, Reich NG (2021) Evaluating epidemic forecasts in an interval format. *PLOS Computational Biology*, 17(2):e1008618. https://doi.org/10.1371/journal.pcbi.1008618

146. Home - COVID 19 forecast hub. https://covid19forecasthub.org/

147. Bröcker J, Smith LA (2007) Increasing the Reliability of Reliability Diagrams. *Weather and Forecasting*, 22(3):651–661. https://doi.org/10.1175/WAF993.1

148. (2022) Columbia COVID-19 Projection in the US, https://github.com/shaman-lab/COVID-19Projection. https://github.com/shaman-lab/COVID-19Projection

149. Lamb MR, Kandula S, Shaman J (2021) Differential COVID-19 case positivity in New York City neighborhoods: Socioeconomic factors and mobility. *Influenza and Other Respiratory Viruses*, 15(2):209–217. https://doi.org/10.1111/irv.12816

150. Kandula S, Shaman J (2021) Investigating associations between COVID-19 mortality and population-level health and socioeconomic indicators in the United States: A modeling study. *PLOS Medicine*, 18(7):e1003693. https://doi.org/10.1371/journal.pmed.1003693

151. Kannoth S, Kandula S, Shaman J (2022) The association between early country-level COVID-19 testing capacity and later COVID-19 mortality outcomes. *Influenza and Other Respiratory Viruses*, 16(1):56–62. https://doi.org/10.1111/irv.12906

152. Cribbin WL, Galanti M, Shaman J (2021) Socioeconomic Disparities in SARS-CoV-2 Serology Testing in New York City. *ISEE Conference Abstracts*, https://doi.org/10.1289/isee.2021.P-679