Information Systems Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Smart Testing with Vaccination: A Bandit Algorithm for Active Sampling for Managing COVID-19

Yingfei Wang, Inbal Yahav, Balaji Padmanabhan

To cite this article:
Yingfei Wang, Inbal Yahav, Balaji Padmanabhan (2024) Smart Testing with Vaccination: A Bandit Algorithm for Active Sampling for Managing COVID-19. Information Systems Research 35(1):120-144. https://doi.org/10.1287/isre.2023.1215

https://pubsonline.informs.org/journal/isre

INFORMATION SYSTEMS RESEARCH
Vol. 35, No. 1, March 2024, pp. 120–144
ISSN 1047-7047 (print), ISSN 1526-5536 (online)

Smart Testing with Vaccination: A Bandit Algorithm for Active Sampling for Managing COVID-19

Yingfei Wang,[a,*] Inbal Yahav,[b] Balaji Padmanabhan[c]

[a] Foster School of Business, University of Washington, Seattle, Washington 98195; [b] Coller School of Management, Tel Aviv University, Tel Aviv 6997801, Israel; [c] Muma College of Business, University of South Florida, Tampa, Florida 33620

*Corresponding author

Contact: yingfei@uw.edu, https://orcid.org/0000-0002-3634-3617 (YW); inbalyahav@tauex.tau.ac.il, https://orcid.org/0000-0002-1513-017X (IY); bp@usf.edu, https://orcid.org/0000-0002-3498-0778 (BP)

Received: August 10, 2021
Revised: June 20, 2022; December 8, 2022
Accepted: February 1, 2023
Published Online in Articles in Advance: May 31, 2023

https://doi.org/10.1287/isre.2023.1215

Abstract. This paper presents methods to choose individuals to test for infection during a pandemic such as COVID-19, characterized by high contagion and presence of asymptomatic carriers. The smart-testing ideas presented here are motivated by active learning and multi-armed bandit techniques in machine learning. Our active sampling method works in conjunction with quarantine policies, can handle different objectives, and is dynamic and adaptive in the sense that it continually adapts to changes in real-time data. The bandit algorithm uses contact tracing, location-based sampling and random sampling in order to select specific individuals to test. Using a data-driven agent-based model simulating New York City we show that the algorithm samples individuals to test in a manner that rapidly traces infected individuals. Experiments also suggest that smart-testing can significantly reduce the death rates as compared with current methods, with or without vaccination. While smart testing strategies can help mitigate disease spread, there could be unintended consequences with fairness or bias when deployed in real-world settings. To this end we show how procedural fairness can be incorporated into our method and present results that show that this can be done without hurting the effectiveness of the mitigation that can be achieved.

History: Ahmed Abbasi, Senior Editor; Maytal Saar-Tsechansky, Associate Editor.

Funding: W. Yahav is supported by the Jeremy Coller Foundation and the Henry Crown Institute of Business Research in Israel.

Supplemental Material: The e-companion is available at https://doi.org/10.1287/isre.2023.1215.

Keywords: active sampling • multi-armed bandits • agent-based models • COVID-19 • non-stationarity • algorithmic fairness

1. Introduction

This paper presents a method to actively sample individuals in a population as a way to mitigate the spread of pandemics such as COVID-19. Sampling algorithms are commonly used in machine learning to acquire training data labels for classification ("active learning"; Cohn et al. 1996, Zheng and Padmanabhan 2006) and in bandit algorithms (Macready and Wolpert 1998, Auer et al. 2002) to explore complex search spaces through exploration and exploitation. Doing so in the context of improving decisions generally has been stressed in the information science literature before (Saar-Tsechansky and Provost 2007). The method we present in this paper builds on these ideas, but does so in the context of containing the spread of an epidemic in a population, and thus contributes to the nascent literature on the role of information systems in disaster management as well (Abbasi et al. 2021).

The literature on managing disease spread through ideas based on active sampling is primarily from the public health area. There are two reasons why this literature has considered sampling, although both are sometimes intertwined in the context of population surveillance (Lee et al. 2010). The first is to estimate the actual incidence or spread of a disease, such as HIV, in a population. In the case of estimating HIV incidence, methods such as a population survey and "sentinel surveillance" (Fylkesnes et al. 1998, Magnani et al. 2005) have been shown to be useful. The population survey in a catchment area (i.e., where there is likely disease) is (stratified/cluster) random sampling; it is generally considered the gold standard, but it is also known to be expensive (Lee et al. 2010). Sentinel

surveillance, in contrast, gathers data from a subset of venues (the "sentinels"; e.g., clinics) where the subset with the disease is likely to visit for treatment. Recognizing that sensitive diseases like HIV may be prevalent in hard-to-reach populations that might stay under the radar because of activities that are illegal or illicit (e.g. sex workers, drug users), the literature (Magnani et al. 2005, Goel and Salganik 2010) has critically studied other techniques such as snowball sampling (starting with a few seeds who have the disease and asking them to name others who might), respondent-driven sampling (similar to snowball, but longer chains with fewer referrals per seed), and time-location sampling (e.g., sampling users at nights in districts with brothels). With sensor data, newer forms of sampling are emerging as well. It has been shown, for instance, that to quantify disease spread in livestock that sampling the nodes (physical locations) where there is greatest movement can be effective (Dawson et al. 2015).

The second reason for considering sampling strategies is to mitigate the spread of disease by identifying individuals who need to be quarantined or tested. One method to do this, common in the public health literature is contact tracing (Huerta and Tsimring 2002, Eames and Keeling 2003, Underwood et al. 2003). The idea is to sample contacts of known infected individuals for presence of disease, and, along with quarantine, is the main mechanism used worldwide during the COVID-19 pandemic (Hellewell et al. 2020, Salathé et al. 2020). Traditionally, contact tracing is done through questionnaires, but newer automated contact tracing based on tracking cell phone proximities have also been proposed (Ferretti et al. 2020).

The SARS-Cov-2 pathogen and its transmission have the following unique characteristics, though, that make none of these current sampling methodologies sufficient by themselves.

- The possibility of presymptomatic and asymptomatic spread rule out pure sentinel-based strategies that would test those who report at clinics with symptoms. For the same reason, contact tracing alone would be insufficient. This is because contact tracing pursues only contacts of those known to have tested positive. Asymptomatic carriers may have also been transmitting the disease to their contacts, but those will not be actively sought by this strategy. Hence, random sampling is also useful.
- The pathogen has also been shown to survive on surfaces or in air in specific locations for extended periods of time. Hence, location-based sampling may also be useful.
- There is uncertainty about whether individuals develop long-lasting immunity. Hence sampling strategies may have to allow for retesting the same individuals repeatedly.
- Information about how covariates influence getting the disease (and its variants) is still evolving. For instance, while we know that comorbidities and age affect outcomes, there is still uncertainty about whether this knowledge is complete. Moreover, how covariates might affect contracting the disease is still uncertain. In this environment, covariate-based sampling strategies alone might be insufficient. Dynamic sampling strategies that target specific *individuals* to test is therefore critical.
- The highly infectious nature of the pathogen and the extent of worldwide spread have placed constraints on testing capacities, requiring prioritizing tests if they are to be useful in mitigating the spread.

This paper addresses all five aspects noted above, by presenting dynamic active sampling strategies based on multi-armed bandit algorithms that can generate real-time lists of whom to sample. Bandit algorithms are particularly good at balancing exploration with exploitation and have seen success in many scenarios where this tradeoff is important. This is the case with SARS-Cov-2, where exploration (random sampling) has to be effectively combined with exploitation (contact tracing, location-based sampling) in a dynamic manner based on data. One of the key aspects of our work is that we identify specific individuals to test/sample, as opposed to criteria or covariate-based sampling strategies.

Furthermore, our algorithm operates in a high uncertainty network environment. Specifically, the contacts between individuals are initially only partially observable and are revealed gradually based on individual sampling. Our proposed bandit algorithm, therefore, first uses the Thompson sampling strategy to trade off between expansion in unobserved nodes to identify possible new hot spots, versus densification on the observed portion to utilize the knowledge learned for testing. If densification is chosen, then an inner-level upper confidence bounding (UCB) policy is designed to balance between the individuals having higher probability of infection versus those having greater information uncertainty. In order to model how the sampling affects, and in turn is affected by, the underlying environment, we implement an agent-based model constructed from realistic data from New York City.

Several recent COVID-19 studies have focused on the ability to predict and flatten the spread of the disease by applying different intervention policies (Atkeson 2020, Chang et al. 2020, Topirceanu et al. 2020, Baxter et al. 2022). Among the policies that were examined were case isolation, home quarantine, social distancing, restrictions on air travels, and school closures and reopening. To assess the efficiency of the restrictions, most of these studies coupled S/I/R (susceptible/infected/recovered) diffusion modeling (Hethcote 2000, Cooper et al. 2020), with agent-based (simulation) models (ABM) (Perez and Dragicevic 2009) of infectious spread under different restrictions scenarios. Among these studies, an ABM of the COVID-19 spread in the city of New York under different quarantine policies was offered by Hoertel et al.

(2020). This model serves as a basis for the ABM that we developed in this work.

There is a similar question of whom to vaccinate (Chen et al. 2020). Also in contemporaneous work, Grushka-Cohen et al. (2020) presented a framework that uses historical data to build a classifier that computes a "risk score," which is the basis of determining how to combine exploration and exploitation. Our framework is general and allows for various vaccination policies to be incorporated into the background dynamics in order to test the effectiveness of combinational strategies such as smart testing with vaccination.

We make the following contributions. First, we present a novel active sampling algorithm to effectively manage pandemics such as COVID-19. We develop a general multi-armed bandit framework for this problem that (a) can leverage information of different types such as individuals and locations, (b) can handle uncertainty in both the underlying disease dynamic as well as information, and (c) is built to proactively sample based on individuals and not covariates. Second, our work adds to the literature on multi-armed bandits. Beyond classic bandit problems, most of the existing work regarding exploration/exploitation trade-offs under graph structures assumes that the complete network structure and underlying process is known beforehand. We model the partially observed scenarios and show how the network can be strategically expanded over time to support the active sampling strategy. Third, we show that smart testing is still important, even when vaccination is available, assuming that some portion of the individuals cannot or will not get the vaccine in a timely manner. Finally, given our understanding today that algorithms could have unintended consequences such as potential bias or fairness issues, we present an analysis of the fairness impact of our proposed method and show that, in our context, it does exhibit fairness based on outcomes. However, real-world contexts are likely different than the specific ABM that we study here in terms of how attributes such as race or gender influence the underlying dynamics. Therefore, we also present two variants of our method that can be implemented to have procedural fairness.

Before presenting the details, it is worth emphasizing that smart testing strategies are needed, even if tests are cheap and/or vaccines are available for the following reasons.

- Tests are never truly free; there is an infrastructure behind the testing (generating the tests, distributing, obtaining and storing results) that is expensive.
- In active infections that have high economic impact, policy makers may need to provide incentives to make individuals take these tests, particularly if they are asymptomatic.
- Pandemics such as COVID-19 are likely going to recur in the future, with no guarantees that tests for future infections will be cheap (particularly in the early stages, when active testing can play a critical role). The methodology presented in this paper applies to mitigating the spread of any pandemic.
- We view vaccines as complementary to testing, and, as we show, active testing strategies remain important, even with vaccinations as a combinational strategy to manage pandemics.

2. Problem Formulation

Taking a policy maker's perspective, our setting has three components. The *data setting* represents knowledge about individuals and their spatiotemporal behavior. The *process setting* represents knowledge about the underlying disease spread. The *policy makers' setting* represents what policy makers are assumed to have access to, and, accordingly, what the sampling algorithm will use. These are described below.

Data setting:

- a population of individuals P,
- a set of locations L in the city,
- spatiotemporal data about individuals in P over L.

Process setting:

- initial actual disease state (e.g., S/I/R) of each individual in $p \in P$,
- a spatiotemporal model of contagion that represents how the disease spreads in the population.

The policy makers' setting involves two components, the information setting and the policy setting. The information setting includes:

- complete list of individuals P,
- complete list of locations L,
- initial disease state of each individual in P (this can be incomplete),
- the contagion process is partially known—the policy maker (algorithm) is assumed to be aware of only the high-level factors affecting disease diffusion, such as contacts with exposed individuals or the role of locations where infected individuals have been.

The policy setting includes:

- a quarantine policy that determines when and how individuals will isolate in the population,
- a vaccination policy that determines which and when individuals will receive vaccine,
- an objective function over time (this paper focuses on sampling strategies for bridging the gap between the actual infected population and the known infected individuals, as a step toward controlling disease spread),
- constraints on number of tests per day (D_{max}) that a policy maker can proactively use to sample individuals.
- Sampling an individual to test reveals the following:
  - — disease state (infected or not) of the individual being tested.
  - — spatiotemporal data about the individual, if infected (in some cases, e.g., access to full GPS data, this will be complete information on the places visited and the full contact network; if this is by self-reporting,

as is common in most places, then the contact network will be incomplete and can potentially be gradually learned, in accordance with the current contact tracing program with interviewers to determine their close contacts.

Given this, the policy maker seeks to proactively sample individuals for testing in order to optimize the objective. The question that our research addresses can be posed, more generally, as follows: *Given a partially observed contact network with no information about how it was observed, and a budget to query the partially observed network nodes to either expand the network and/or reveal the nodes status, can we learn to sequentially ask optimal queries?*

3. Active Sampling Framework and Algorithm

The fundamental challenge policy makers face in mitigating COVID-19 is how to make wise decisions under uncertainty, with only partially known dynamics of contagion process, partially observed contacts between individuals, and disease states of individuals only if tested (among others mentioned in Section 2). This leads to natural policy considerations: Should they go for a decision that seems to be optimal for now based on known information, or acknowledge that their knowledge could be inaccurate and gather new information that could lead to improvements in the long run? This trade-off will then be intensified by limited available resources and considerations such as fairness and the cost of the intervention. In what follows, we present dynamic active sampling strategies based on multi-armed bandits that balance exploitation and exploration through sequentially and dynamically generating real-time lists of whom to test.

3.1. Background

The multi-armed bandit (MAB) is a generic framework to address the problem of decision making under uncertainty. In this setting, the learner must choose among a variety of actions and only observes partial feedback from the environment, without prior knowledge of which action is the best. In the classic stochastic K-armed bandit problem, at each time step t, the learner selects a single action/arm x_t among a set of K actions and observes some payoff $r_t(x_t)$. The reward of each arm is assumed to be drawn stochastically from some unknown probability distribution. The goal of the learner is to maximize the cumulative payoff obtained in a sequence of n allocations over time or, equivalently, minimize the *regret* (Bubeck and Cesa-Bianchi 2012), which is defined as the difference between the cumulative reward obtained by always playing the optimal arm in hindsight and the cumulative reward achieved by the learning policy,

$$\mathcal{R}_n = \max_{i=1,\dots,K} \mathbb{E}\left[\sum_{t=1}^{n} r_t(i) - \sum_{t=1}^{n} r_t(x_t)\right].$$

The fundamental exploration/exploitation dilemma is to: (1) gain as many rewards as possible in the current round, but also (2) have a high probability of correctly identifying the better arm. Thompson sampling (Thompson 1933, Chapelle and Li 2011) and upper confidence bounding policies (UCB) (Auer et al. 2002, Bubeck and Cesa-Bianchi 2012) have been recognized as the best performing policies both theoretically and empirically. To better leverage the widely available side information, there are studies on contextual bandits that reply on feature vectors (Li et al. 2010, Chu et al. 2011), and graphical bandits (e.g., active search on graph, bandits with graph-structured side observations).

The objective of active search on graph is to find as many target nodes as possible. Most of the existing work assumes that the complete network structure is known beforehand, or at least that a substantial part of the graph is revealed (Gu and Han 2014, Alon et al. 2015, Ma et al. 2015), with the exception of a few recent papers that consider partially observed networks (e.g., Singla et al. 2015, Soundarajan et al. 2017, Madhawa and Murata 2019). However, these approaches do not consider the node-sampling scenario needed for the testing problem considered in this paper. For example, if we can make an analogy to our setting, the feedback structure in Carpentier and Valko (2016) requires identification of all the individuals who caught the virus directly from the queried node, which is not applicable here. There is also related literature on bandits with graph-structured side observations. However, in their feedback loop, upon choosing an action, the decision maker gets to observe the rewards of all its "neighboring" arms (Mannor and Shamir 2011, Singh et al. 2020, Chen et al. 2021), which is significantly different from the settings we consider in this work. The general goal of other recent learning literature on graph exploration and discovery is to learn the topology of the underlying network using as fewer queries as possible (Dai et al. 2019, Kamarthi et al. 2020, Chiotellis and Cremers 2020), which is more relevant to robotics planning, navigation, or network coverage. In comparison, our objective is not to traverse or represent the underlying contact network but to mitigate the disease spread. Meanwhile, contact network will grow only if the node is tested positive.

As in what follows, we offer important modeling contributions in order to be able to cast the smart testing problem into a MAB framework, and we propose using network embedding with a two-level exploration/exploitation trade-off structure.

3.2. The Sampling Framework

We model contact networks as an undirected, time-dependent, network $\mathcal{G}_t =< V, E, D >$, where V is the node representation of the individuals P. The network is evolving over time with new contacts reported to the

policy maker. For example, if the policy maker aims to make daily decisions, then $\mathcal{G}_t$ represents the daily snapshot of the network on day t. An edge $e_{ij} \in E$ represents a direct contact between individuals v_i and v_j. At any time point, every node v in our graph $\mathcal{G}_t$ has one disease state in D. When node v is tested, a reward $Y_t(v) \in \{0,1\}$ (representing tested positive or not) is obtained, where $Y_t(v)$ is Bernoulli distributed with expectation $f_t(v)$ that depends on the true underlying topology $\mathcal{G}_t$ of the contact network (which is only partially known to the policy maker as $\hat{\mathcal{G}}_t$) and the diffusion process. When a node tests "positive," its *known* contacts will be revealed at once, with the understanding that there might be other contacts that are latent and not readily observed. We note that $Y_t(v)$ will be revealed only for nodes that are self-reported with coronavirus symptoms and/or actively selected to be tested by a testing policy.

Mathematically, this problem of active sampling can be formalized as a stochastic combinatorial optimization problem with $< V_t, \mathcal{U}, f_t >$, where V_t is the feasible set of nodes to be sampled from at time t, $\mathcal{U} = \{U \subset V_t : |U| \leq D_{max}\}$ is a family of subsets of V_t with up to D_{max} (daily testing capacity) nodes. The feasible choice set V_t excludes nodes that are known to be currently infected, previously dead, and/or immunized. The objective function for all the nodes in a set $U \subset V$ is defined by $F(U, f_t)$. For example, if the objective is to find as many infected nodes as possible in each day, $F(U, f_t)$ can be defined as

$$F(U, f_t) = \sum_{v \in U} f_t(v). \tag{1}$$

The goal of the sampling policy π is to adaptively select a sequence of subsets U_t^π to test, in a way that maximizes the expected cumulative rewards of $F(U_t^\pi, f_t)$ over time, recognizing that we observe, on testing, the realized reward $f_t(v)$ of each node $v \in U_t$ immediately, that is, whether the tested individual is infected or not.

An important observation is that immediate contacts alone (as done in previous work, e.g., Singla et al. 2015, Soundarajan et al. 2017, Madhawa and Murata 2019) may be insufficient to infer the risk of infection $f_t(v)$, and we need a richer representation of topology of the contact network. We seek to model *similarity* such that if one agent is infected, then a *similar* node is also possibly infected. This can help identify potential regions at high risk of outbreaks given previous outbreak locations, the mobility of agents, and their network structure. Local structure features, like degree, number of triads, and centrality, have been used in previous network analysis on finding structurally similar nodes (Henderson et al. 2012). Although these features can help infer roles of each node, such as super spreaders, or periphery nodes, they do not capture the information about the neighborhood similarity, social relations, and community membership. For example, because of a large amount of presymptomatic and asymptomatic disease transmissions, even if there is no direct contact link between two agents, if they constantly went to the same supermarket, or live in nearby neighborhoods, the agents can end up infecting each other. This complex interplay of social relationships and locations has resulted in several infections from the same gathering (superspreaders, hot spots) and brought up the importance of even seemingly minor occurrences such as how even a few agents with the same travel history may have caused unknown breakouts along the way.

Our goal is to quantify these aforementioned social relations by a continuous latent representation of nodes, which can then be exploited to guide the sampling policy to more effectively allocate testing kits of limited capacity. To this end, we formulate the active sampling problem as a contextual bandit, in which before making the choice of action, the learner observes the currently known disease states, as well as a feature vector x_i associated with each individual node v_i encoded from the contact network topology.

A pioneering method, DeepWalk (Perozzi et al. 2014), uses language modeling approaches to learn latent node embedding in the following two steps (please refer to Section 3.2.1 for the technical details of this approach). First, it traverses the network with random walks to infer local structures by neighborhood relations, and then uses a SkipGram model (Mikolov et al. 2013) to learn node embedding based on the produced samples. In our context, by translating the nodes in the network into a continuous space this way, the node embedding thus generated provides a method to efficiently sample individuals to test based on how "close" they are to others.

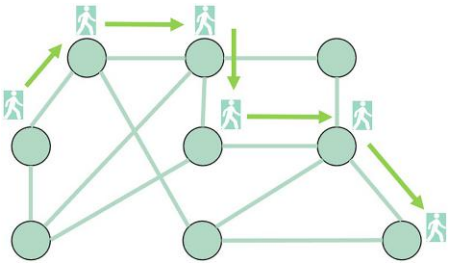
3.2.1. Technical Details of the DeepWalk Algorithm.

DeepWalk (Perozzi et al. 2014) uses language modeling approaches to learn latent node embedding in two steps: random walk and word embedding.

Random Walk on Graph. The first step of DeepWalk is to identify the context nodes for each node. Starting at any given node, it identifies all its neighbors, randomly selects one, and walks, with predefined path length, as illustrated in Figure 1. Repeat until there are enough samples. By generating truncated random walks from the network, the context nodes of $v \in V$ can be defined as the k neighboring nodes in each random walk sequence, which is a combination of nodes from v's 1-hop, 2-hop, … , k-hop neighbors. The idea is that the techniques used to model natural language (where the symbol frequency follows a power law distribution) can be repurposed for nodes appearing in short random walks whose frequencies also follow a power-law distribution.

SkipGram. The SkipGram algorithm is used to learn word embeddings thatwhich represent words in vector space so that if the word embeddings are close to one another, those words are semantically similar to one

Figure 1. (Color online) An Example Random Walk of Length 6 on a Network



other (Mikolov et al. 2013). The key idea of SkipGram is to quantify the similarity between any two words by how frequently they share the same surrounding words. SkipGram first identifies the neighboring words for a given target word within a predefined window size. Then it predicts the context word by maximizing the conditional probability of observing the context words given the target word.

3.3. The Sampling Algorithm

The sampling algorithm faces two major challenges: (1) the policy maker (through contact tracing from positive tests) observes only a portion of the network at any time t ($\hat{\mathcal{G}}_t$, as illustrated in Figure 2(a)), and (2) proactively sampling individuals, as opposed to group sampling, entails a combinatorial number of possible arms to sample with the cardinality constraint D_{max} (daily testing capacity).

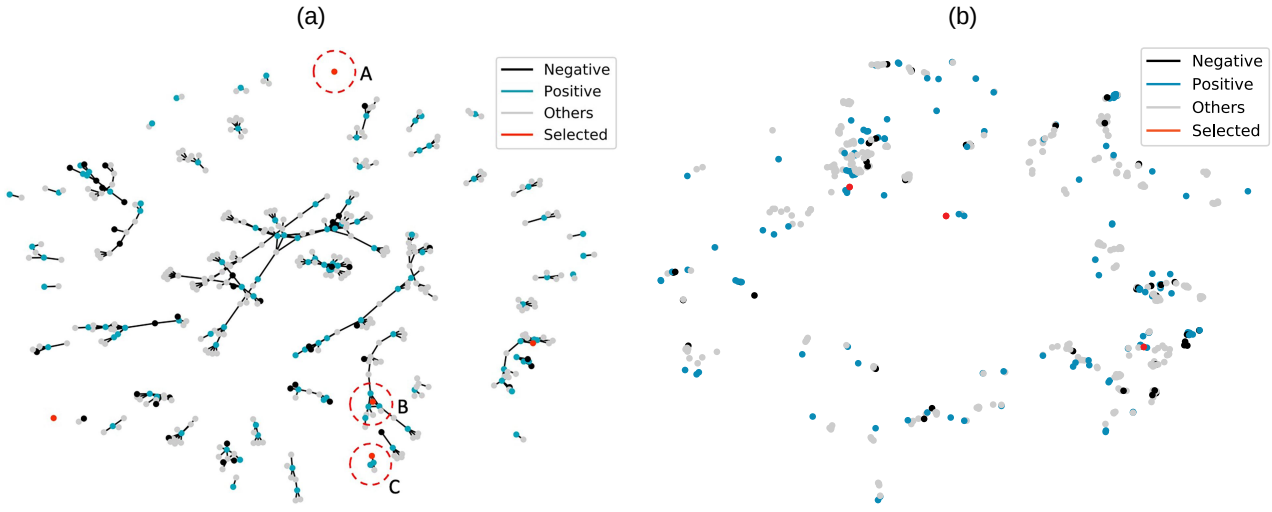
There are, in turn, three sources of uncertainty. First, whether the reported contacts are infected is not known, and needs to be learned as we continuously sample individuals and observe their test results $Y_t(v)$. Second, since the contact network is only partially observed, other possible paths of disease transmissions are not known to the policy maker. Third, the vast majority of the population is not reported in the current contact network $\hat{\mathcal{G}}_t$, and it is worthwhile to observe a broader area of the underlying network to detect and control possible outbreaks in unreported areas.

Correspondingly, the design of the proposed sampling policy concerns two levels of exploration/exploitation trade-off.

3.3.1. Outer Level: Expansion vs. Proliferation. In regard to the third source of uncertainty, the outer level is modeled as a two-armed bandit, balancing between the two choices of expansion outside the current known contact network, and/or proliferation within the current network (which is generated through contact tracing of known infected nodes). For example, in Figure 2(a), individual A is not observed in the current contact network, and is selected by the active sampling policy based on the choice of expansion.

We note here that, on each day, the sampling policy needs to proactively select D_{max} individuals all at once and the rewards are observed only at the end of the batch. Even though the regret analysis for batched stochastic multi-armed bandits still remains underexplored, it is empirically reported in various contexts that randomized policies turn out to be more robust to the impact of batching, and the improvement of a randomized policy over other methods becomes larger as batch size D_{max} increases (Kandasamy et al. 2018; Provodin et al. 2022). To this end, we choose to use Thompson sampling policy with

Figure 2. (Color online) An Illustrative Example of the Partially Observed Contact Network and the Two Levels of Exploration/Exploitation Trade-Off



Notes. Known testing results are reported, together with the individuals selected by the proposed sampling policy. (a) Partially observed contact network $\hat{\mathcal{G}}_t$ with some known testing results. (b) Partially observed contact network $\hat{\mathcal{G}}_t$ in the embedding space (t-SNE plot).

Beta-Bernoulli distribution (Thompson 1933, Chapelle and Li 2011). The expected rewards θ of the two choices are modeled with a Beta distribution. After one choice is chosen, the realized reward (whether an individual is tested positive) is sampled from a Bernoulli distribution. We repeatedly sample D_{max} realizations and choose between the two choices of expansion/proliferation that has the highest sampled value. If expansion is chosen, then our policy samples a node v that was not in $\hat{\mathcal{G}}_t$ uniformly at random. If proliferation is chosen, then our policy uses the inner-level exploration/exploitation algorithm to choose individuals v to sample. We use $\overline{D}_{max}$ to represent the total number of tests that are allocated to proliferation. After the test result $Y_t(v)$ for each chosen individual is revealed, the posterior Beta distribution will be updated, according to whether node v is from expansion or proliferation.

3.3.2. Inner Level: Individual Selection Within the Current Observed Network (Combinatorial *k*NN-UCB). Within the current observed contact network, the first two sources of uncertainty give rise to a trade-off between choosing individuals that are *seemingly* more prone to be infected, and/or learning the network to get more contact information to detect more infected in the future. As illustrated in Figure 2(a), given the current observed network, the probability of individual B being infected is very high since three of his or her direct contacts tested positive, together with other confirmed individuals nearby. In comparison, the neighboring information of individual C is very limited, rendering in higher uncertainty. It might be possible that individual C has a higher chance of being infected and we need to learn his or her surroundings. An inner-level bandit algorithm thus adaptively allocates $\overline{D}_{max}$ laboratory tests to individuals $U_t \subset V$ on the observed contact network with the goal of maximizing the cumulative rewards of $F(U_t, f_t)$ over time through a balance of exploitation versus exploration. Considering the inherent complexity of measuring individuals' heterogeneous risks of being infected given the network structure and the uncertainty around it, we adopt nonparametric modeling as follows.

On each day, we first use network embedding $x_v \in \mathcal{X}$ ($\mathcal{X} \subset \mathbb{R}^d$) from Section 3.2 to find the node representation. The node embedding encodes the network structure (i.e., neighbors and connection information) and quantifies the similarity between two nodes in the sense that if one is infected then the other is highly likely to be infected from potential contact chains. Figure 2(b) is an illustration of the observed nodes in the embedding space, which translates the network structure to continuous space for direct sampling from the entire population. Given that the representation space as a result of DeepWalk can change over time, we resort to nonparametric models. In our case, as each individual is represented by a lower-dimensional embedding ("lying in a manifold"), *k*NN regression can have provable high probability bounds (Jiang 2019) with mild assumptions such as Lipschitz mean reward, and compact support of the embedding space bounded below away from zero.

Definition 1 (*k*NN). Let the *k*NN radius of $x \in \mathcal{X} \subset \mathbb{R}^d$ be $\rho_k(x) := \inf\{\rho : |B(x,\rho) \cap \mathcal{X}| \geq k\}$, where $B(x,\rho)$ denotes the open metric ball of radius $\rho > 0$, centered at x. Let $\mathcal{N}_k(x)$ denote the k-neighborhood of x, such that $\mathcal{N}_k(x) := B(x, \rho_k(x)) \cap \mathcal{X}$. Then, for all $x \in \mathcal{X}$, the *k*NN regressor is defined with respect to the training data $(x_1, y_1), \ldots, (x_n, y_n)$,

$$\hat{f}(x) := \frac{1}{|\mathcal{N}_k(x)|} \sum_{i=1}^{n} y_i \mathbb{1}\{x_i \in \mathcal{N}_k(x)\}.$$

We then express explicitly the uncertainty about our knowledge with confidence intervals and follow the strong uniform consistency theorem for *k*NN regression to guide exploration.

Theorem 1 (Uniform Consistency (Jiang 2019)). *Let $\delta > 0$. There exists N_0 and universal constant C such that for $n \geq N_0$ and $k = \lfloor n^{2/(2+d)} \rfloor$, we have*

$$\mathbb{P}\left(\sup_{x \in \mathcal{X}} |f(x) - \hat{f}(x)| \geq C \cdot n^{-1/(2+d)} \sqrt{\log n \log\left(\frac{1}{\delta}\right)} \right) \leq \delta.$$

As illustrated in Figure 2, on each day t, some of the testing results (negative/positive) are known to the policy maker, either by self-reporting, or by the active sampling policy. It is also worth noting that the network embeddings for two unconnected nodes are not comparable. Thus, when constructing the *k*NN prediction for each node, we need to first find the reported (either positive or negative) individuals that are connected to the focal node. At the same time, based on the uniform consistency theorem, the size of the neighborhood (the value of k) needs to be node-dependent.

We now describe our combinatorial *k*NN-UCB as follows. On day t, for each node v_i in the current contact network, we use $N_t(i)$ to denote the number of known individuals (either tested positive or negative) that are connected to the focal node v_i. The neighborhood size of the focal node is chosen as $k_t(i) = \lfloor (N_t(i))^{2/(2+d)} \rfloor$. We use $\hat{f}_t(i)$ to denote the *k*NN estimate of the probability of infection of the focal node v_i. At each time step, if only one individual will be sampled, based on the principle of *optimism in the face of uncertainty* (Auer et al. 2002), the upper confidence bound (UCB) policies define the largest *plausible estimate* of each individual node as

$$\overline{f}_t(i) := \hat{f}_t(i) + \eta \cdot (N_t(i))^{-1/(2+d)} \sqrt{\log N_t(i) \log\left(\frac{1}{\delta}\right)}. \quad (2)$$

To ensure that the upper confidence bound decreases with respect to the size $N_t(i)$ of the neighborhood, we

modified the UCB for any $N_t(i) \le \lfloor e^{1+\frac{d}{2}} \rfloor + 1$ to be $(2e)^{-\frac{1}{2}}\sqrt{(2+d)\log\frac{1}{\delta}}$, based on Lemma EC.1 in the e-companion. Then the algorithm chooses the arm that maximizes the above quantity.

However, as mentioned above, in our specific COVID-19 active testing scenario, allocating $\overline{D}_{max}$ laboratory tests to individuals is a case of combinatorial optimization. To deal with this challenge, we allow the policy maker to use any exact/approximation/randomized algorithm, termed as ORACLE, to find solutions for the corresponding offline combinatorial optimization problem (if all the rewards $f(v)$ were known beforehand to the policy maker),

$$U^{\mathrm{OPT}} \in \arg\max_{U\in\mathcal{U}} F(U,f), |U| \le \overline{D}_{max}. \tag{3}$$

We denote the solution as $U^* = \mathrm{ORACLE}(V_t, \mathcal{U}, f, \overline{D}_{max})$. We here propose a combinatorial kNN-UCB algorithm that makes use of a greedy $\mathrm{ORACLE}(V_t, \mathcal{U}, \overline{f}_t, \overline{D}_{max})$ to provide $(1,e)$-approximate solutions for the offline optimization problem given that $F(U, f)$ is a submodular set function and monotone in U (Nemhauser et al. 1978). For example, if the objective is to find as many infected nodes as possible (as defined in Equation (1)), it achieves the case of equal sign in the submodularity definition. In general, submodular functions can be understood as natural diminishing returns properties; as such, the benefit of adding one additional tested individual decreases as the number of total tests increases.

The confidence level δ should be small enough to ensure optimism with high probability, but not so small that the larger upper confidence bounds will result in excessive exploration of suboptimal arms. We thus choose $1/\delta = 1 + \bar{t}\log^2(\bar{t})$, where $\bar{t}$ is the total number of known testing results, and δ is decreasing to zero slightly faster than $1/\bar{t}$ to ensure sublinear regret, and which is shown to improve the regret and empirical performance (Lattimore and Szepesvári 2020). The pseudocode is provided in Figure 3.

4. Agent-Based Model

In the real-world setting, as long as we have daily spatiotemporal data on individuals in a population, the algorithm presented in the previous section can be used to select individuals to test. However, real-world data alone is insufficient to fully understand the impact of different sampling policies. The "real world" represents a single run of how reality is shaped, and does not offer the advantage of running experiments to test counterfactuals. For instance, would the sampling algorithm have worked as well under a different quarantine policy, or under various compliance scenarios? Agent-based models, on the other hand, are particularly effective to answer these types of more general questions.

In this paper, we couple a disease progression model (Hethcote 2000) with an agent-based simulation model (ABM) (Perez and Dragicevic 2009) of infectious spread under different settings. Disease progression is commonly modeled with the deterministic compartmental S/I/R (Hethcote 2000, Cooper et al. 2020), which divides the population into three compartments—susceptible to the disease (S), actively infected with the disease (I), and recovered (or dead) and no longer contagious (R)—and defines transmission rates between the compartments. To account for additional aspects of the COVID-19 transmission, we use an extended S/I/R model that includes individuals that are exposed (E) to the disease (Prem et al. 2020, Silva et al. 2020), and asymptomatic (A) individuals (Koo et al. 2020, Manchein et al. 2020)—an S/E/A/I/R model.

We constructed a data-driven stochastic agent-based model (ABM) of the COVID-19 epidemic in New York City. Based on Hoertel et al. (2020), our ABM model includes four components: (1) synthetic population with demographic characteristics and spatial information that is generated to resemble the New York City; (2) daily interaction network between individuals (agents) in the population; (3) disease dynamics that spreads via interactions, and progresses as an S/E/A/I/R model; and (4) policy makers, as agents, who influence the environment based on the sampling strategies and quarantine policies in place. In the reminder of this section, we describe components (2)–(4) of the ABM. The population generating approach in component (1) is detailed in Appendix EC.1 in the e-companion. In Figure 4, we provide the pseudocode of our model.

4.1. Interaction Data Structure

Our data are composed of a set of individual agents $A = \{a_1, \ldots, a_n\}$, a set of geographic locations $L = \{l_1, ..l_m\}$, and the propensity of the interactions between the agents and the locations: $AL = \{\{\phi_{a_i,l_j}\}_{1\le i\le n, 1\le j\le m}\}$ (how often an agent goes to the location).

Agents in our data are characterized by age, gender, and a list of locations that the agents go to, at a given probability. Locations can take different location types, and are spread across the neighborhoods of New York City. For this paper, we set the following location types: household, workplace, school, station, and supermarket. Two agents may meet if they go to the same location at the same time. The meeting probability at a given location is a function of the location type (e.g., agents that live at the same house will meet with a probability of $\phi_{household} = 1$, whereas agents that go to the same supermarket will meet at a much lower probability: $\phi_{supermarket} \ll 1$). In addition to the different location types, we add a special location type that we call a *mixing location*, in which agents can meet at random.

Figure 3. Algorithm 1 (Pseudocode for Active Sampling (Combinatorial kNN-UCB))

input : Past 14-day contact network $\hat{\mathcal{G}}_0 =< V, E, Z, D >$, oracle ORACLE, trade-off parameter η, daily constraint D_{max}, reward function f, α, β for prior Beta distribution, $S_i = 0$, $F_i = 0$ for $i = 1, 2$

for $t = 1$ *to* T **do**
- Update the contact network $\hat{\mathcal{G}}_t$ based on contact tracing
- Find the network embedding of each node x_v
- Clear any outdated testing results y_v after 14 days
- Initialize the proportion allocated to proliferation $\bar{D}_{max} = 0$
- **for** $l = 1$ *to* D_{max} **do**
  - **[Outer Level]**
  - Draw $\hat{\theta}_i$ from Beta$(S_i + \alpha, F_i + \beta)$ for $i = 1, 2$
  - Choose arm $a_t = \arg\max_i \hat{\theta}_i$
  - **if** a_t == *expansion* **then**
    - Randomly sample a node v that is not in $\hat{\mathcal{G}}_t$ to test, and update y_v
  - **end**
  - **else**
    - $\bar{D}_{max} = \bar{D}_{max} + 1$
  - **end**
- **end**
- **[Inner Level]**
- Get the list of connected components in $\hat{\mathcal{G}}_t$
- **for** *each individual node* v_i *in* $\hat{\mathcal{G}}_t$ **do**
  - Within the connected component, find the number $N_t(i)$ of nodes with known y_v's
  - Define $k_t(i) = \lfloor (N_t(i))^{2/(2+d)} \rfloor$, and build the kNN prediction $\hat{f}_t(i)$
  - Compute the UCB index as

$$\bar{f}_t(i) := \hat{f}_t(i) + \eta \cdot \begin{cases} \frac{1}{\sqrt{e}} \sqrt{(2+d)\log\left(1 + \bar{t}\log^2(\bar{t})\right)}, & \text{if } N_t(i) \le \lfloor e^{1+\frac{d}{2}} \rfloor + 1 \\ \left(N_t(i)\right)^{-1/(2+d)} \sqrt{\log N_t(i) \log\left(1 + \bar{t}\log^2(\bar{t})\right)}, & \text{otherwise} \end{cases}$$

- **end**
- Compute $U_t \leftarrow \text{ORACLE}(V, \mathcal{U}, \bar{f}_t, \bar{D}_{max})$
- Test individuals in set U_t
- Observe the testing results and update y_v
- Compute the number of positive cases under expansion and proliferation, respectively, $\bar{S}_1, \bar{S}_2$
- Compute the number of negative cases under expansion and proliferation, respectively, $\bar{F}_1, \bar{F}_2$
- Update posterior Beta distribution:

$$S_i \leftarrow S_i + \bar{S}_i$$
$$F_i \leftarrow F_i + \bar{F}_i, \text{ for } i = 1, 2$$

end

Mixing locations in our model simulates meeting at public spaces, such as at parks, shops, and theaters.

Agents in our model can potentially be in one of seven disease states, as described below. We formally define $D(A) = \{d(a_1), \ldots d(a_n)\}$ to be the set of disease states of all agents.

4.2. Disease Dynamic

The disease process used in this paper follows the deterministic *S/E/A/I/R* (Susceptible (*S*) / Exposed (*E*) / Asymptomatic (*A*) / Infectious (*I*) / Recovered (*R*)) compartmental model, with death (*D*) rate. We further split the infectious state into three disease stages: Infected asymptomatic (*Ia*), Infected symptomatic (*Is*), and Infected critical (*Ic*). Beyond being a more accurate description of COVID-19, this allows us to flexibly define different policies and different behaviors toward infected individuals according to the state of their disease. Similar to *S/E/A/I/R*, we assume that Recovered (*R*) is an absorbing state, implying that infected individuals can either

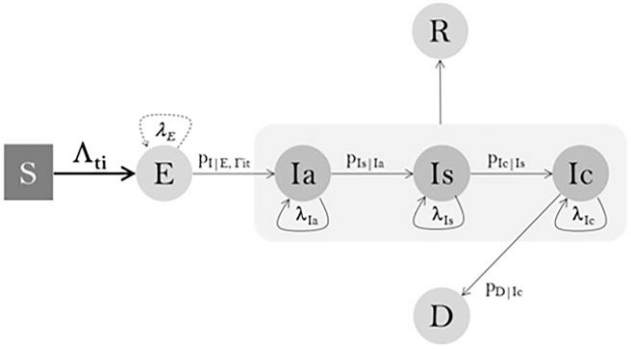
Figure 4. Algorithm 2 (Pseudocode for the ABM)

Input:
  Agents $A = \{a_1, ..., a_n\}$; Locations $L = \{l_1, ..l_m\}$; Exposure risk factor per location $\Gamma(l_j)$;
  Agents-locations interaction $AL = \left\{\{\phi_{a_i,l_j}\}_{1\le i\le n, 1\le j\le m}\right\}$; Number of initial infected agents $\tilde{n}$;
Output:
  Disease state $D(A) = \{d(a_1), ...d(a_n)\}$; meetings log;
Initialize:
(1) Set initial disease state $D(A)$ such that:
  for $\tilde{n}$ agents $d(a) \leftarrow Ia$,
  for $N - \tilde{n}$ agents $k(a) \leftarrow S$;
(2) Set simulation day $\tau \leftarrow 1$;
(3) Initialize set of agents in quarantine $Q \leftarrow \{\}$;
(4) Initialize agent-location log;
while *some agents are still infected* **do**
  // Draw locations per agents:
  for $a_i \in A - Q$ **do**
    Draw daily locations l_{a_i} according to the agent-location interactions propensity AL;
    Append tuples $\left\{\{a_i, l_j \in l_{a_i}\}, \tau\right\}$ to the agent-location log;
  end
  // Get new infected agents:
  for $l_j \in L$ **do**
    Get the list of agents a_{l_j} that visited l_j at time τ from the agent-location log;
    If there are $n_{l_j} > 0$ infected agents in l_j $(d(a \in a_{l_j}) \in \{Ia, Is, Ic\}$, then (1) set the disease status of all ***Susceptible (S)*** agents $\in a_{l_j}$ as ***Exposed (E)***, with ***Exposure risk*** $= \Gamma(l_j)^{n_{l_j}}$; and (2) increase the exposure risk of all ***Exposed (E)*** agents $\in a_{l_j}$ by a factor of $\Gamma(l_j)^{n_{l_j}}$;
  end
  // Progress disease state:
  Progress disease state d_i for all agents according to Figure 5;
  // Put critically infected in self-quarantine:
  for $d_i \in D$ **do**
    switch d_i **do**
      case *Is:* **do**
        Append a_i to Q
      end
      case *R or D:* **do**
        Remove a_i from Q
      end
    end
  end
  // Put agents in enforced-quarantine:
  Interact with policy-makers (sampling policy)
    Send: $\tilde{A} \leftarrow$ set of externally sampled agents to be tested
    Receive: $\tilde{\tilde{A}} \leftarrow$ set of agents to be set in quarantine
    Respond: Append $\tilde{\tilde{A}}$ to Q
  end
  $\tau \leftarrow \tau + 1$ *// Advance simulation time*
end

die or become immune. Figure 5 presents the disease states and the transmission dynamics in this paper.

Transmission probabilities are assumed to be constant across individuals. This is because of the fact that the impact of a patient's covariates on the disease dynamics is still unknown. An exception is the transmission rate between states S and E—Λ_{it}—that determines the rate at which uninfected individuals are exposed to infected individuals, and depends on the individuals' ego-network structure. Once exposed, they become infected with probability of $p_{I|E,\Gamma_{it}}$. The argument Γ_{it} indicates the exposure risk factor, which depends on the number of infected

Figure 5. Transmission Dynamics of the COVID-19 Pandemic



agents that individual i met at time t, and the strength of his or her relationship with them. Specifically, following Madewell et al. (2020), we differentiate the probability of being infected from close connectors (a total of $n_{1_{it}}$), such as housemates or colleagues, from the probability of being infected from random contacts (a total of $n_{2_{it}}$), such as meetings at supermarkets, or metro stations: $\Gamma_{it} = \Gamma_1^{n_{1_{it}}} \times \Gamma_2^{n_{2_{it}}}$.

In infected individuals, the disease may escalate and become mild (Is), at a probability of $p_{Is|Ia}$, and then critical (Ic), at a probability of $p_{Ic|Is}$. At any stage of the disease, the infected patient can recover. The critical disease state (Ic) can result in death, at a rate of $p_{D|Ic}$.

In the results presented in the paper, we set $\lambda_{Ia} = \lambda_{Is} = \lambda_{Ic} = 0.9$, following the average infection period reported by Lauer et al. (2020). Because of the difficulties in differentiating the incubation period from the asymptomatic period, as reported by Roda et al. (2020), we set $\lambda_E = 0$, effectively merging states E and Ia. The transition probabilities used in our results are $\Gamma_1 = 0.02$, $\Gamma_2 = 0.004$, $p_{Ic|Is} = 0.006$, and $p_{D|Ic} = 0.05$. The numbers were set according to the disease statistics, as reporter by the World Meter (World Meter 2020).

We note that the sampling algorithm presented in the previous section only uses data observed as a result of the actual disease diffusion process, and does not depend specifically on the details of how this is implemented in the ABM—all that is required is for the ABM to simulate contacts and movements of individuals in a population. In that sense, our approach will work with any extensions of the model implemented here.

4.3. Policy Makers as Agents

At the end of each ABM simulation epoch, we sample a set of agents to be tested for COVID-19. Once an agent is sampled, his or her disease state and (possibly incomplete) contact network are revealed to policy makers, who can then decide which agents should be put into quarantine. Other than sampled agents, we assume that critically infected agents self-quarantine themselves. The quarantine duration is set for 14 days, following the current global practice. While in quarantine, an agent's disease dynamics continues, but his or her interactions with others stop.

5. Empirical Results

We compare three policies: Contact Tracing (CT), Random testing with CT, and Active Sampling with CT as presented here. Among these options, CT is the current baseline in the United States, since neither random or active testing is currently not implemented. We assume that symptomatic individuals are self-reported and thereafter quarantined, and contacts of infected individuals with a positive test quarantine through standard contact-tracing processes. Beyond CT, other methods considered in comparison throughout the manuscript are proactively testing other individuals beyond self-reports, with the same daily testing capacity (0.5% of the population, to be consistent with daily tests in the United States, as reported by the Covid Tracking Project).

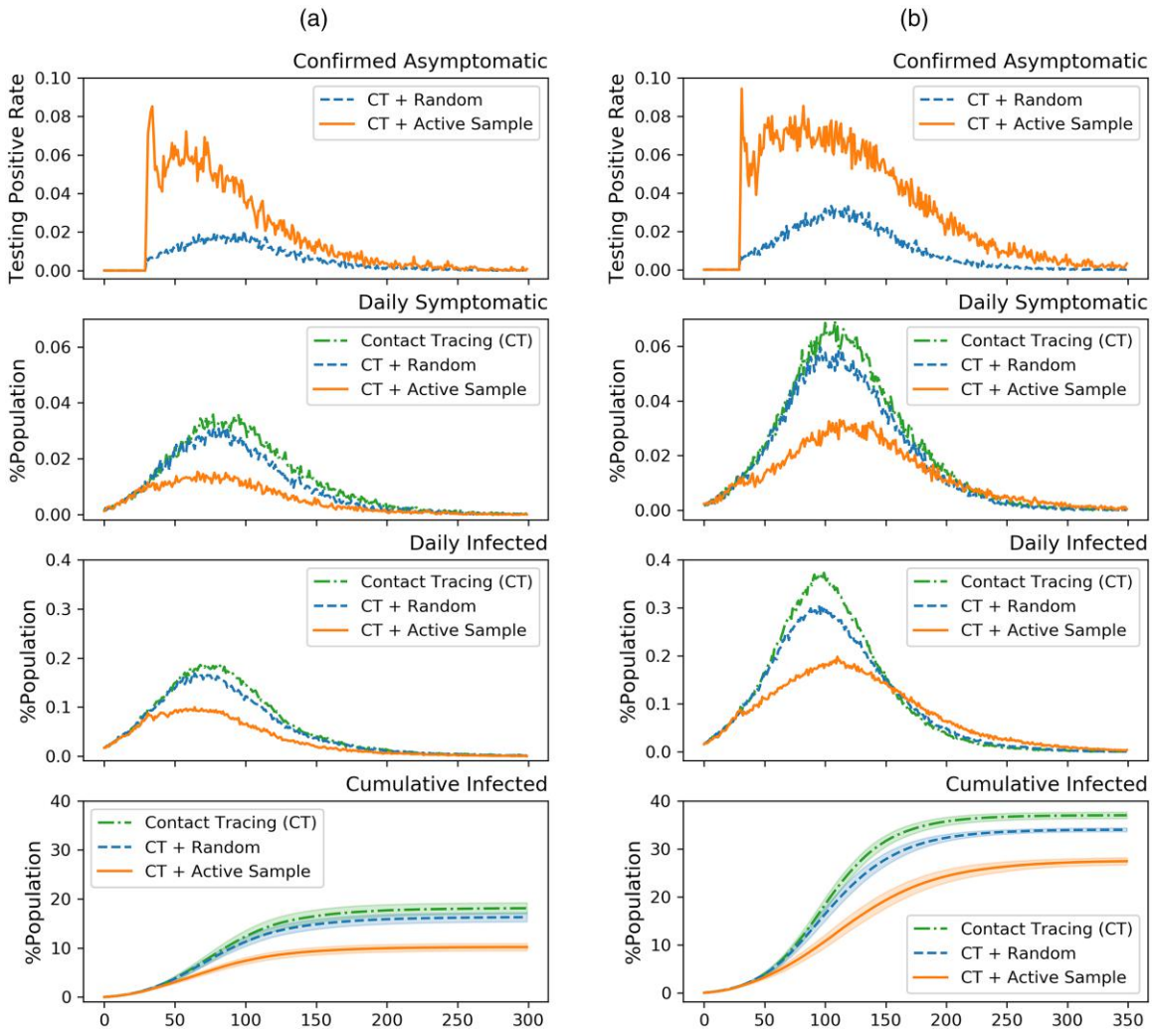
We implemented active testing strategies under different vaccination scenarios and show that a combinational strategy is critical. We use a random vaccination policy that runs parallel to active testing and consider different extents of vaccination that eventually (after 100 days) reach to either $\kappa = 25\%$ or $\kappa = 50\%$ of the population. We chose these to closely reflect numbers in the United States (four months after the start of vaccinations, the United States had fully vaccinated 25% of the population, and the annual percentage of those vaccinated for flu each year is approximately 50%). All of these can of course be varied as needed to reflect realities in different countries. We further vary the compliance level of individuals to self-isolate and quarantine in our empirical study ($\chi \in \{10\%, 50\%\}$). Unfortunately, there is no scientific consensus yet on the percentage of the U.S. population that complies. We discuss the implications of this further below.

We simulate each policy for 50 runs, under each of four scenarios, namely, low/high compliance and low/high vaccination rate. We also simulate the corresponding scenarios without any vaccination in place. We simulate the disease spread for 350 days or until the spread terminates. To account for the fact that testing tool kits and vaccinations are not developed at the beginning of the pandemic, we start the testing and vaccination on day 30. The running time is around 126 hours for 50 runs, under an AWS EC2 r4.4xlarge instance, with 122 GB memory and four 2.3-GHz Intel Xeon E5-2686 v4 processors.

5.1. Main Results

We report our study results in Figures 6 and 7 and in Tables 1 and 2. Figure 6 reports the simulation results under high compliance level and low/no vaccination rate, averaged over 50 runs. It depicts the following: (1) the percentage of the allocated tests returned with a

Figure 6. (Color online) [Infections] Comparison Between Three Testing Policies, Under High Compliance Level ($\chi = 50\%$), with (Low Vaccination Rate ($\kappa = 25\%$)) or Without Vaccination



Notes. (a) With random vaccination policy. (b) With no vaccination policy.

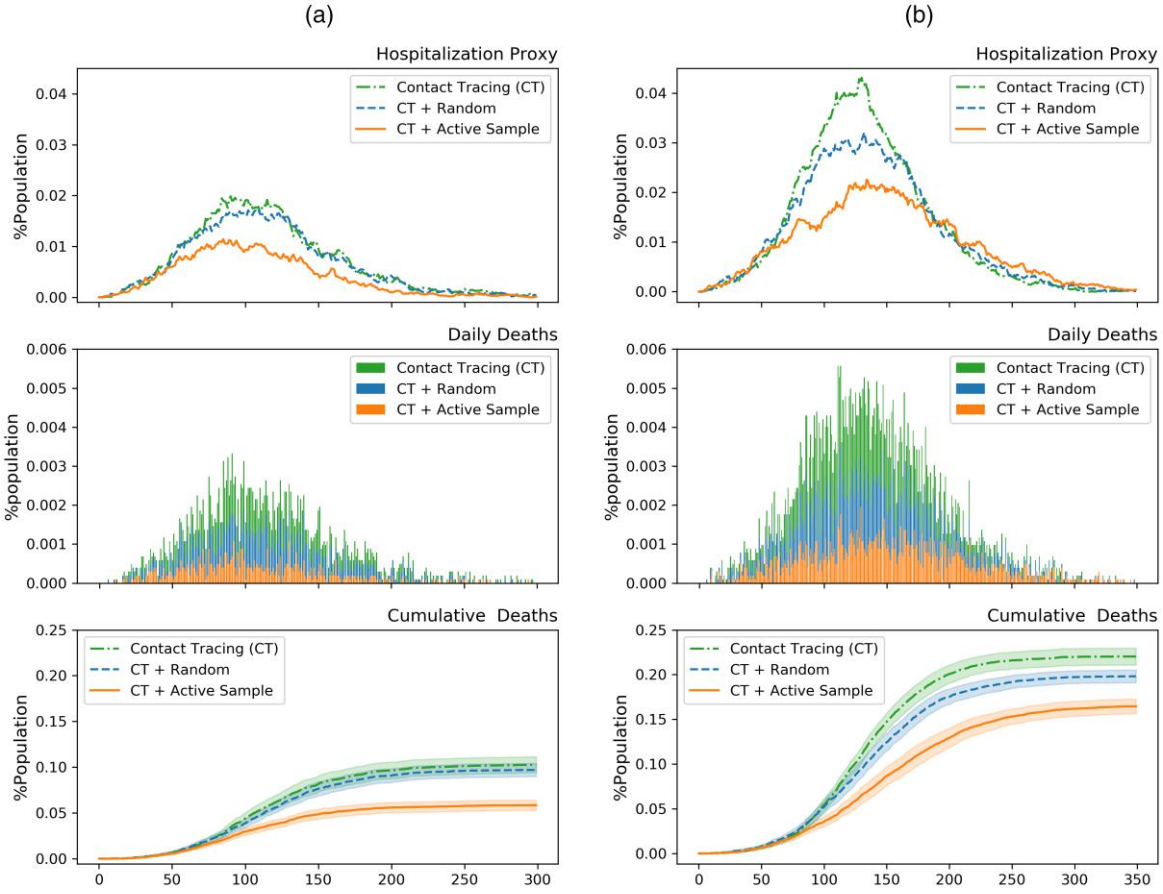
positive result; (2) the percentage of daily confirmed new symptomatic patients; (3) the percentage of daily new infected; and (4) the percentage of cumulative infected within the entire population. Because of the fact that a vast majority of the infected are asymptomatic, the last two statistics, daily/cumulative infected, are, in reality, not directly observable by the policy maker. Figure 7 further examines the corresponding economic measures, namely, hospitalization rates (proxied via the number of critically and severely daily infected individuals) and death rates, with the y-axis depicting percentage of population.

Tables 1 and 2 reports the full comparisons of three testing policies, namely, contact tracing (CT), CT + Random, and CT + Active sample, under four compliance/vaccination scenarios, and without a vaccination policy in place. The corresponding standard errors (SEs) are reported in parentheses. Significance levels are marked relatively to CT to indicate whether the difference in performance is statistically significant.

The main takeaways from these results are the following:

- Even with a high vaccination rate, active sampling is still useful in reducing the extent of the disease. Specifically (from Tables 1 and 2), the death rate significantly drops by 20%–30%; the percentage of infected drops by approximately 30%. These numbers are very significant. In the U.S. population, these numbers map to approximately 10 million fewer cases and 150,000 fewer deaths. With a low vaccination rate, the effect of active sampling is even more notable. For instance, with low vaccination rate and high compliance, the percentage infected drops by ~ 45%.
- Asymptomatic transmission is a significant aspect of COVID-19. Our results show that smart testing strategies generate four times the test positivity rate as

Figure 7. (Color online) [Hospitalization and Deaths] Comparison Between Three Testing Policies, Under High Compliance ($\chi = 50\%$), with (Low Vaccination Rate ($\kappa = 25\%$)) or Without Vaccination



Notes. Hospitalization is proxied based on how many are critically ill on each given day. (a) With random vaccination policy. (b) With no vaccination policy.

Table 1. Cumulative Percentages of Infected Across Different Compliance Level and Vaccination Rate

		Random vaccination		No vaccination	
Scenarios	Testing policy	% Population	SE	% Population	SE
Low compliance level	Contact tracing (CT)	22.8311	(0.5748)	46.0067	(0.1907)
Low vaccination rate	CT + Random	22.2713	(0.4952)	44.6116***	(0.1952)
	CT + Active sample	17.5112***	(0.5343)	41.3715***	(0.2123)
Low compliance level	Contact tracing (CT)	11.6922	(0.5338)		
High vaccination rate	CT + Random	10.3761*	(0.5569)		
	CT + Active sample	8.4913***	(0.4547)		
High compliance level	Contact tracing (CT)	18.1235	(0.5831)	37.0027	(0.3491)
Low vaccination rate	CT + Random	16.2956**	(0.4838)	33.9982***	(0.2088)
	CT + Active sample	10.1960***	(0.4857)	27.4378***	(0.3999)
High compliance level	Contact tracing (CT)	9.1383	(0.5171)		
High vaccination rate	CT + Random	8.1184†	(0.4622)		
	CT + Active sample	5.9998***	(0.3738)		

Notes. Significance levels are marked relatively to CT. % Population column reports the mean percentage of population infected under different compliance levels (10% and/or 50%) and different vaccination rates (~ 0.25% and/or ~ 0.5% per day), averaged over 50 simulation runs. The corresponding standard errors (SE) are reported in parentheses.

†$p < 0.1$; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

Table 2. Cumulative Percentages of Deaths Across Different Compliance Level and Vaccination Rate

		Random vaccination		No vaccination	
Scenarios	Testing policy	% Population	SE	% Population	SE
Low compliance level	Contact tracing (CT)	0.1370	(0.0046)	0.2750	(0.0061)
Low vaccination rate	CT + Random	0.1344	(0.0042)	0.2660	(0.0057)
	CT + Active sample	0.1029***	(0.0045)	0.2455***	(0.0051)
Low compliance level	Contact tracing (CT)	0.0623	(0.0039)		
High vaccination rate	CT + Random	0.0575	(0.0037)		
	CT + Active sample	0.0498**	(0.0032)		
High compliance level	Contact tracing (CT)	0.1027	(0.0046)	0.2202	(0.0050)
Low vaccination rate	CT + Random	0.0970	(0.0038)	0.1981***	(0.0037)
	CT + Active sample	0.0573***	(0.0033)	0.1645***	(0.0043)
High compliance level	Contact tracing (CT)	0.0524	(0.0038)		
High vaccination rate	CT + Random	0.0446†	(0.0029)		
	CT + Active sample	0.0353***	(0.0030)		

Notes. Significance levels are marked relatively to CT. % Population column reports the mean percentage of deaths under different compliance levels (10% and/or 50%) and different vaccination rates (~ 0.25% and/or ~ 0.5% per day), averaged over 50 simulation runs. The corresponding standard errors (SE) are reported in parentheses.
$^{\dagger}p < 0.1$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

random testing. Catching the asymptomatic carriers early is a critical part of curbing the spread and decreasing infection/death rates; our active sampling method is effective at doing this, which is a key reason for its effectiveness.

- Load on hospitals at peak times, a crucial aspect of managing COVID-19, drops, according to Figure 7, by 50% when implementing smart testing (with or without vaccinations).
- Whereas the focus of this paper is not entirely on vaccination, the results emphasize the importance of vaccination. Even with low vaccination rates, the death rate can drop by over 50% (see Table 2).
- Is random testing useful? Our results provide mixed evidence. It is useful at identifying asymptomatic individuals when the spread is large, but with reasonably large vaccination numbers, it does not necessarily help in significantly lowering death rates. In pockets where compliance rates are high, the spread is increasing, but vaccination rates are low (perhaps because of availability), so random testing is likely to be an effective complement to contact tracing alone.
- One major unknown is the extent of compliance in the population (research has shown that this varies from low single digits to as high as 100%) to self-quarantining requirements. The results show that testing policies are less effective when compliance is low. This is not surprising, since one of the main advantages of early identification of asymptomatic individuals is cutting the infection chain through quarantine policies. When the compliance is low, this advantage diminishes significantly. One possibility for policy, which we turn to next, is to consider compensation strategies that can increase compliance (Bodas and Peleg 2020).

5.2. Ablation Analysis

We conduct an ablation analysis to examine the effectiveness of each major design component. To test whether the outer-level exploration (outside of the observed contact network) is needed, we remove the outer-level TS, and only use *k*NN-UCB to select individuals from the observed network (*k*NN-UCB). To test the proposed quantification of uncertainty, we remove the optimistic UCB term and only choose nodes based on *k*NN prediction as the inner-level selection policy (*k*NN-TS). We then remove the aforementioned two components, and test based directly on *k*NN predictions on observed contact network without consideration of the unobserved uncertainty (*k*NN). As an alternative specification, we consider *k*NN-ϵ-greedy, which represents the outer level as an additional arm. To examine whether the node representation is effective, we replace the constructed embeddings with traditional network measures (including observed degree centrality of focal node, average degree centrality of its neighbors, average betweenness of its infected neighbors, number of infected neighbors) (NM-UCB-TS) and compare the model performance.

We also test whether the specific nonparametric specification is effective by comparison with other cost-effective baseline methods. First, we incorporate two simpler heuristics, maximum observed degree (MOD), which greedily chooses the node having the maximum observed degree, and maximum number of infected neighbors (MIN) that selects the node having the highest number of infected neighbors. Second, we consider a state-of-the-art parametric (linear) bandit algorithm (SW-LinUCB) proposed for a nonstationary environment (Cheung et al. 2022). As the representation space resulting from DeepWalk can change over time, the node

embeddings cannot be incorporated in a parametric model (which assumes a static contextual space). With this being said, we hence use the network measures as the contextual representation of each individual. Although the original SW-LinUCB policy requires the knowledge of variation budget and/or simulation time T in advance, without knowing this information we set the window size to be 14 days according to the special characteristics of COVID-19.

We note that the number of possible scenarios to consider can be large (even potentially unbounded). Given this, we suggest that the ablation analyses are informative mainly for the specific scenario(s) under consideration. Here, we present these results (in Table 3) for two scenarios (high compliance, with and without vaccination). For example, under the *k*NN-UCB testing policy, death rates increase significantly. The MOD and MIN policies are poor in terms of both infections and death rates, compared with the active sample policy. The results show that all other testing policies fall short compared with the main active sampling policy; in this sense, this suggests that no single component alone can be dropped and that the combination as presented here is important to drive value.

5.3. Testing Accuracy and Compliance

In the main results above, we assumed that test results were accurate and that users would consent to being tested when selected. Since in practice these may not always be the case, this section presents results under imperfect testing and less-than-perfect user compliance to testing. We also experimented with an alternative contact network structure, various lockdown policies, and a scenario that during the disease diffusion, there are times that the individuals in the *Ic* state exceeds the hospitalization capacity. Additional results and discussions are presented in Appendix EC.2 in the e-companion.

5.3.1. Imperfect Testing Results. In practice, particularly in the early stages of a new infection, testing might not be perfect. In this section, we study what happens in our framework with a test that is 90% accurate. Specifically, we assume that 90% of the time the test will provide an accurate result and that 10% of the time the result will be obtained randomly as tossing a 50-50 coin (of course, any of these can be varied to explore their effects). For illustration, the set of experiments shown in Figure 8 is run under high compliance ($\chi = 50\%$) and a low vaccination rate ($\kappa = 25\%$). Interestingly, both the random testing and active sampling are better than their counterparts before (with perfect testing results). But this better performance comes at the cost of putting more individuals in quarantine than under perfect testing. For instance, consider the case when approximately 10% of the population is infected. In this case, random testing of 100 individuals under this imperfect testing will result, in expectation, of 9 positive individuals correctly identified, but 4.5 individuals who are not infected also sent to quarantine (along with their contacts). This increase in the quarantined group will likely control infection rates—not because the correct group was identified, but simply due to more people quarantined. Whereas active testing could also "benefit" from this, the effect is less pronounced since the group that active testing identifies tends to typically have a higher incidence of infection than the base rate. Nevertheless, our results show that active sampling is still significantly better than the alternatives.

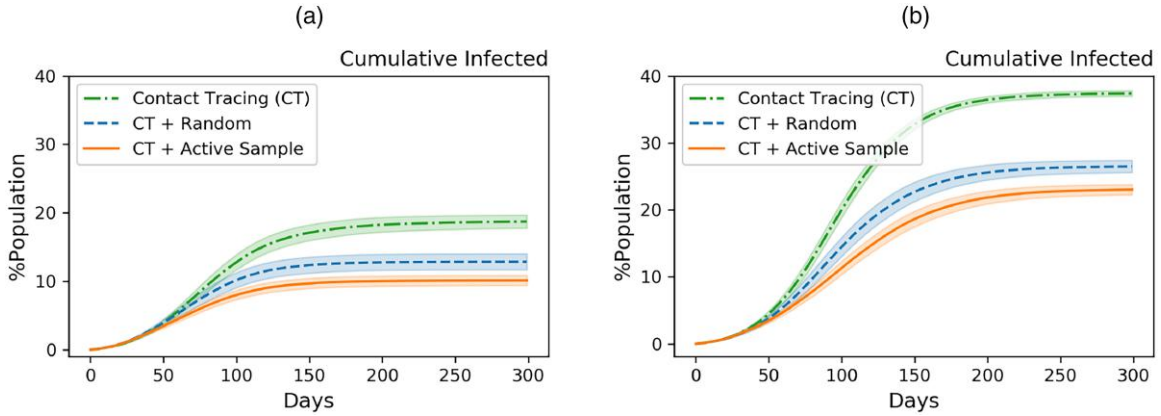
5.3.2. Partial Compliance to Testing. The effectiveness of policy measures to slow the spread of COVID-19 will inevitably depend on the commitment of individuals' compliance and their change of behavior. Whereas we have studied a different compliance level to quarantine recommendation, here we additionally consider an individual's potential noncompliance to testing when presymptomatic or asymptomatic. Table 4 is run under a high compliance ($\chi = 50\%$) level to quarantine order and a low vaccination rate ($\kappa = 25\%$). Specifically, in the high compliance setting, we assume that with probability of 80%, any individual being selected for testing will actually get tested, whereas the compliance probability is

Table 3. Cumulative Percentages (% Population) of Infected/Deaths Under High Compliance

	Percentages of infected		Percentages of deaths	
Testing policy	Vaccination	No vaccination	Vaccination	No vaccination
CT + Active sample	**10.1960**	**27.4378**	**0.0573**	**0.1645**
*k*NN-UCB	11.3987*	28.5452*	0.0684*	0.1850**
*k*NN-TS	11.1882†	28.1924†	0.0654†	0.1766*
*k*NN	12.3278***	29.1807***	0.0749***	0.1884***
*k*NN-ϵ-greedy	11.2205†	28.3884*	0.0675*	0.1772*
NM-UCB-TS	11.2672†	31.0350***	0.0696*	0.2110***
MOD	16.2421***	35.8405***	0.1007***	0.2395***
MIN	13.2738***	32.6216***	0.0833***	0.2117***
SW-LinUCB	15.9699***	33.9083***	0.0952***	0.2194***

Notes. Significance levels are marked relatively to vanilla active sampling policy (CT + Active sample). % Population column reports the mean percentage of infected/deaths, averaged over 50 simulation runs.
$^{\dagger}p < 0.1$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

Figure 8. (Color online) Comparison Between Three Testing Policies, Under Imperfect Testing Results



Notes. (a) With random vaccination policy. (b) With no vaccination policy.

lowered to 40% in the low compliance scenario.[1] The corresponding infection and death rates are reported in Table 4. It is worth noting that the mitigation effectiveness of random testing quickly diminishes to the performance of CT in the setting of low compliance to testing, whereas the smart testing algorithm that adaptively allocates the testing kits can still generate a significant improvement. This is likely due to the fact that the nodes identified by the active testing procedure are mostly informative ones; hence, if even a moderate percentage of those comply, we see significant benefits from a disease mitigation perspective.

6. Extensions to Non-stationarity

Although the risk of infection of each individual is changing over time according to disease diffusion and movement patterns, the scenarios we considered before in general assume that the system remains stationary over time (i.e., human behaviors remain unchanged, diffusion process remains the same). In this section, we consider scenarios that capture some specific types of non-stationarity. First, we consider the case where individuals' compliance to quarantine is changing corresponding to the spreading of the disease, which can be a result of either a change of the government regulation, or social norm. Second, we modify the ABM to explicitly model the emergence of a second variant with different diffusion characteristics. We show that, under these two scenarios, the smarting testing framework can still significantly reduce the total number of infected and deaths compared with contact tracing. We also present variants of our framework to potentially better accommodate non-stationarity.

Table 4. Cumulative Percentages of Infected/Deaths Under Partial Compliance to Testing

		Low vaccination rate		No vaccination	
Compliance level	Testing policy	% Population	SE	% Population	SE
Cumulative percentages of infected					
High compliance	Contact tracing (CT)	18.4318	(0.5304)	37.1822	(0.2173)
	CT + Random	16.9150**	(0.3653)	34.6014***	(0.2120)
	CT + Active sample	11.8168***	(0.5104)	28.6462***	(0.3563)
Low compliance	CT + Random	18.2694	(0.4768)	36.0340***	(0.2161)
	CT + Active sample	12.4671***	(0.6318)	32.8213***	(0.2323)
Cumulative percentages of deaths					
High compliance	Contact tracing (CT)	0.1182	(0.0044)	0.2407	(0.0059)
	CT + Random	0.1070*	(0.0042)	0.2262*	(0.0044)
	CT + Active sample	0.0703***	(0.0039)	0.1838***	(0.0048)
Low compliance	CT + Random	0.1134	(0.0041)	0.2407	(0.0056)
	CT + Active sample	0.0721***	(0.0048)	0.2185***	(0.0051)

Notes. Significance levels are marked relatively to CT. % Population column reports the mean percentage of infected/deaths, averaged over 50 simulation runs.
$^{\dagger}p < 0.1$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

6.1. Algorithmic Extensions to Non-stationarity

We implemented two variants of our proposed algorithm that are designed to adapt to nonstationary environments. One approach is to ignore historical observations made beyond some number of time periods in the past (Besbes et al. 2014, Trovo et al. 2020). Correspondingly, our first extension imposes a sliding window ω (named *SW*) on the MAB learning process. Whereas the vanilla combinatorial *k*NN-UCB proposed before has already taken the 14-day (with $\omega = 14$) moving average into consideration (see line 4 of Figure 3), here we additionally modify the exploration term such that the total number of data sample $\bar{t}$ is recomputed as the number of known testing results within the past ω days. In addition, for the outer-level Thompson sampling, a sliding window of size ω is also imposed. Specifically, when updating the posterior distribution, we use $S_{i,t,\omega}$ to record the number of positive cases within the sliding window ω under expansion and proliferation, respectively; and $F_{i,t,\omega}$ denotes the number of negative cases within the past ω days. The posterior distribution at day t is then $\text{Beta}(S_{i,t,\omega}+\alpha, F_{i,t,\omega}+\beta)$ for $i=1, 2$, where the prior is chosen as an uninformative prior. The pseudocode is provided in Algorithm EC.3.1 in the e-companion.

As in what follows, we explicitly model a second COVID variant, creating an *abruptly changing* environment. The framework itself can start to "work" automatically from a depletion of its window (of max ω) when the change point is detected; or it can directly be informed by the policy maker who detects this and "restarts" the framework at some point. To this end, we additionally consider a variant of the sliding window UCB (named *SW-Restart*, with pseudocode presented in Algorithm EC.3.2 in the e-companion) that restarts and refreshes the buffer (after three days of the first occurrence of the second variant to give the policy maker some time to identify the new variant).

Next, we consider discounted combinatorial *k*NN-UCB (named *Discounted*) that is based on the ideas of Garivier and Moulines (2011) and Raj and Kalyani (2017). Garivier and Moulines (2011) and Raj and Kalyani (2017) proposed nonstationary extensions of UCB and/or Thompson sampling policies for *k*-armed bandits that rely on a discount factor $\gamma \in (0,1)$ to discount the effect of past observations. In line with this research, we modify the posterior Beta distribution of the outer level (line 29 in Figure 3) to $S_i \leftarrow \gamma S_i + \overline{S}_i$, $F_i \leftarrow \gamma F_i + \overline{F}_i$, for $i = 1, 2$.

For the inner level, on day t, the discounted number of known individuals that are connected to the focal node v_i is given by $N_t(\gamma,i) = \sum_j \gamma^{t-s_j}$, where s_j is the time when node v_j was tested. The corresponding neighborhood size is $k_t(\gamma,i) = \lfloor (N_t(\gamma,i))^{2/(2+d)} \rfloor$. Then the discounted estimate of the reward is given by

$$\hat{f}_t(\gamma,i) = \frac{1}{K_t(\gamma,i)} \sum_{j=1}^{n} \gamma^{t-s_j} y_j \mathbb{1}\{v_j \in \mathcal{N}_{k_t(\gamma,i)}\},$$

$$K_t(\gamma,i) = \sum_{j}^{n} \gamma^{t-s_j} \mathbb{1}\{v_j \in \mathcal{N}_{k_t(\gamma,i)}\},$$

where $\mathcal{N}_{k_t(\gamma,i)}$ denotes the $k_t(\gamma,i)$-neighborhood of node v_i. With the discounted number of cumulative known cases as $\bar{t}(\gamma) = \sum_j \gamma^{t-s_j}$, the discounted exploration term is thus

$$(N_t(\gamma,i))^{-1/(2+d)} \sqrt{\log N_t(\gamma,i) \log(1+\bar{t}(\gamma)\log^2(\bar{t}(\gamma)))}, \quad \text{or}$$
$$\frac{1}{\sqrt{e}} \sqrt{(2+d)\log(1+\bar{t}(\gamma)\log^2(\bar{t}(\gamma)))}, \text{ if } N_t(\gamma,i) \le e^{1+\frac{d}{2}}+1.$$

The pseudocode of this nonstationary variant is provided in Algorithm EC.3.3 in the e-companion.

Lastly, this discounted algorithm considers all the past observations. However, in mitigating COVID-19, cases over 14 days old are seldom related to further testing results. To this end, we experimented with one additional discounted version (Algorithm EC.3.4 in the e-companion) which is combined with sliding window with only the past 14 days records when computing the exploration/exploitation terms (named: *14-day Discounted*). In the experiments, the discounted factor is set to be 0.975 following the work by Garivier and Cappé (2011).

In summary, we implement two main variants: SW (with and without restart), and Discounted (with and without SW).

6.2. Changing Compliance Level

The existing literature has studied the impact of risk perception on citizen's compliance with preventive measures during the outbreak of COVID-19, and earlier SARS and H1N1 (Asnakew and Kerebih Asrese 2020, Ning et al. 2020, Franzen and Wöhner 2021). Beca-Martínez et al. (2021) found that having some concerns about how fast COVID-19 was spreading was significantly affecting the adherence to quarantine order. In what follows, we follow the findings by Beca-Martínez et al. (2021) and set the compliance level for fast/very fast disease spreading (measured by daily new cases) to $\chi = 50\%$ and let the compliance level change (with an adjusted odds ratio of 1.59) dynamically based on the number of daily new cases throughout the disease diffusion process.

We experimented with both a low vaccination rate (~ 0.25% per day) and no vaccination and report the results in Table 5 (the significance marked relative to CT + Active sample). The results show that by explicitly modeling additional sources of non-stationarity, our proposed framework still yields significant improvement

Table 5. Cumulative Percentages of Infected/Deaths Under Changing Compliance Level

	Low vaccination rate		No vaccination	
Testing policy	% Population	SE	% Population	SE
Cumulative percentages of infected				
Contact tracing (CT)	19.2129***	(0.4744)	37.8082***	(0.2174)
CT + Random	16.3221***	(0.4679)	34.7623***	(0.1881)
CT + Active sample	11.0529	(0.4389)	27.9864	(0.2681)
CT + Active sample (SW)	10.7945	(0.5193)	27.9664	(0.2681)
CT + Active sample (Discounted)	10.7921	(0.3642)	27.6743	(0.2079)
CT + Active Sample (14-day Discounted)	10.1983†	(0.4968)	27.3944†	(0.2566)
Cumulative percentages of deaths				
Contact tracing (CT)	0.1275***	(0.0048)	0.2508***	(0.0057)
CT + Random	0.1009***	(0.0046)	0.2260***	(0.0044)
CT + Active sample	0.0689	(0.0043)	0.1817	(0.0045)
CT + Active sample (SW)	0.0676	(0.0046)	0.1803	(0.0052)
CT + Active sample (Discounted)	0.0666	(0.0038)	0.1795	(0.0044)
CT + Active sample (14-day Discounted)	0.0633	(0.0039)	0.1794	(0.0051)

Notes. Significance levels are marked relatively to vanilla active sampling policy (CT + Active sample). % Population column reports the mean percentage of infected/deaths, averaged over 50 simulation runs. The corresponding standard errors (SE) are reported in parentheses.
†$p < 0.1$; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

over CT and CT + Random, similar to the magnitude of our main results.

It is worth noting that even though the SW and Discounted variants are slightly better than the vanilla active sampling algorithm, the improvement is not significant. The 14-day Discounted version yields the best performance, which is marginally significant when compared with the vanilla active sampling algorithm in terms of infection with or without vaccination. No significant difference was found regarding death rates.

6.3. New COVID-19 Variant

Similar to the breakout of Omicron, in this section we modified the diffusion process to model a second COVID variant, with the diffusion parameters set according to the statistic regarding the Delta variant versus the Omicron variant, as reported in recent literature. Specifically, people infected with Delta spread the infection to roughly 11% of their household members, whereas those who had Omicron spread it to almost 16%; and people infected with Delta spread the infection to roughly 4% of people they came in contact with outside their home, whereas those with Omicron passed it to 8% of people (Allen et al. 2022, Mallapaty 2022). Compared with Delta, proportion of asymptomatic infections is doubled for omicron, as high as 80%–90% (Murray 2022). In a matched study, there were 53 hospitalizations (0.6%) and 3 deaths (0.03%) among matched Omicron cases compared with 129 hospitalizations (1.4%) and 26 deaths (0.3%) among matched Delta cases (Ulloa et al. 2022). We thus define the S/E/A/I/R transition probabilities between disease states accordingly.

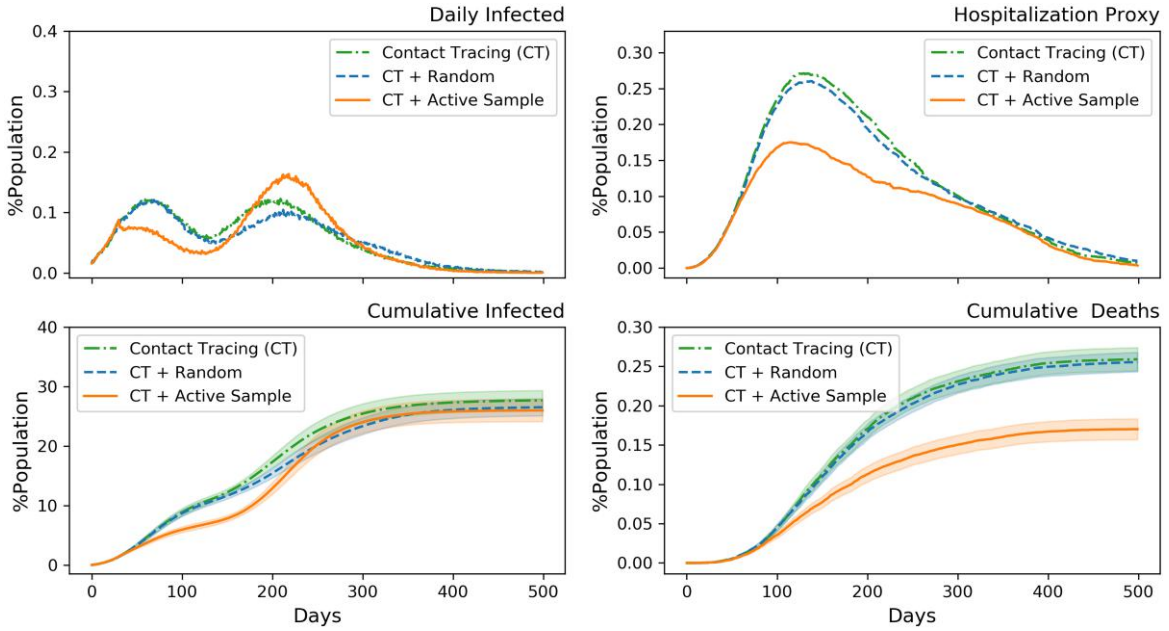
In the following experiments, we simulate the second variant to start after 90 days, with the first variant continuing to still spread simultaneously. The results are reported in Table 6 and Figure 9. Here, again, the results show that our proposed vanilla algorithms perform as well as their nonstationary variants. In fact, the 14-day

Table 6. Cumulative Percentages of Infected/Deaths with Two COVID Variants

	Percentage of infected		Percentage of deaths	
Testing policy	% Population	SE	% Population	SE
Contact tracing (CT)	27.9485†	(1.2265)	0.2643***	(0.0117)
CT + Random	25.8009	(1.0696)	0.2612***	(0.0097)
CT + Active sample	24.7428	(1.4749)	0.1669	(0.0099)
CT + Active sample (SW)	24.7124	(1.1118)	0.1645	(0.0101)
CT + Active sample (SW-Restart)	24.6875	(1.3309)	0.1633	(0.0084)
CT + Active sample (Discounted)	27.3710	(1.2125)	0.1732	(0.0071)
CT + Active sample (14-day Discounted)	27.4725†	(1.1228)	0.1717	(0.0071)

Notes. Significance levels are marked relatively to vanilla active sampling policy (CT + Active sample). % Population column reports the mean percentage of infected/deaths, averaged over 50 simulation runs.
†$p < 0.1$; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

Figure 9. (Color online) Comparison Between Three Testing Policies, Under High Compliance ($\chi = 50\%$) and Low Vaccination Rate ($\kappa = 25\%$), Averaged over 50 Runs



Note. The second variant occurs at day 90.

Discounted model performs (marginally) significantly worse than the vanilla model.

Our main problem setting is more similar to slowly-varying nonstationary bandit formulation, which allows time variations of the unknown mean values (or the variations of the unknown parameters) (Besbes et al. 2014, Cheung et al. 2022) with no restriction on the number of times the mean value can change. A second COVID-19 variant at a different diffusion rate then can be treated as a source of abruptly-changing environment aside from the slowly-varying non-stationarity. Taken together, the results for nonstationary environments show that the previous algorithm performs quite well with no modifications needed. The reason is that the algorithm already takes into account (implicitly) the changes in the environment, by clearing outdated tested samples more than 14 days old (line 4 of Figure 3), in a similar manner as using a sliding window. This, coupled with the fact that knowledge of the contact network is growing, helps the algorithm focus on nodes that are likely to be infected given the current known infections in the population. Also, the k-Nearest Neighbor is shown to yield a robust nonstationary classification performance, provided it is trained based on only a certain number of the most recent data points received (Alippi and Roveri 2008, Bifet et al. 2013, Cervantes et al. 2018). Thus, the additional explicit formulation of non-stationarity appears to add only marginal information that is not reflected in the performance. It is possible, of course, that our current non-stationary variants can be improved, and also the types of non-stationarity can be different; we will leave a more comprehensive treatment of these questions to future work.

7. Fairness Considerations

We have seen that algorithmic approaches that optimize various metrics could suffer from major fairness issues (Ahmad et al. 2020). To this end, we present a detailed treatment of how the active sampling framework affects fairness, including offering variants of the approach that can take some fairness constraints directly into the process. We note at the outset that defining fairness itself is an important consideration, and the literature offers many different perspectives, including the distinction between procedural and outcome-based fairness (De Cremer et al. 2010), which is the perspective we take here. In outcome-based fairness, we would examine how infection rates and/or death rates vary based on attributes such as race; procedural fairness instead focuses on whether these races are allocated tests proportionally—either to their base rate in the population (*FairPopulation* in our implementation) or to the proportion of known active cases in each race (*FairActiveCases* in our implementation). As we show below, our proposed active sampling method demonstrates outcome-based fairness (for reasons we note later). We also show that our framework allows us to incorporate constraints that can ensure procedural fairness, which is important in practice as well.

We examine the fairness of our proposed algorithm across five different race/ethnicity (Asian, Black,

Hispanic, White, and Other) reported by the NYC Department for the Aging. and by family structure (see more details in Appendix EC.1 in the e-companion). We primarily associate race with location in the city, and the meeting probabilities that this affects. Some races are in more dense areas, and if a subpopulation in that geography is affected, then it is likely to rapidly spread to others (within the same race). We do not model race-specific disease diffusion probabilities here and acknowledge that those may also be important to consider in future work. We also note that in practice race/ethnicity might be correlated with movement and/or work patterns, but there are no detailed data on these patterns. Hence, the results reported here should be mainly used as an illustration of how our framework can be made to be fairness-aware with the right data.

The results in Table 7 show that our vanilla method already demonstrates outcome-based fairness; the infection rate across races is not significantly different. To quantify the extent of fairness, we compute the entropy of the infection rates across races under each simulation run, and Table 7 reports the average over 50 runs. Compared with the perfect outcome-based fairness that achieves the maximum possible entropy of 1.6094, and the scenario of no intervention policy (entropy = 1.6084), our proposed method does not seem to add significant disparate impact to the outcome. The reason is that the vanilla active sampling policy does do a good job at following hot spots with some small delays. The exploitation part focuses on the current hot spots, and our design of the exploration compartments helps identify new hot spots beyond the known contact network.

Still, the fact that our framework allowed for the relatively straightforward incorporation of procedural fairness into the sampling policy is important. We implemented two variants of the smart testing algorithm with two fairness constraints. On each day, the fairness constraints were imposed separately on the inner level and outer level with a daily ratio ϱ_t across the races/ethnicity. For example, out of the D testing kits that are allocated to outer-level expansion, $\varrho_{t,Asian} \times D$ capacity is allocated to the Asian population. The two variants then consider two possible daily ratios ϱ_t. The first fairness constraint (*FairPopulation*) is based on the initial population ratio in the data (Asian ≈ 0.1241, Black ≈ 0.1073, Hispanic ≈ 0.2670, White ≈ 0.4691, and Other ≈ 0.0325). The second fairness constraint (*FairActiveCases*) is set dynamically according to the *observed* daily active cases within each race/ethnicity group. As reported in Table 7, these two variants also demonstrate outcome-based fairness as well, but we also know (by design) that they are procedurally fair as well. The distinction may be more important in real-life scenarios where the race may influence aspects such as access to healthcare (which we do not model here).

Recall from Table 1 that under the active sample policy without fairness constraint (vanilla policy), the overall infected population rate was on average 10.1960% (SE = 0.4857). In comparison, with fairness constraints in place, the mitigation effectiveness is slightly lower than the vanilla smart testing policy, with infections rate of 10.8317% (SE = 0.3992) under *FairActiveCases* and 11.3415% (SE = 0.4788) under *FairPopulation*. Meanwhile, the death rate under *FairActiveCases* is 0.0699% (SE = 0.0040) and 0.0725% (SE = 0.0042) under *FairPopulation*, compared with a lower rate of 0.0573% (SE = 0.0033) under the vanilla active sampling policy.

Next, we raise the following question: By how much do we need to increase the testing capacity to achieve a statistically similar infectious and death rate, with the fairness constraint in place. In Table 8, we show that roughly by increasing the testing capacity by 6%–7%, there is no statistically significant difference in mean mitigation effectiveness under the fairness constraint compared with the vanilla active sampling policy. At a high level, this suggests two things. First, imposing fairness on testing alone can provide (fairness) benefits but at a slight cost to the mitigation performance, which is not unexpected. Second, and more importantly, this shows that the active testing policy can achieve the same outcomes and with fairness if we are willing to slightly increase daily testing capacities, which is a small price to pay relative to the cost of being potentially unfair in a public health context.

Table 7. Cumulative Percentages of Infection per Race

	No policy		Vanilla		*FairActiveCases*		*FairPopulation*	
Races	Percentage	SE	Percentage	SE	Percentage	SE	Percentage	SE
Asian	0.2478	(0.0358)	0.1047	(0.0358)	0.1108	(0.0292)	0.1157	(0.0351)
Black	0.2339	(0.0332)	0.0991	(0.0324)	0.1058	(0.0283)	0.1100	(0.0333)
Hispanic	0.3405	(0.0343)	0.1009	(0.0348)	0.1065	(0.0284)	0.1114	(0.0322)
White	0.2455	(0.0338)	0.1026	(0.0348)	0.1090	(0.0387)	0.1146	(0.0353)
Other	0.2457	(0.0346)	0.1007	(0.0352)	0.1124	(0.0324)	0.1153	(0.0355)
Entropy	1.6084		1.6060		1.6056		1.6067	

Notes. Simulation was run under high compliance levels (50%) and low vaccination rate (~ 0.25%), averaged over 50 simulation runs. The corresponding standard errors (SE) are reported in parentheses.

Table 8. Cumulative Percentages of Infected and Deaths Under Different Daily Testing Budgets

Cumulative percentages of infected							
	Vanilla	*FairActiveCases*					
Daily budget	**100**	**100**	**103**	**105**	**106**	**107**	**108**
% Population	10.1960	10.8317	10.7463	10.3946	10.2937	10.2706	10.1130
	(0.4857)	(0.3992)	(0.4128)	(0.4805)	(0.4709)	(0.4946)	(0.4345)
Cumulative percentages of deaths							
	Vanilla	*FairActiveCases*					
Daily budget	**100**	**100**	**103**	**105**	**106**	**107**	**108**
% Population	0.0573	0.0699**	0.0694**	0.0657*	0.0638*	0.0630	0.0624
	(0.0033)	(0.0040)	(0.0038)	(0.0037)	(0.0041)	(0.0043)	(0.0033)

Notes. Significance levels are marked relatively to vanilla active sample policy. % Population reports the mean percentage of infected/deaths, averaged over 50 simulation runs. The corresponding standard errors (SE) are reported in parentheses.
$^{\dagger}p < 0.1$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

8. Policy Implications and Conclusion

The method presented here can be used by policy makers to proactively determine whom to test if the following information can be provided: (1) a list of all individuals in the population; (2) the set of known infected individuals in the population; (3) the daily contact network for individuals—ideally, this should be provided for all individuals that were tested positive, but the method can work even with partial information (daily contact network, for instance, can be computed easily from readily available mobile tracking data); and (4) if available, a partial mapping of individuals to locations that they can be associated with (e.g., household, workplaces, schools). Our method can also be used by policy makers to examine carefully constructed scenarios in a world with smart testing strategies.

Active sampling, as we demonstrated here, should become an important aspect of testing programs for public health. Policy makers worldwide should start developing frameworks to operationalize these ideas in a manner that suits their local culture and context (e.g., specific privacy constraints, compliance scenarios, and how much technology support they might have to both operationalize and track progress). Currently, to the best of our knowledge, there is very limited active sampling ideas both in practice, as well as in the public health literature, and this paper, therefore, lays a lot of the groundwork that can be built on, both in practice and in research. We highlight specific policy implications below.

- In general, it is true that "testing more" can help control disease spread, particularly, if compliance to quarantine measures is reasonable. However, active testing is not the same as "testing more," as we showed in this work. Under the same discretionary testing capacity, the active testing strategy presented here provides significantly greater disease control than other alternatives by more efficiently allocating limited resources. Policy makers need to start developing the infrastructures needed to design and deploy such strategies when testing capacity has constraints.
- Vaccination policies and active testing policies do work in conjunction with each other, and policy makers need an integrated plan that combines these two rather than viewing these as separate ideas for managing a pandemic.
- Explainability and fairness are both critical for public policy in general, and health in particular. The method presented here offers the ability to incorporate both of these objectives into an active testing framework. The MAB approach here can explicitly provide information on whether an individual is being sampled for exploration (random) or exploitation (with specific details that can be provided based on infections in that immediate contact network). Some alternate methods (such as deep reinforcement learning) do not have this property. Whereas we showed that the proposed method here demonstrates outcome-based fairness, we stressed that this could also be because of how race influenced disease spread in our own implementation. Instead, we emphasized that our method can incorporate procedural fairness constraints readily into an active testing policy.
- Compliance to testing, and compliance to quarantine are both important aspects of the active testing methodology, as well as of the benefits that can be realized from this approach. This can be an issue if users selected for testing refuse to do so on grounds that this was algorithmically generated. Here, we suggest three options. First, as noted previously, our method can provide explanations as to why a specific individual was chosen by the algorithm. Such explanations can also help manage any legal challenges that may arise. Second, policy makers will have to resort to incentive

design schemes that can help individuals to both agree to be tested when selected, as well as to stay in quarantine when they need do. These incentives will be expensive to implement but will likely be significantly less costly than entire lockdowns that have hurt the global economy significantly. There is potential for high-impact research along these lines that can also bring in the ethical aspects of doing so in the case of vaccinations (Jecker 2021). Third, policy makes can use public health announcements as a way to explain to citizens why participating in such active testing protocols can be beneficial to the society at large; such announcements are commonly used in public health but need to be carefully designed (Whitney and Viswanath 2004).

- The broader methodology used here can also serve as a framework for countries and policy makers to think about how to evaluate different pandemic mitigation strategies in the future. ABM models can be built specific to each location along with the policies (agents) and their strategies; this can help evaluate different vaccination/testing strategies before actual rollout of expensive policies. It can also help policy makers adopt a more rational response strategy to pandemic management than what might otherwise be viewed as an ad hoc response.
- Finally, we do note that there are privacy implications that have to be considered carefully. This is one where our work here does not provide specific guidance. There may be implementation-based ideas that can offer some degree of privacy protection that need to be explored. For example, the individual contact network can be stored with anonymized user IDs, with a separate lookup table that maps the anonymized IDs to specific individuals. From a technology perspective, it then becomes possible to implement role-based access to the database that can prevent the exact name lookups unless absolutely required. In most cases, the only time this needs to happen is when the individual needs to be identified for testing or contact tracing, and those restrictions can also be operationalized using existing database controls. We leave a more comprehensive treatment of privacy-protecting designs of the backend database to future work but feel it is essential to point this out here as an issue that needs to be explicitly recognized and addressed as well.

In summary, this paper makes the following contributions. We presented a novel active sampling algorithm to effectively manage pandemics such as COVID-19 by sampling individuals and not covariates. We also add to the MAB literature, as most of the existing work assumes that the complete network structure and underlying process is known beforehand. Our ideas here modeled the partially observed scenarios within this framework and showed how the contact network can be strategically expanded over time to support the active sampling strategy, even under a nonstationary environment. From a practical perspective, we showed that smart testing is still important even when vaccination is available, assuming that some portion of the individuals cannot or will not get the vaccine in a timely manner—this issue is particularly important given recent findings that significant percentages of the population are resisting vaccination.

Finally, we would like to point out that a more complete model of our problem can fall into the umbrella of reinforcement learning (RL), which is a strict generalization of contextual bandits. In our setting of mitigating COVID-19, one difficulty is that many contextual sequences are exogenous and are not controlled or influenced by the smart testing algorithm. As suggested by Chen et al. (2020) and Long et al. (2018), it is challenging to compute the optimal dynamic allocation using RL, even for 35 age-compartment pairs, and the heuristic obtained from RL performs worse than simple heuristics such as a single-step myopic policy, let alone individual sampling policies considered in our work. The other difficulty is the large/infinite space of the context, intensified by data scarcity and unclear disease evolution patterns. Along with the disease diffusion, the learning agent might not encounter the same context twice. The critical issue is then how to generalize across contexts and how to obtain reliable estimates of value functions (Sutton and Barto 2018, Lattimore and Szepesvári 2020). Recent development has focused on deep learning for powerful function approximation (Mnih et al. 2015, Silver et al. 2016, Wang et al. 2016), and yet they require strong domain knowledge and large amounts of data to be successful. For the policy maker, having specific explanations as to why specific individuals were chosen is also important to incorporate in any RL approach. We thus leave the design of full RL algorithms for this problem to future research.

Acknowledgments

The authors thank the senior editor, the associate editor, and the anonymous reviewers for taking the time and effort necessary to review the manuscript. The authors sincerely appreciate all the insightful comments and constructive suggestions.

Endnote

[1] See https://www.ucl.ac.uk/news/2021/jan/lockdown-compliance-improving-low-take-covid-tests-worrying.

References

Abbasi A, Dillon-Merrill R, Rao HR, Sheng O, Chen R (2021) Call for papers—Special issue of information systems research—Unleashing the power of information technology for strategic management of disasters. *Inform. Systems Res.* 32(4):1490–1493.

Ahmad MA, Patel A, Eckert C, Kumar V, Teredesai A (2020) Fairness in machine learning for healthcare. *Proc. 26th ACM SIGKDD Internat. Conf. Knowledge Discovery Data Mining* (ACM, New York), 3529–3530.

Alippi C, Roveri M (2008) Just-in-time adaptive classifiers—Part i: Detecting nonstationary changes. *IEEE Trans. Neural Networks* 19(7):1145–1153.

Allen H, Tessier E, Turner C, Anderson C, Blomquist P, Simons D, Lochen A, et al (2022) Comparative transmission of SARS-CoV-2 Omicron (B.1.1.529) and Delta (b.1.617.2) variants and the impact of vaccination: National cohort study, England. Preprint, submitted February 17, https://doi.org/10.1101/2022.02.15.22271001.

Alon N, Cesa-Bianchi N, Dekel O, Koren T (2015) Online learning with feedback graphs: Beyond bandits. *Proc. 28th Conf. Learning Theory*, PMLR 40:23–35.

Asnakew Z, Kerebih Asrese MA (2020) Community risk perception and compliance with preventive measures for COVID-19 pandemic in Ethiopia. *Risk Management Healthcare Policy* 13:2887–2897.

Atkeson A (2020) What will be the economic impact of COVID-19 in the US? Rough estimates of disease scenarios. NBER Working Paper No. 26867, National Bureau of Economic Research, Cambridge, MA.

Auer P, Cesa-Bianchi N, Fischer P (2002) Finite-time analysis of the multi-armed bandit problem. *Machine Learning* 47(2–3):235–256.

Baxter A, Oruc BE, Asplund J, Keskinocak P, Serban N (2022) Evaluating scenarios for school reopening under COVID19. *BMC Public Health* 22(1):1–10.

Beca-Martínez MT, Romay-Barja M, Falcón-Romero M, Rodríguez-Blázquez C, Benito-Llanes A, Forjaz MJ (2021) Compliance with the main preventive measures of covid-19 in spain: The role of knowledge, attitudes, practices, and risk perception. *Transboundary Emerging Diseases* 69(4):e871–e882.

Besbes O, Gur Y, Zeevi A (2014) Stochastic multi-armed-bandit problem with non-stationary rewards. *Adv. Neural Inform. Processing Systems* 27.

Bifet A, Pfahringer B, Read J, Holmes G (2013) Efficient data stream classification via probabilistic adaptive windows. *Proc. 28th Annual ACM Sympos. Appl. Comput.* (ACM, New York), 801–806.

Bodas M, Peleg K (2020) Self-isolation compliance in the COVID-19 era influenced by compensation: Findings from a recent survey in Israel: Public attitudes toward the COVID-19 outbreak and self-isolation: A cross sectional study of the adult population of Israel. *Health Affairs* 39(6):936–941.

Bubeck S, Cesa-Bianchi N (2012) Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations Trends Machine Learn.* 5(1):1–122.

Carpentier A, Valko M (2016) Revealing graph bandits for maximizing local influence. *Proc. 19th Internat. Conf. Artificial Intelligence Statistics*, PMLR 51:10–18.

Cervantes A, Gagné C, Isasi P, Parizeau M (2018) Evaluating and characterizing incremental learning from non-stationary data. Preprint, submitted June 18, https://doi.org/10.48550/arXiv.1806.06610.

Chang SL, Harding N, Zachreson C, Cliff OM, Prokopenko M (2020) Modelling transmission and control of the COVID-19 pandemic in Australia. *Nature Comm.* 11(1):5710.

Chapelle O, Li L (2011) An empirical evaluation of Thompson sampling. *Adv. Neural Inform. Processing Systems* 24:2249–2257.

Chen H, Huang Z, Li S, Zhang C (2021) Understanding bandits with graph feedback. *Adv. Neural Inform. Processing Systems* 34:24659–24669.

Chen X, Li M, Simchi-Levi D, Zhao T (2020) Allocation of COVID-19 vaccines under limited supply. Preprint, submitted August 31, https://dx.doi.org/10.2139/ssrn.3678986.

Cheung WC, Simchi-Levi D, Zhu R (2022) Hedging the drift: Learning to optimize under non stationarity. *Management Sci.* 68(3):1696–1713.

Chiotellis I, Cremers D (2020) Neural online graph exploration. Preprint, submitted December 6, https://doi.org/10.48550/arXiv.2012.03345.

Chu W, Li L, Reyzin L, Schapire RE (2011) Contextual bandits with linear payoff functions. *Proc. 14th Internat.. Conf. Artificial Intelligence Statistics*, PMLR 15:208–214.

Cohn DA, Ghahramani Z, Jordan MI (1996) Active learning with statistical models. *J. Artificial Intelligence Res.* 4(1):129–145.

Cooper I, Mondal A, Antonopoulos CG (2020) A SIR model assumption for the spread of COVID-19 in different communities. *Chaos Solitons Fractals* 139:110057.

Dai H, Li Y, Wang C, Singh R, Huang PS, Kohli P (2019) Learning transferable graph exploration. *Adv. Neural Inform. Processing Systems* 32.

Dawson P, Werkman M, Brooks-Pollock E, Tildesley M (2015) Epidemic predictions in an imperfect world: Modelling disease spread with partial data. *Proc. Royal Soc. B Biol. Sci.* 282:20150205.

De Cremer D, Brockner J, Fishman A, Van Dijke M, Van Olffen W, Mayer DM (2010) When do procedural fairness and outcome fairness interact to influence employees' work attitudes and behaviors? The moderating effect of uncertainty. *J. Appl. Psych.* 95(2):291.

Eames KT, Keeling MJ (2003) Contact tracing and disease control. *Proc. Royal Soc. B Biol. Sci.* 270(1533):2565–2571.

Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, Parker M, Bonsall D, Fraser C (2020) Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science* 368(6491):eabb6936.

Franzen A, Wöhner F (2021) Coronavirus risk perception and compliance with social distancing measures in a sample of young adults: Evidence from Switzerland. *PLoS One.* 16(2):e0247447.

Fylkesnes K, Ndhlovu Z, Kasumba K, Musonda RM, Sichone M (1998) Studying dynamics of the HIV epidemic: Population-based data compared with sentinel surveillance in Zambia. *AIDS* 12(10):1227–1242.

Garivier A, Cappé O (2011) The KL-UCB algorithm for bounded stochastic bandits and beyond. *Proc. 24th Annual Conf. Learning Theory* (PMLR, New York) 19:359–376.

Garivier A, Moulines E (2011) On upper-confidence bound policies for switching bandit problems. *Internat. Conf. Algorithmic Learning Theory* (Springer, Berlin), 174–188.

Goel S, Salganik MJ (2010) Assessing respondent-driven sampling. *Proc. Natl. Acad. Sci. USA* 107(15):6743–6747.

Grushka-Cohen H, Cohen R, Shapira B, Moran-Gilad J, Rokach L (2020) A framework for optimizing covid-19 testing policy using a multi armed bandit approach. Preprint, submitted July 28, https://doi.org/10.48550/arXiv.2007.14805.

Gu Q, Han J (2014) Online spectral learning on a graph with bandit feedback. *Proc. 2014 IEEE Internat. Conf. Data Mining* (IEEE, Piscataway, NJ), 833–838.

Hellewell J, Abbott S, Gimma A, Bosse NI, Jarvis CI, Russell TW, Munday JD, et al. (2020) Feasibility of controlling covid-19 outbreaks by isolation of cases and contacts. *Lancet Global Health* 8(4):E488–E496.

Henderson K, Gallagher B, Eliassi-Rad T, Tong H, Basu S, Akoglu L, Koutra D, Faloutsos C, Li L (2012) RolX: Structural role extraction & mining in large graphs. *Proc. 18th ACM SIGKDD Internat. Conf. Knowledge Discovery Data Mining* (ACM, New York), 1231–1239.

Hethcote HW (2000) The mathematics of infectious diseases. *SIAM Rev.* 42(4):599–653.

Hoertel N, Blachier M, Blanco C, Olfson M, Massetti M, Limosin F, Leleu H (2020) Facing the COVID-19 epidemic in NYC: A stochastic agent-based model of various intervention strategies. Preprint, submitted April 28, https://doi.org/10.1101/2020.04.23.20076885.

Huerta R, Tsimring LS (2002) Contact tracing and epidemics control in social networks. *Physical Rev. E* 66:056115.

Jecker NS (2021) Cash incentives, ethics, and COVID-19 vaccination. *Science* 374(6569):819–820.

Jiang H (2019) Non-asymptotic uniform rates of consistency for k-NN regression. *Proc. AAAI Conf. Artificial Intelligence* (AAAI Press, Palo Alto, CA), 3999–4006.

Kamarthi H, Vijayan P, Wilder B, Ravindran B, Tambe M (2020) Influence maximization in unknown social networks: Learning policies for effective graph sampling. *Proc. 19th Internat. Conf. Autonomous Agents and MultiAgent Systems*, 575–583.

Kandasamy K, Krishnamurthy A, Schneider J, Póczos B (2018) Parallelised Bayesian optimisation via Thompson sampling. *Proc. 21st Internat. Conf. Artificial Intelligence Statistics* (PMLR, New York), 133–142.

Koo JR, Cook AR, Park M, Sun Y, Sun H, Lim JT, Tam C, Dickens BL (2020) Interventions to mitigate early spread of SAR-CoV-2 in Singapore: A modelling study. *Lancet Infectious Diseases* 20(6):678–688.

Lattimore T, Szepesvári C (2020) *Bandit Algorithms* (Cambridge University Press, Cambridge, UK).

Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, Azman AS, Reich NG, Lessler J (2020) The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: Estimation and application. *Ann. Internal Medicine* 172(9):577–582.

Lee LM, Thacker SB, Louis MES (2010) *Principles and Practice of Public Health Surveillance* (Oxford University Press, New York).

Li L, Chu W, Langford J, Schapire RE (2010) A contextual-bandit approach to personalized news article recommendation. *Proc. 19th Internat. Conf. World Wide Web* (ACM, New York), 661–670.

Long EF, Nohdurft E, Spinler S (2018) Spatial resource allocation for emerging epidemics: A comparison of greedy, myopic, and dynamic policies. *Manufacturing Service Oper. Management* 20(2):181–198.

Ma Y, Huang TK, Schneider JG (2015) Active search and bandits on graphs using sigma-optimality. *Proc. 31st Conf. Uncertainty Artificial Intelligence* (AUAI, Amsterdam), 542–551.

Macready WG, Wolpert DH (1998) Bandit problems and the exploration/exploitation tradeoff. *IEEE Trans. Evolutionary Comput.* 2(1):2–22.

Madewell ZJ, Yang Y, Longini IM, Halloran ME, Dean NE (2020) Household transmission of SARS-CoV-2: A systematic review and meta-analysis. *JAMA Network Open* 3(12):e2031756.

Madhawa K, Murata T (2019) A multi-armed bandit approach for exploring partially observed networks. *Appl. Network Sci.* 4(1):26.

Magnani R, Sabin K, Saidel T, Heckathorn D (2005) Review of sampling hard-to-reach and hidden populations for HIV surveillance. *AIDS* 19:S67–S72.

Mallapaty S (2022) COVID-19: How Omicron overtook Delta in three charts. *Nature* 10.

Manchein C, Brugnago EL, da Silva RM, Mendes CF, Beims MW (2020) Strong correlations between power-law growth of covid-19 in four continents and the inefficiency of soft quarantine strategies. *Chaos* 30(4):041102.

Mannor S, Shamir O (2011) From bandits to experts: On the value of side-observations. *Adv. Neural Inform. Processing Systems* 24:684–692.

Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J (2013) Distributed representations of words and phrases and their compositionality. *Adv. Neural Inform. Processing Systems* 26:3111–3119.

Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, Graves A, Riedmiller M, Fidjeland AK, Ostrovski G, et al. (2015) Human-level control through deep reinforcement learning. *Nature* 518(7540):529–533.

Murray CJ (2022) COVID-19 will continue but the end of the pandemic is near. *Lancet* 399(10323):417–419.

Nemhauser GL, Wolsey LA, Fisher ML (1978) An analysis of approximations for maximizing submodular set functions—I. *Math. Programming* 14(1):265–294.

Ning L, Niu J, Bi X, Yang C, Liu Z, Wu Q, Ning N, et al. (2020) The impacts of knowledge, risk perception, emotion and information on citizens' protective behaviors during the outbreak of COVID-19: A cross-sectional study in China. *BMC Public Health* 20(1):1–12.

Perez L, Dragicevic S (2009) An agent-based approach for modeling dynamics of contagious disease spread. *Internat. J. Health Geographics* 8(1):50.

Perozzi B, Al-Rfou R, Skiena S (2014) Deepwalk: Online learning of social representations. *Proc. 20th ACM SIGKDD Internat. Conf. Knowledge Discovery Data Mining* (ACM, New York), 701–710.

Prem K, Liu Y, Russell TW, Kucharski AJ, Eggo RM, Davies N, Flasche S, et al. (2020) The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: A modelling study. *Lancet Public Health* 5(5):e261–e270.

Provodin D, Gajane P, Pechenizkiy M, Kaptein M (2022) The impact of batch learning in stochastic linear bandits. Preprint, submitted February 14, https://doi.org/10.48550/arXiv.2202.06657.

Raj V, Kalyani S (2017) Taming non-stationary bandits: A Bayesian approach. Preprint, submitted July 31, https://doi.org/10.48550/arXiv.1707.09727.

Roda WC, Varughese MB, Han D, Li MY (2020) Why is it difficult to accurately predict the COVID-19 epidemic? *Infectious Disease Modelling* 5:271–281.

Saar-Tsechansky M, Provost FJ (2007) Decision-centric active learning of binary-outcome models. *Inform. Systems Res.* 18(1):4–22.

Salathé M, Althaus CL, Neher R, Stringhini S, Hodcroft E, Fellay J, Zwahlen M, et al. (2020) COVID-19 epidemic in Switzerland: On the importance of testing, contact tracing and isolation. *Swiss Medical Weekly* 150(11–12):w20225.

Silva PC, Batista PV, Lima HS, Alves MA, Guimarães FG, Silva RC (2020) COVID-ABS: An agent-based model of COVID-19 epidemic to simulate health and economic effects of social distancing interventions. *Chaos Solitons Fractals* 139:110088.

Silver D, Huang A, Maddison CJ, Guez A, Sifre L, Van Den Driessche G, Schrittwieser J, Antonoglou I, Panneershelvam V, Lanctot M, et al. (2016) Mastering the game of go with deep neural networks and tree search. *Nature* 529(7587):484–489.

Singh R, Liu F, Liu X, Shroff N (2020) Contextual bandits with side-observations. Preprint, submitted June 6, https://doi.org/10.48550/arXiv.2006.03951.

Singla A, Horvitz E, Kohli P, White R, Krause A (2015) Information gathering in networks via active exploration. *Proc. 24th Internat. Conf. Artificial Intelligence* (AAAI Press, Palo Alto, CA), 981–988.

Soundarajan S, Eliassi-Rad T, Gallagher B, Pinar A (2017) ε-WGX: Adaptive edge probing for enhancing incomplete networks. *Proc. 2017 ACM on Web Science Conf.* (ACM, New York), 161–170.

Sutton RS, Barto AG (2018) *Reinforcement Learning: An Introduction* (MIT Press, Cambridge, MA).

Thompson WR (1933) On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika* 25(3–4):285–294.

Topirceanu A, Udrescu M, Marculescu R (2020) Centralized and decentralized isolation strategies and their impact on the COVID-19 pandemic dynamics. Preprint, submitted April 8, https://doi.org/10.48550/arXiv.2004.04222.

Trovo F, Paladino S, Restelli M, Gatti N (2020) Sliding-window Thompson sampling for non-stationary settings. *J. Artificial Intelligence Res.* 68:311–364.

Ulloa AC, Buchan SA, Daneman N, Brown KA (2022) Estimates of SARS-CoV-2 Omicron variant severity in Ontario, Canada. *JAMA* 327(13):1286–1288.

Underwood BR, White VLC, Baker T, Law M, Moore-Gillon JC (2003) Contact tracing and population screening for tuberculosis – Who should be assessed? *J. Public Health* 25(1):59–61.

Wang Z, Schaul T, Hessel M, Hasselt H, Lanctot M, Freitas N (2016) Dueling network architectures for deep reinforcement

learning. *Proc. 33rd Internat. Conf. Machine Learning*, PMLR 48:1995–2003.

Whitney R, Viswanath K (2004) Lessons learned from public health mass media campaigns: Marketing health in a crowded media world. *Annual Rev. Public Health* 25:419.

World Meter (2020) COVID-19 coronavirus pandemic. Retrieved August 31 from https://www.worldometers.info/coronavirus/.

Zheng Z, Padmanabhan B (2006) Selectively acquiring customer information: A new data acquisition problem and an active learning-based solution. *Management Sci.* 52(5):697–712.