SCIENTIFIC REPORTS

OPEN

# A decision-support framework to optimize border control for global outbreak mitigation

Received: 21 August 2018
Accepted: 28 December 2018
Published online: 18 February 2019

Aleksa Zlojutro[1], David Rey[1] & Lauren Gardner[1,2]

**The introduction and spread of emerging infectious diseases is increasing in both prevalence and scale. Whether naturally, accidentally or maliciously introduced, the substantial uncertainty surrounding the emergence of novel viruses, specifically where they may come from and how they will spread, demands robust and quantifiably validated outbreak control policies that can be implemented in real time. This work presents a novel mathematical modeling framework that integrates both outbreak dynamics and outbreak control into a decision support tool for mitigating infectious disease pandemics that spread through passenger air travel. An ensemble of border control strategies that exploit properties of the air traffic network structure and expected outbreak behavior are proposed. A stochastic metapopulation epidemic model is developed to evaluate and rank the control strategies based on their effectiveness in reducing the spread of outbreaks. Sensitivity analyses are conducted to illustrate the robustness of the proposed control strategies across a range of outbreak scenarios, and a case study is presented for the 2009 H1N1 influenza pandemic. This study highlights the importance of strategically allocating outbreak control resources, and the results can be used to identify the most robust border control policy that can be implemented in the early stages of an outbreak.**

The potential harm posed by the introduction and spread of emerging infectious diseases has been recently illustrated by the 2009 H1N1[1], SARS[2], and Zika[3] epidemics. There is considerable evidence that such pandemics are likely to become more frequent unless action is taken to mitigate their spread at a global scale[4–7]. For this reason, resilience management for outbreaks has attracted a growing body of literature, both from epidemic modelers and the optimization and control community. One of the most critical aspects of resilience management is the need to combine accurate epidemic growth models with detailed outbreak control strategies[8], representing a gap in the literature which this study aims to fill.

The availability of large scale data and growing computational capabilities has significantly advanced infectious disease spread models in recent years[9–12]. The community has notably explored the impact of network topology, epidemic thresholds and diffusion models on outbreak spread patterns[13–18]. A range of epidemic models have been developed, which increase in complexity from single-population, deterministic models to metapopulation, stochastic models[6,19–21]. Deterministic models provide efficient mathematical representation but lack the realism of stochastic simulation models[20,22–24]. On the other hand, detailed computational and visualization tools such as GLEaM[25] and STEM[26], among others[27–30], have emerged as powerful solutions to model the spread of infectious diseases with a high level of accuracy and even measure the impact of control strategies[31], albeit at a high computational cost. As an alternative to simulation-based models, analytical global epidemic models have also been developed to characterize the stochastic spread of infectious diseases in metapopulation networks[5,32,33].

As highlighted by[4], the intensive restriction of human activities in metapopulation networks may undermine the system's functionality and lead to significant societal costs. Hence, in efforts to mitigate large-scale pandemics, it is critical to cautiously deploy control strategies to minimize disruptions and maximize the reactiveness of the system[34–37]. At a global scale, passenger air travel is known to play a critical role in the spread of infectious disease[16,27,38–41]. Additionally, border control has been shown to play a pivotal role in mitigating epidemics, especially during the emerging stage of outbreaks[5,42–45]. Border control is typically deployed at airports in an attempt to prevent the spread of an infectious disease between cities, states and countries through passenger air travel[46].

[1]School of Civil and Environmental Engineering, University of New South Wales (UNSW) Sydney, Sydney, NSW, 2052, Australia. [2]Department of Civil Engineering, Johns Hopkins University, Baltimore, MD, 21218, USA. Correspondence and requests for materials should be addressed to L.G. (email: l.gardner@unsw.edu.au)

However, identifying the optimal set of airports for deploying border control is challenging, especially when only limited control resources are available, *e.g.*, budget constraints.

In this work we present a novel mathematical modeling framework that integrates both outbreak dynamics and outbreak control into a decision-support tool for mitigating infectious disease pandemics at the onset of an outbreak through border control. The border control mechanism considered in this work is passenger screening upon arrival at airports (entry screening), which is used to identify infected or at-risk individuals and provide immediate treatment and isolation to reduce the risk of further transmission[47]. The uncertainty of exit screening effectiveness in other countries combined with the possible development of symptoms during a flight[48] has prompted several governments to deem entry screening as crucial to the protection of their countries[49], and further motivates its use in this study. The proposed model seeks to determine the optimal set of airports that should be allocated screening resources (technology and personnel), and the corresponding amount (thus dictating the proportion of arriving passengers that can be screened) such that the outbreak risk is minimized. We propose an ensemble of control strategies that exploit the heterogeneity of the air traffic network structure and outputs from an outbreak simulation model in the allocation of control resources. We evaluate each control strategy using a stochastic metapopulation epidemic model, and compare the strategies based on their effectiveness in reducing the spread of outbreaks. Note that we are not proposing or evaluating air travel restrictions, which have substantial economic costs as well as recognized limitations in their ability to prevent or reduce the scale of pandemics[24,33,50–54]. The goal of the proposed decision-support framework is to optimize the use of airport screening for border control to minimize the harm, *i.e.*, number of infected persons and cities, posed by the introduction of a new disease into susceptible cities, thus providing local public health authorities more time to plan, prepare and distribute local control strategies, *e.g.*, anti-virals, vaccines, source exit screening etc., which must be rapidly administered if/when infection is introduced.

This study builds upon previous work[55,56] that presented a mathematical modeling framework to integrate control and outbreak dynamics. We extend this line of work through the following substantial contributions: (1) we model outbreak dynamics using a stochastic metapopulation framework as opposed to a deterministic model, (2) we propose and evaluate a novel set of control strategies, (3) the control strategies are embedded within a resource constrained decision-support framework, (4) the model is calibrated using historical outbreak data, and (5) a case study is presented for the 2009 H1N1 outbreak. Additionally, a range of sensitivity analysis is conducted to illustrate the robustness of the model with regards to variability across outbreak scenarios, disease parameters, policy considerations, and modelling assumptions. The analysis elucidates key trade-offs in terms of budget availability and outbreak mitigation which can be leveraged to inform on public health policy for global epidemic preparedness and control.

## Materials and Methods

**Mathematical Model.** In this section, we present both the stochastic metapopulation epidemic model used to simulate outbreak dynamics, and its integration within the proposed decision-support framework. Our metapopulation model is based on a global air travel network which connects local, city-level, populations. Formally, the proposed metapopulation network can be represented by a graph $G=(V, E)$ where $V$ is the set of nodes and $E$ is the set of directed edges in the network. Nodes represent cities and edges represent passenger travel routes, possibly including stopovers, among cities. At each node of the network, we locally model outbreak dynamics using a discrete-time Susceptible-Exposed-Infected-Recovered (SEIR) compartmental model[57]. The time steps are set to be $t \in T=\{1, 2, \ldots, t_{obs}\}$ where $t_{obs}$ is the time step where the state of the outbreak is being evaluated. Local and global outbreak dynamics models are coupled by indexing compartmental states by network nodes $i \in V$ and time steps $t$. Specifically, we denote $S_{i,t}$, $E_{i,t}$, $I_{i,t}$ and $R_{i,t}$ the susceptible, exposed, infectious and recovered compartments at node $i$ at time $t$. Because our focus is on the early stages of an outbreak (*e.g.*, weeks or months), we assume that nodes have time-independent populations and we denote $N_i$ the population at node $i \in V$. We use this metapopulation model to capture day-to-day global travel dynamics, wherein the time steps are assumed to be of the order of magnitude of a day in length, which is consistent with other studies that simulate infectious diseases dynamics at a global scale[31,35,36,43].

We use a multi-commodity network flow model with time-dependent edge flows to model passenger movements from their origin node to their destination node. Let $\Pi_{ij}$ be the set of paths from $i \in V$ to $j \in V$. We denote $f^k_{ij,t}$ the average passenger flow on the route from $i \in V$ to $j \in V$ using path $k \in \Pi_{ij}$ at time step $t$ and we assume symmetric passenger flows for all pairs of origins and destinations, i.e. $f^k_{ij,t} = f^k_{ji,t}$. We denote $\Gamma^-_i = \{j \in V\colon \exists\ k \in \Pi_{ij},\ t \in T, f^k_{ji,t} > 0\}$ and $\Gamma^+_i = \{j \in V\colon \exists\ k \in \Pi_{ij},\ t \in T, f^k_{ij,t} > 0\}$ the sets of nodes connected to and from node $i \in V$, respectively. Each path $k \in \Pi_{ij}$ is an ordered sequence of nodes starting at $i$ and ending at $j$, i.e. $k=\{i, n_1, n_2, \ldots, j\}$. The path-based formulation, while more complex, enables more effective control decisions to be identified by the model. Specifically, the model is able to accurately capture the effect of controlling at stopover airports along a route, as well as identify the most cost-effective control decisions which utilize information about the entire path. For example, consider a group of travelers departing a high-risk infected region, *e.g.*, Sierra Leon during an Ebola outbreak, destined for ten different destination cities; and all ten routes include an initial stopover at the same airport, *e.g.*, JFK. By accounting for the complete path information, the model is able it identify the critical role played by JFK in this scenario, which further enables it to be optimally selected for control. Without accounting for the full travel paths of the routes, the critical role of JFK would be unidentifiable by the model.

The governing infection dynamics of the SEIR model[57] are used to model local outbreak dynamics in each city. For the purposes of this work the contact rate is assumed to be constant across populations. We denote $\beta_i$ the (local) contact rate at node $i$, $\gamma$ the transition or recovery rate and $\alpha$ the exposed parameter. In addition, we define $\lambda \in [0, 1]$ the likelihood to travel when infectious, with $\lambda = 1$ representing the case where infected and healthy

individuals are equally likely to travel. This parameter aims to represent the impact of reduced travel demand when infectious individuals are unable to travel due to severe symptoms. Finally, we assume that compartmental edge flows are proportional to tail node states, i.e. the number of travelers in a state is proportional to the number of individuals in this state at the origin node. Discrete-time stochastic metapopulation outbreak dynamics are summarized in Eq. (1) below.

$$S_{i,t+1} = S_{i,t} - \frac{\beta_i I_{i,t} S_{i,t}}{N_i} + \sum_{j\in\Gamma_i^-}\sum_{k\in\Pi_{ji}} S_{ji,t}^k - \sum_{j\in\Gamma_i^+}\sum_{k\in\Pi_{ij}} S_{ij,t}^k \tag{1a}$$

$$E_{i,t+1} = E_{i,t} + \frac{\beta_i I_{i,t} S_{i,t}}{N_i} - \alpha E_{i,t} + \sum_{j\in\Gamma_i^-}\sum_{k\in\Pi_{ji}} E_{ji,t}^k - \sum_{j\in\Gamma_i^+}\sum_{k\in\Pi_{ij}} E_{ij,t}^k \tag{1b}$$

$$I_{i,t+1} = I_{i,t} - \gamma I_{i,t} + \alpha E_{i,t} + \sum_{j\in\Gamma_i^-}\sum_{k\in\Pi_{ji}} I_{ji,t}^k - \lambda \sum_{j\in\Gamma_i^+}\sum_{k\in\Pi_{ij}} I_{ij,t}^k \tag{1c}$$

$$R_{i,t+1} = R_{i,t} + \gamma I_{i,t} + \sum_{j\in\Gamma_i^-}\sum_{k\in\Pi_{ji}} R_{ji,t}^k - \sum_{j\in\Gamma_i^+}\sum_{k\in\Pi_{ij}} R_{ij,t}^k \tag{1d}$$

The symbols $S_{ij,t}^k$, $E_{ij,t}^k$, $I_{ij,t}^k$ and $R_{ij,t}^k$ represent compartmental edge flows on $(i, j)$ with destination $k$ at time step $t$. For compartments $S$ and $R$, compartmental edge flows are assumed deterministic and equal to their expected values, i.e.: $S_{ij,t}^k = f_{ij,t}^k \frac{S_{i,t}}{N_i}$ and $R_{ij,t}^k = f_{ij,t}^k \frac{R_{i,t}}{N_i}$. However, since the compartmental edge flows of exposed and infectious passengers may be considerably smaller than that of other compartments, we model $E_{ij,t}^k$ and $I_{ij,t}^k$ as discrete random variables, as the stochastic allocation of infected individuals to destinations is critical when modeling the early stages of an outbreak. Specifically, we define these compartmental edge flows as follows:

$$E_{ij,t}^k = \left\lfloor f_{ij,t}^k \frac{E_{i,t}}{N_i} \right\rfloor + \widetilde{E}_{ij,t}^k \tag{2a}$$

$$I_{ij,t}^k = \left\lfloor f_{ij,t}^k \frac{I_{i,t}}{N_i} \right\rfloor + \widetilde{I}_{ij,t}^k \tag{2b}$$

where $\widetilde{E}_{ij,t}^k$ and $\widetilde{I}_{ij,t}^k$ are discrete random variables representing the number of exposed and infectious passengers, respectively, beyond the integer-part of their respective compartmental edge flows. Further, let $m = |\Gamma_i^+|$ be the number of destination nodes from node $i \in V$, the vector $\widetilde{\boldsymbol{E}}_{i,t}^k = (\widetilde{E}_{i1,t}^k, \ldots, \widetilde{E}_{im,t}^k)$ (resp. $\widetilde{\boldsymbol{I}}_{i,t}^k = (\widetilde{I}_{i1,t}^k, \ldots, \widetilde{I}_{im,t}^k)$) follows a multinomial distribution with a number of trials $n_{i,t}^{k,E} = \left\lfloor f_{ij,t}^k \frac{E_{i,t}}{N_i} - \left\lfloor f_{ij,t}^k \frac{E_{i,t}}{N_i} \right\rfloor \right\rfloor$ (resp. $n_{i,t}^{k,I} = \left\lfloor f_{ij,t}^k \frac{I_{i,t}}{N_i} - \left\lfloor f_{ij,t}^k \frac{I_{i,t}}{N_i} \right\rfloor \right\rfloor$) and probability vector $\boldsymbol{p}_{i,t}^k = \left( \frac{f_{i1,t}^k}{\sum_{j\in\Gamma_i^+} f_{ij,t}^k}, \ldots, \frac{f_{im,t}^k}{\sum_{j\in\Gamma_i^+} f_{ij,t}^k} \right)$. This implementation enables a more computationally efficient global epidemic simulation tool compared with one fully simulating all epidemic compartments as random variables following multinomial distributions, while still capturing the critical and inherent uncertainty of the destination of the first infected travelers.

This stochastic formulation aims to model integer, compartmental edge flows and prevent the movement of fractional exposed or infectious individuals which may result in unrealistic epidemic behavior at a global scale[58]. Consequently, the compartmental states $S_{i,t}$, $E_{i,t}$, $I_{i,t}$ and $R_{i,t}$ are also random variables representative of the evolution of the outbreak over time and space.

To integrate control decisions within the above stochastic metapopulation network we model passenger screening upon arrival at airports as a control variable. Passenger screening can be done through visual inspections of passengers, health declaration cards and/or infrared thermal image scanners[59]. In this work the specific type of screening is not the focus; as the framework is applicable to multiple control mechanisms. We propose to use airport screening rates as the main control variables, which are representative of the proportion of arriving passengers successfully screened at a given airport. We denote $x_{i,t} \in [0, 1]$ the control rate at node $i$ at time step $t$. Control variables can be incorporated in the proposed metapopulation epidemic model by re-defining Eqs (1c) and (1d) as follows:

$$I_{i,t+1} = I_{i,t} - \gamma I_{i,t} + \alpha E_{i,t} + \sum_{j\in\Gamma_i^-}\sum_{k\in\Pi_{ji}} I_{ji,t}^k \left( \prod_{p\in k\setminus\{j\}} (1 - x_{p,t}) \right) - \lambda \sum_{j\in\Gamma_i^+}\sum_{k\in\Pi_{ij}} I_{ij,t}^k \tag{3c}$$

$$R_{i,t+1} = R_{i,t} + \gamma I_{i,t} + \sum_{j\in\Gamma_i^-}\sum_{k\in\Pi_{ji}} \left( R_{ji,t}^k + I_{ji,t}^k \left( 1 - \prod_{p\in k\setminus\{j\}} (1 - x_{p,t}) \right) \right) - \sum_{j\in\Gamma_i^+}\sum_{k\in\Pi_{ij}} R_{ij,t}^k \tag{3d}$$

This formulation is able to capture the combined effects of screening passengers at multiple nodes along their travel route. The combination of Eqs (1a), (1b), (3c) and (3d), hereby to as (3), can be viewed as a control-driven stochastic metapopulation epidemic model wherein variables $x_{i,t}$ represent the level of control over time space in the network. An illustration of the combined effect of node controls along a path with stop-overs is provided in the Supplementary Material (Section A). A control rate of less than one can be interpreted as shortcomings of the methods or technology involved with screening. We assume that passengers coming into a controlled airport who are successfully identified as infected individuals are isolated for treatment, and hence are no longer able to spread infection. In terms of the model, infected individuals screened at a controlled node are assumed to transition to the recovered state $R$. The main objective function is to minimize the expected cumulative number of infected individuals at the observation time[31]. This can be stated as follows:

$$\min \mathbb{E}\left[\sum_{i\in V} E_{i,t_{obs}} + I_{i,t_{obs}} + R_{i,t_{obs}}\right] \tag{4}$$

The challenge in policy decision making results because of a constraint on available resources. To address this challenge we introduce a budget and cost for control. The cost of using outbreak control resources is modeled using a generic cost function which consists of a fixed cost and a variable cost. The fixed cost represents the setup costs associated with the deployment of control resources, e.g., installation of new screening technologies and training of personnel, whereas the variable cost depends on the level of control deployed.

Formally, to model the activation of control at a node, we introduce binary variables $y_i \in \{0, 1\}$ for each node $i \in V$ which take value 1 if node $i$ is allocated a non-zero amount of control resources and 0 otherwise. Let $s_i$ be the setup cost associated to node $i \in V$, the fixed cost of the cost function is modeled as $y_i s_i$, i.e. if control resources are deployed at $i$ then a cost of $s_i$ monetary units is incurred. To ensure that variable $y_i$ is correctly adjusted, we introduce linking inequalities $x_{i,t} \leq y_i$ for each node $i \in V$ and for each time period $t \in T$ imposing that variable $y_i$ must take value 1 if control resources are deployed at node $i$ at any time $t$ during the control period (recall that $x_{i,t}$ represents the level of control at node $i$ at time $t$ and belongs to the interval [0,1]).

To model the variable cost of the cost function, we only require a generic real function $g(\boldsymbol{x}_i)$: $\mathbb{R}^{|V|} \rightarrow \mathbb{R}$ that represents the monetary costs associated with the local control vector $\boldsymbol{x}_i \in [0, 1]^{|T|}$at node $i \in V$ over the control period. Although no assumption is made on the shape of function $g(\boldsymbol{x}_i)$ is it reasonable to assume that it is non-decreasing with regards to vector components $x_{i,t}$. Further, in our numerical experiments we will assume that both $s_i$ and $g(\boldsymbol{x}_i)$ are functions of the total average incoming edge flow to $i$ at each time step $t$, i.e. $\sum_{j\in\Gamma_i^-}\sum_{k\in\Pi_{ji}} f_{ji}^k$, where $f_{ji}^k$ is the average flow on path $k$ over the control period. Let $c_i(y_i, \boldsymbol{x}_i)$ be the cost function at node $i \in V$, we assume the following generic form for this cost function:

$$c_i(y_i, \boldsymbol{x}_i) = y_i s_i + g(\boldsymbol{x}_i) \tag{5}$$

Finally, we assume that a budget $B$ is available for deploying control resources which translates into the budget constraint:

$$\sum_{i\in V} c_i(y_i, \boldsymbol{x}_i) \leq B \tag{6}$$

Our objective is to optimize the control vector $\boldsymbol{x} \in [0, 1]^{|V||T|}$ such that the impact of the outbreak at $t_{obs}$ as represented by (4) is minimized subject to control resource constraints and outbreak dynamics as governed by the stochastic metapopulation epidemic model (3) wherein compartmental edge flows $E_{ij,t}$ and $I_{ij,t}$, as well as compartmental states $S_{i,t}$, $E_{i,t}$, $I_{i,t}$ and $R_{i,t}$ are discrete random variables. This optimization formulation is summarized in Eq. (7) below.

$$\min \mathbb{E}\left[\sum_{i\in V} E_{i,t_{obs}} + I_{i,t_{obs}} + R_{i,t_{obs}}\right] \tag{7a}$$

Subject to:

$$\sum_{i\in V} y_i s_i + g(\boldsymbol{x}_i) \leq B \tag{7b}$$

$$x_{i,t} \leq y_i \quad \forall\, i \in V,\, t \in T \tag{7c}$$

$$S_{i,t+1} = S_{i,t} - \frac{\beta_i I_{i,t} S_{i,t}}{N_i} + \sum_{j\in\Gamma_i^-}\sum_{k\in\Pi_{ji}} S_{ji,t}^k - \sum_{j\in\Gamma_i^+}\sum_{k\in\Pi_{ij}} S_{ij,t}^k \quad \forall i \in V,\ t \in T \tag{7d}$$

$$E_{i,t+1} = E_{i,t} + \frac{\beta_i I_{i,t} S_{i,t}}{N_i} - \alpha E_{i,t} + \sum_{j\in\Gamma_i^-}\sum_{k\in\Pi_{ji}} E_{ji,t}^k - \sum_{j\in\Gamma_i^+}\sum_{k\in\Pi_{ij}} E_{ij,t}^k \quad \forall i \in V,\ t \in T \tag{7e}$$

| Strategy (Abbreviation) | Method of ranking airports |
|---|---|
| Baseline (B) | No control |
| Largest Population (LP) | Airports are ranked in descending order based on the population of the city they serve. When one city is serviced by multiple airports, those airports are further ranked in descending order based on travel volume. |
| Most Travelled (MT) | Airports are ranked in descending order based on travel volume. Travel volume is defined as the sum of daily incoming and outgoing flows. |
| Most Connected (MC) | Airports are ranked in descending order based on total incoming flow from source node(s). |
| Effective Path (EP) | Airports are ranked in ascending order based on their minimum effective path distance from the source(s). For a given outbreak scenario, the effective path distance for each airport is defined as the length of the shortest path from the source, where the length of a path is the sum of the effective distance of each link included in the path. The effective distance of link $(i, j)$ is defined as $d_{ij} = 1 - \log\left(\frac{\sum_{k\in V} f_{ji}^k}{\sum_{k\in V}\sum_{m\in\Gamma_i^-} f_{mi}^k}\right)$[16] where $f_{ji}^k$ is the average flow of path $k$ over the control period. When there is more than one source, the minimum effective path distance over all sources is assigned to each airport. |
| First Case (1C) | For a given outbreak scenario the baseline case is simulated a fixed number of times (*i.e.*, 200), and for each run we record the time at which the first infected individual reached each airport, i.e., the initial infection time. Airports are then ranked in ascending order based on their most frequently observed initial infection time. Ties are broken based on the MT criterion. |
| First Order Uniform (1OU) | For a given outbreak scenario the effect of fully controlling each airport independently, i.e. $x_i = 1$ for a selected node $i \in V$ and $x_i = 0$ otherwise, is simulated, and the reduction in cumulative number of infected individuals at the observation time $t_{obs}$ is compared to the baseline case (no control). Airports are then ranked in descending order based on the relative reduction in cases, *i.e.*, first-order effects. |

**Table 1.** List of control strategies and their description.

$$I_{i,t+1} = I_{i,t} - \gamma I_{i,t} + \alpha E_{i,t} + \sum_{j\in\Gamma_i^-}\sum_{k\in\Pi_{ji}} I_{ji,t}^k\left(\prod_{p\in k\setminus\{j\}}(1 - x_{p,t})\right) - \lambda\sum_{j\in\Gamma_i^+}\sum_{k\in\Pi_{ij}} I_{ij,t}^k \quad \forall i \in V,\ t \in T \tag{7f}$$

$$R_{i,t+1} = R_{i,t} + \gamma I_{i,t} + \sum_{j\in\Gamma_i^-}\sum_{k\in\Pi_{ji}}\left(R_{ji,t}^k + I_{ji,t}^k\left(1 - \prod_{p\in k\setminus\{j\}}(1 - x_{p,t})\right)\right) - \sum_{j\in\Gamma_i^+}\sum_{k\in\Pi_{ij}} R_{ij,t}^k \quad \forall i \in V,\ t \in T \tag{7g}$$

$$x_{i,t} \in [0, 1] \quad \forall i \in V,\ t \in T \tag{7h}$$

$$y_i \in \{0, 1\} \quad \forall i \in V \tag{7i}$$

**Proposed Control Strategies.** The final outbreak dynamics model with control decisions incorporated can be used as a tool to solve the resource allocation problem, and evaluate various control strategies. We propose a set of control strategies to optimize the control vector $\boldsymbol{x}$ subject to resource constraints. In this work, each proposed control strategy relies on a different metric to rank airports, and this ranking is then used to allocate control resources to a select set of airports. The objective of all strategies is to minimize the impact of the outbreak at a pre-selected future date we call the observation time, $t_{obs}$, at which impact is measured both in terms of total cumulative cases and number of infected cities.

Although the proposed model can accommodate dynamic control strategies, for the purposes of this work we focus on static control strategies, wherein nodes are controlled at the same level throughout the period of observation. Hence, we define and use the following static-equivalent cost function $g_s(x_i)$: $\mathbb{R} \rightarrow \mathbb{R}$ which we assume to be invertible over its domain. Our approach is based on a greedy resource allocation algorithm which operates as follows: the algorithm works from a set of sorted nodes $\overline{V}$ and iterates over the nodes in $\overline{V}$. The sorted node set is assumed to be pre-processed by one of the sorting criteria discussed in Table 1. At each iteration, the algorithm picks the next node $i$ of the sorted set and determines whether this node can be fully controlled ($x_i = 1$) without exceeding the available budget $B$. Variable $U$ represents the consumed budget. If the node can be fully controlled, the consumed budget is incremented by $s_i + g_s(1)$ which corresponds to the cost of control for $x_i = 1$. Otherwise, the algorithm checks whether node $i$ can be partially controlled: this is possible only if there is enough budget remaining to cover the node setup cost $s_i$. If this is possible, the remaining budget is depleted and used to deploy control resources at node $i$. The control level of node $i$ is determined by taking the inverse of the variable cost function $g_s(x_i)$ corresponding to the remaining budget available, i.e. $B - U - s_i$. After this last pass, the remaining budget is null and the algorithm stops.

The pseudo-code of this resource allocation procedure is summarized in Algorithm 1.

**Algorithm 1.** Greedy outbreak control resource allocation.

**Input:** Sorted node set $\bar{V}$, budget $B$
**Output:** Control vector $\boldsymbol{x}$
$\boldsymbol{x} \leftarrow \mathbf{0}$
$U \leftarrow 0$
**for** $i$ in $\bar{V}$:
  **if** $U + s_i + g_s(1) < B$ **then**:
    $x_i \leftarrow 1$
    $U \leftarrow U + s_i + g_s(1)$
  **else if** $U + s_i < B$ **then**
    $x_i \leftarrow g_s^{-1}(B - U - s_i)$
    $U \leftarrow B$
   **break**

We consider multiple ranking strategies to determine $\bar{V}$. Each control strategy tested is presented in Table 1, as well as the Baseline (B), in which no control is implemented. The first control strategy, Largest Population (LP) simply targets control at airports in cities with the largest population. The next three strategies exploit known properties of the air traffic network, specifically its hub and spoke structure. The Most Travelled (MT) strategy seeks to target control at airports that are highly trafficked based on total incoming and outgoing volumes, while the Most Connected (MC) and Effective Path (EP) strategies seek to target control at airports that are highly connected to the outbreak source based on travel volumes. The last two strategies utilize learned outcomes from a stochastic epidemic simulation model to inform outbreak control. The First Case (1C) strategy aims to control the set of airports most likely to see the first infected passenger (for a given outbreak scenario), while the First Order Uniform (1OU) targets control at the airports where it is likely to have the largest relative marginal impact. The four strategies, MC, EP, 1C and 1OU utilize knowledge of initial outbreak conditions, while MT and LP airport sets are selected independent of the outbreak state.

## Data

The metapopulation network is constructed using global passenger air travel data from 2015 provided by the International Air Transport Association (IATA)[60]. IATA data consists of global ticket sales which account for true origins and final destinations, and represents 90% of all commercial flights. The remaining 10% of trips are modeled using airline market intelligence. The data provided from IATA includes monthly passenger travel volumes for all travel routes connecting airport pairs (including stopovers), representing nearly 83% of global traffic volumes. The final network used in this study contains the top 99% of the travelled routes provided, resulting in a network with approximately 500,000 routes, 2,908 cities, and 3,267 airports. The city populations served by each airport are based on the population densities provided by Oak Ridge National Laboratory's LandScan[61]. The population size for each city was based on a 50 km radius centered on each airport as was done previously[62,63] and computed using open source Geographic Information Systems software QGIS (https://qgis.org/). In some cases, multiple cities are serviced by more than one airport, for which all the assigned airport flows are mapped to the same population.

## Results

We demonstrate the performance of the proposed control strategies to mitigate global outbreak spread for a set of hypothetical outbreak scenarios and present results from a cost-benefit analysis, which characterizes the marginal gains in outbreak reduction with respect to increases in available resources, *i.e.*, budget. Further, all strategies are applied to a case study representative of the 2009 H1N1 Pandemic Influenza to illustrate the hypothetical impact of each in a similar outbreak setting.

In this study, only U.S. cities are considered for control, intended to provide an example of how the model may be used by a country's public health authority, which is constrained by an available federal budget for control. The two metrics used to compare the performance of each strategy are i) the total cumulative number of infected individuals in the U.S. at observation time, $t_{obs}$, and ii) the number of infected cities in the U.S. at observation time, $t_{obs}$, where an infected city has at least one infected individual.

**Base Case Analysis.** For the base case analysis, all of the proposed control strategies are implemented and compared for three independent hypothetical outbreak scenarios that vary by the outbreak source location. Three source cities are selected (1) Orlando, Florida, (2) Portland, Oregon and (3) Honolulu, Hawaii, and each is initialized with 100 infected individuals at $t = 0$. These cities were chosen because they represent a range of geographic and travel profiles. In the remainder of this work these three cities are denoted by their assigned airport IATA codes, MCO, PDX and HNL, respectively.

To model the cost of control at airports, we consider linear cost functions. Setup costs $s_i$ represent screening equipment costs based on the total incoming flow at each airport $i \in V$. Let $M$ be the cost of a passenger screening machine and $C$ its capacity, we set $s_i = \frac{M}{C}\sum_{j\in\Gamma_i^-}\sum_{k\in\Pi_{ji}} f_{ji}^k$ , where $f_{ji}^k$ is the average flow on path $k$ over the control

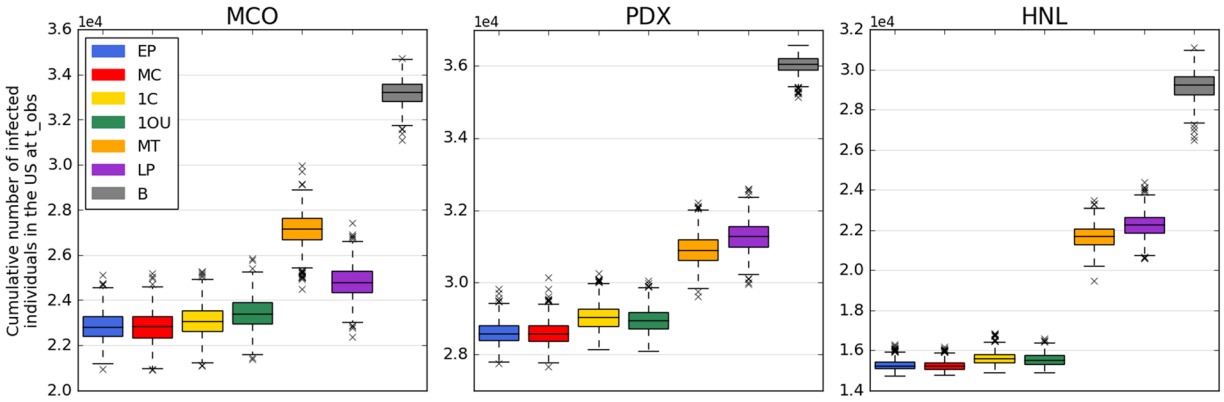


**Figure 1.** Control strategy performance for the base case scenarios MCO, PDX and HNL in terms of number of infected individuals. The figure reports the cumulative number of infected individuals in the U.S. at the observation time $t_{obs} = 50$ days for each control strategy. Each boxplot represents the distribution of the criterion measured over 1,000 simulations of the stochastic metapopulation epidemic model under the corresponding control strategy.

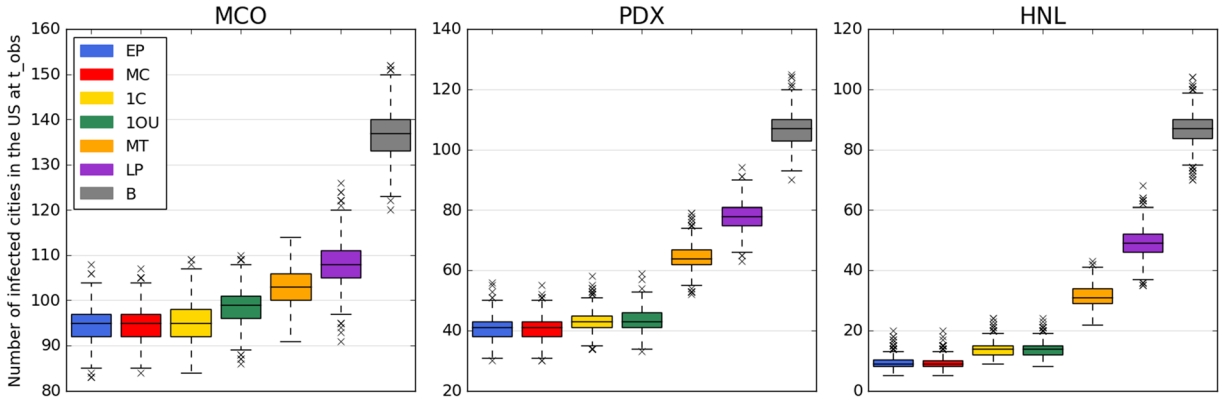


**Figure 2.** Control strategy performance for the base case scenarios MCO, PDX and HNL in terms of number of infected cities. The figure reports the number of infected cities in the U.S. at the observation time $t_{obs} = 50$ days for each control strategy, where a city is categorized as infected if it contains at least one infected individual. Each boxplot represents the distribution of the criterion measured over 1,000 simulations of the stochastic metapopulation epidemic model under the corresponding control strategy.

period. For the variable cost, we assume that the cost of screening a passenger is represented by $P$ and set $g_s(x_i) = x_i t_{obs} P \sum_{j \in \Gamma_i^-} \sum_{k \in \Pi_{ji}} f_{ji}^k$ to model the impact of deploying control resources over a varying time period and at a varying level of control. For the base case analysis, we set the available budget $B$ to \$500 million. The parameter values are $M = \$500{,}000$, $C = 10{,}000$ and $P = \$10$, and based on the existing literature[64]. Under this configuration, the available budget is enough to fully control the 13 most travelled airports in the U.S. for $t_{obs} = 50$.

For the scenarios presented in this section, the simulation is set to begin on June 1 and $t_{obs}$ is set to 50 days. The time period of 50 days is chosen to align with the focus of this study, which is to identify the most strategic allocation of outbreak control resources that provides robust border control in the early stages of an outbreak. Thus, the planning time period considered (50–100 days) is intended to represent the high risk period at the early stages of a potential epidemic, and potentially before local control is fully in effect. We assume after this period of time, alternative (and more effective) local control methods will be put in place, which are not accounted for in this analysis. For the base case analysis, the hypothetical virus has values of $\alpha = 0$, $\beta = 0.25$, $\gamma = 0.143$ and $\lambda = 1$. The chosen baseline parameters correspond to a disease with a reproductive ratio $R_0 = 1.75$. The U.S. level outbreak dynamics (SIR curves) for each of the three source cities is presented in the Supplementary Material (Section B). For the base case, under a do-nothing scenario, the final proportion of the U.S. population infected is around 70%. At $t_{obs} = 50$, the chosen planning time period in this study, only 0.01% of the U.S. population is infected, or around 30 to 35 k individuals on average.

Extensive sensitivity analysis was conducted to assess how the different control strategies respond to differing disease characteristics, model assumptions and policy decisions. Specifically, we explored how changes in the contact rate, planning horizon, control start time, control effectiveness, and source screening impact the performance of each strategy. Results for all sensitivity analysis are provided in the Supplementary Material (Section C). All results are based on 1,000 simulation runs of the stochastic metapopulation epidemic model. Each run of the simulation took an average runtime of three seconds on a desktop computer.

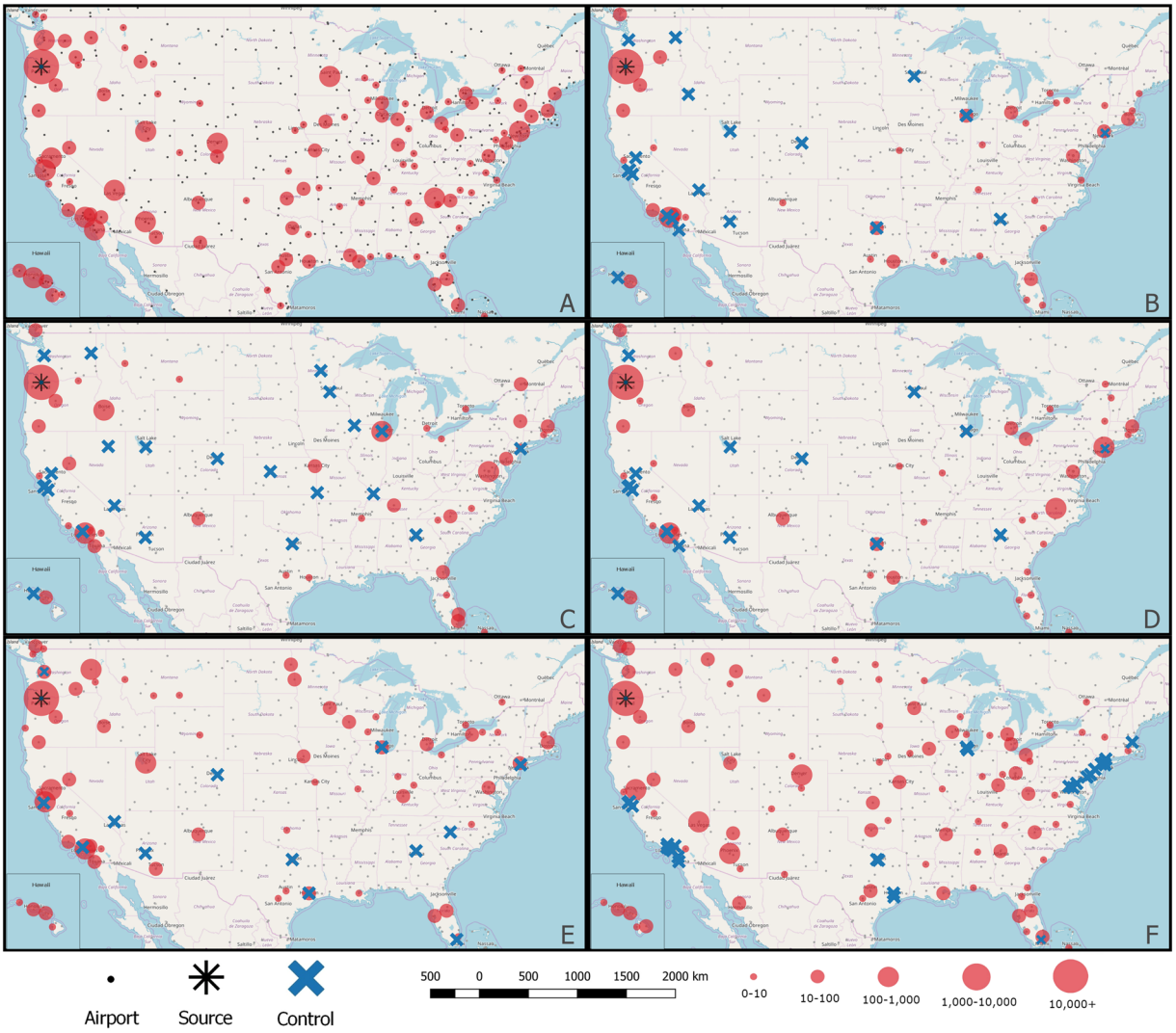


**Figure 3.** Visualization of outbreak control strategies and impact in the U.S. The results are illustrated for the PDX scenario, with an asterisk marking the source of infection. The maps therein display the set of airports selected for control (blue crosses) for each strategy, and the expected size of outbreak in each city at the observation date $t_{obs} = 50$ days (red circles) measured over 1,000 simulations of the stochastic metapopulation epidemic model. The panels correspond to the following strategies: A) No-control, B) MC/EP, C) 1C, D) 1OU, E) MT and F) LP. The red circles are sized proportional to the outbreak size. The grey circles represent the complete set of airports that can be feasibly selected for control. The maps were generated using the standard base map from OpenStreetMap. ©OpenStreetMap contributors, CC-BY-SA https://www.openstreetmap.org/copyright.

Figure 1 provides the cumulative number of cases in the U.S. at $t_{obs}$ for scenarios MCO, PDX and HNL. The boxplots capture the results for all 1,000 simulations conducted, illustrating the robustness of the results and rankings. In each plot the six proposed strategies to allocate screening resources are compared against the baseline (corresponding to no control) for the respective scenario. Similar trends are evident for all scenarios, with the more simplistic strategies of controlling at the airports in the largest cities or most travelled airports performing poorest, and EP and MC performing best.

For the MCO scenario, the cumulative number of infected individuals is on average 33,200 for the baseline configuration (no control). Using control strategy LP leads to a reduction of 25.2% in the number of cases, compared with a reduction of 31.2% using EP or MC. For the PDX scenario, the amount of infected individuals is on average 36,000 in the baseline case and using EP or MC results in a reduction of 20.6%. For the HNL scenario, we report a reduction of 47.7% from the baseline to the best control strategy as well as substantially less variability across outbreak simulations.

Figure 2 provides the number of infected cities in the U.S. at $t_{obs}$ for the three scenarios. This metric is critical to assess the success of the control strategy with regards to preventing spread into new cities. A similar trend is observed across all scenarios, with strategies EP, MC, 1C and 1OU proving superior to MT and LP. For the MCO scenario, the number of cities affected is on average 137 for the baseline case. Using control strategy MT leads to

a reduction of 24% in the number of cities affected, compared with a reduction of 31% for EP/MC. For the PDX scenario, the amount of infected is on average 107 in the baseline case and 40 for the EP/MC cases resulting in a reduction of 63%. For the HNL scenario, we observe a more substantial reduction of 90% in the number of cities affected from the baseline to the best control strategy.

Figure 3 illustrates the decisions and impact of the proposed border control strategies against the baseline no-control case for a particular hypothetical outbreak scenario. Specifically, the maps display the set of airports selected for control (blue crosses), and the expected size of the outbreak in each affected city at the observation date (red circles). The red circles are sized proportional to the cumulative number of cases at $t_{obs}$ at each location. The small grey circles illustrate the full set of airports that can be selected for control, further highlighting the scale and the complexity of the problem. The results shown in Fig. 3 are for the PDX scenario, and the panels correspond to the following strategies: A) No-control, B) MC/EP, C) 1C, D) 1OU, E) MT and F) LP, and. Note: EP and MC have the same airport control set in the PDX scenario. The list of airports selected for control in each strategy (and shown in the figure) is provided in the Supplementary Material (Section E). It is worth noting that both LP and MT airport control sets are fixed for a given budget, and independent of the outbreak scenario, unlike the proposed network- and simulation-driven control strategies.

The results in Fig. 3 indicate both the large spatial variability in the airport sets selected for control across strategies, and their respective impact on outbreak location and outbreak size. Specifically, only five (of the 300 possible) airports are selected across all control strategies, *i.e.*, LAX, ORD, DFW, JFK and SFO. As expected, when no control is deployed (Panel A), there are substantially more cities affected than there are under any of the control strategies considered (Panels B–F). While both MC/EP and MT (Panels B and E, respectively) control relatively few and more spatially sparse airports, EP (Panel B) provides substantially greater reduction in both the outbreak size and number of affected cities (as depicted in Figs. 1 and 2) due to the strategic nature of the path-based metric. In contrast, MT (Panel E) is the second worst performing strategy, and spends large portions of its budget on controlling fewer, more heavily travelled airports, which may not play a critical role in the early stages of the outbreak. The poorest performing strategy is LP (Panel F), which, perhaps surprisingly, controls more airports than all other strategies. However, many of these airports are far-removed from the outbreak source in terms of traffic volumes, and are therefore less likely to contribute to spread early in the outbreak. Strategy 1C (Panel C) suffers from similar faults, indicating that budget is spent on airports that are less expensive to control, but also provide less overall impact. These results further highlight the necessity for optimized resource allocation in the context of outbreak control.

**Cost-Benefit Analysis.** To explore the impact of resource availability on epidemic spread, we conduct a cost-benefit analysis by varying the budget available for control. Specifically, we explore a range of budgets from \$0.25 bil to \$1.25 bil, in \$0.25 bil increments. Note that given the cost functions used in this study, a budget of \$1.35 bil is enough to fully control all airports within the U.S. for a 50-day period. The results are summarized in Fig. 4, which illustrates the impact of budget on the effectiveness of each strategy for all three source scenarios. At \$0 bil and \$1.25 bil the strategies perform nearly the same, which represent the cases of no control and close to full control at all airports, respectively. For budgets in between these values, the impact of all strategies decreases with budget as expected, with control strategy EP consistently outperforming other strategies. The weaker strategies, MT and LP, decrease in a more linear fashion compared to the other four, MC, EP, 1C and 1OU, which show evidence of decreasing marginal returns, *e.g.*, there is negligible improvements after the budget reaches \$0.75 bil. These results have valuable implications for policy makers, and indicate the potentially cost-effective nature of resource allocation if assigned strategically.

**Case Study: 2009 H1N1 Influenza Pandemic.** To illustrate the ability of the simulation model to replicate a realistic pandemic, we consider the 2009 H1N1 influenza pandemic as a case study, and quantify the performance of the proposed strategies under similar outbreak conditions. The simulation model was calibrated to the 2009 H1N1 outbreak data at both the U.S. and global scales. The calibration methodology and results are included in the Supplementary Material (Section D). The final calibrated disease parameters are $\alpha = 1$, $\beta = 0.475$, $\gamma = 0.25$ and $\lambda = 1$. For all simulations the starting date of the outbreak is set as the 5$^{th}$ of February 2009 with 1 infected individual placed into the city of Veracruz, Mexico[1]. Each control strategy is deployed four weeks (28 days) after the first case appeared in Mexico, at which time there were about 100 local cases in Mexico (based on the simulation). The cost function, budget and screening costs are the same as those used in the base case analysis. The results are based on 1,000 simulations of the stochastic model, and the observation time is set to $t_{obs} = 100$ days after the first case.

The case study illustrates the hypothetical impact of implementing the proposed strategies for an outbreak similar in characteristics to the 2009 H1N1 influenza pandemic. The performance of each strategy based on the two metrics used for evaluation, *i.e.*, the cumulative number of cases and number of infected cites in the U.S. at the time of observation, under each of the proposed strategies are illustrated in Fig. 5. The results highlight the EP strategy to once again dominate, and LP and MT to perform the poorest. Specifically, the EP strategy can reduce the final outbreak size by 37.1% relative to the baseline, with the number of infected cities dropping from 115 to 86, representing a 25.2% decrease.

## Discussion

This work addresses the challenge of pandemic mitigation planning through border control, specifically using entry screening at airports to minimize the potential harm posed by an outbreak. We present an ensemble of control strategies that are evaluated based on their ability to reduce the cumulative number of cases and the number of cities infected at a target observation time. The decision-support framework provided can be implemented in real-time at the early stages of a confirmed or suspected outbreak. An in-depth analysis is presented for multiple

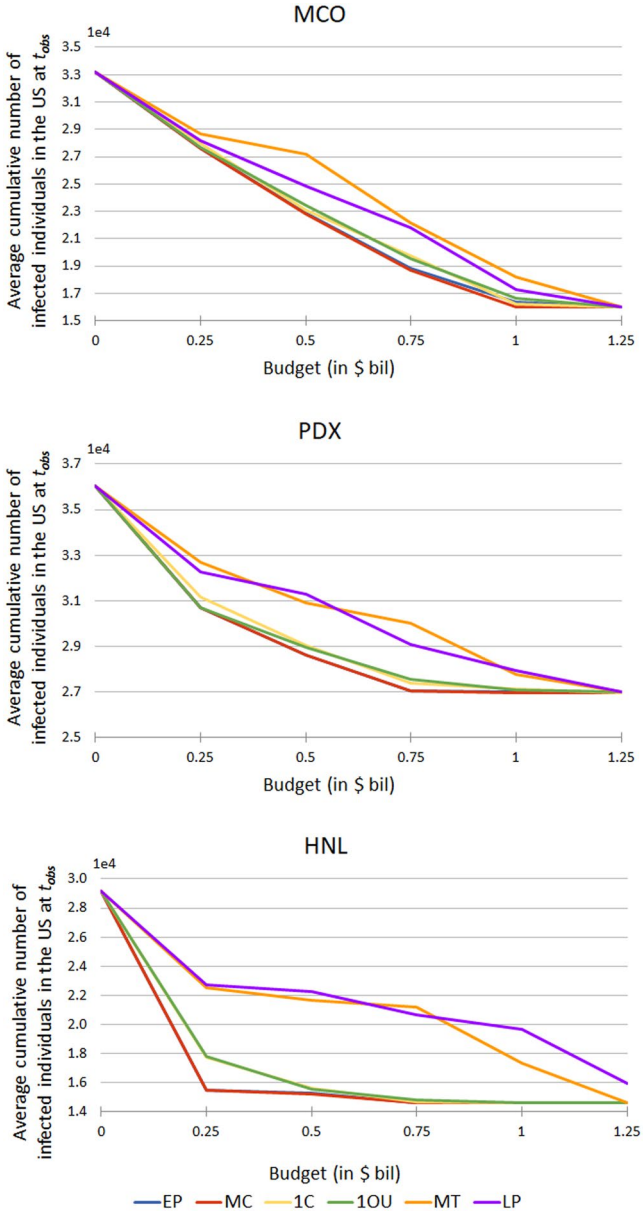


**Figure 4.** Cost-benefit analysis for the base case scenarios MCO, PDX and HNL in terms of number of infected individuals. The figure depicts the average cumulative number of infected individuals in the U.S. at the observation time $t_{obs} = 50$ days based on a varying border control budget $B$ expressed in billions of dollars. The data points converge at $B = 0$ which corresponds to the baseline (no control) case.

hypothetical outbreak scenarios, and a case study is conducted to illustrate the performance of the control strategies for an outbreak similar to the 2009 H1N1 influenza pandemic. Further, extensive sensitivity analysis illustrates the robustness of the control strategies to various modelling parameters and assumptions.

The best performing control strategies are the network-driven strategies, *e.g.*, EP and MC, which are shown to be robust across a range of outbreak scenarios and model assumptions. In contrast, the more simplistic strategies, *e.g.*, controlling the most travelled airports (MT) or the airports in the largest cities (LP), perform the poorest. The superiority of the network-driven strategies highlights the significance of the heterogeneity of the world air traffic network in outbreak spreading[5,16], which can be exploited by policy makers for the purposes of pandemic planning and mitigation. The two control strategies informed by simulation, 1OU and 1C, rank in the middle in terms of performance in most scenarios considered in this work, however the relative performance of the strategies is sensitive to the outbreak initial conditions, as highlighted by the case study. While the results from the analysis presented indicate the EP strategy to be reliably superior, the larger contribution of this work is the modeling framework, which can be implemented in real-time for any outbreak scenario given reported case counts and locations. Additionally, the integrated framework is flexible, and can be easily extended to incorporate and compare the performance of additional strategies if desired.

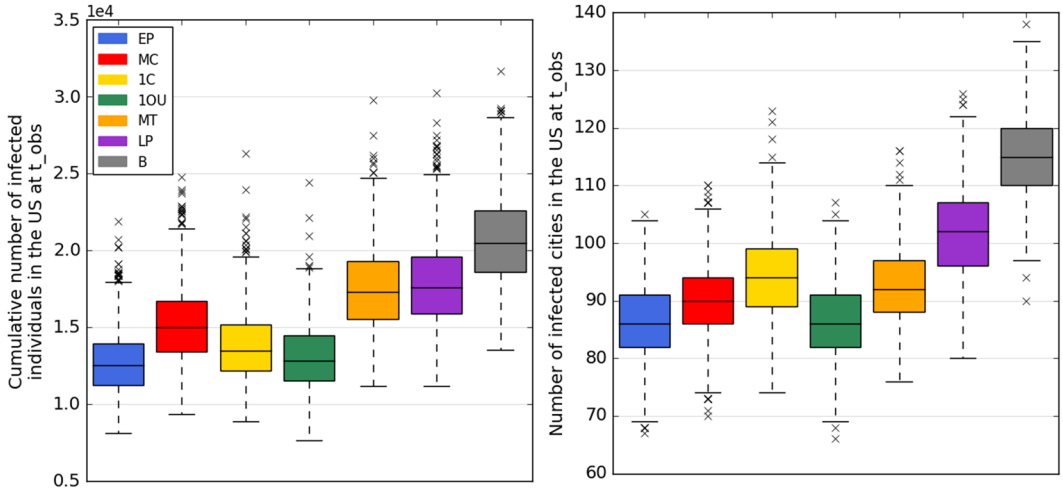


**Figure 5.** Control Strategy performance for the 2009 H1N1 influenza pandemic case study in terms of number of infected individuals (left) and number of infected cities (right) in the U.S. The figure depicts the performance of each control strategies for an observation time $t_{obs} = 100$ days, and assuming border control is deployed at 28 days after the first infected individual was reported in Mexico. Each boxplot represents the distribution of the criterion measured over 1,000 simulations of the stochastic metapopulation epidemic model under the corresponding control strategy.

The cost-benefit analysis highlights two critical issues. Firstly, a given budget can be used more effectively if the control decisions are made strategically. *e.g.*, EP/MC can achieve the same effectiveness as MT/LP for nearly half the budget. Second, there is evidence of decreasing marginal returns for the superior strategies, indicating minimal benefits may be gained by spending more on border control beyond a certain threshold. This cost-effectiveness threshold is critical for policy makers, who can choose to redirect available control resources towards alternative control strategies for outbreak mitigation which may be more effective.

The case study illustrates the expected performance of each control strategy for an outbreak similar in behavior to the 2009 H1N1 influenza pandemic. This analysis critically introduces an asymptomatic exposed state, which poses an additional challenge for control, because even with perfect screening not all infected travelers can be identified. Even so, results from the case study reveal the reliably superior performance of EP as an outbreak control strategy, and once again, the value gained by strategically allocating control resources.

The sensitivity analysis on the infection contact rate (see Section C.1) illustrates that the ranking and effectiveness of the control strategies are robust with regards to the rate of spread. Furthermore, the control strategies are observed to be more effective (relative to the no control scenario) and more robust for faster spreading viruses. The more reliable performance for higher contact rates can be attributed to the highly stochastic nature of outbreaks in their early stages, which is heavily dependent on where the first few infected cases spread to, whereas a faster spreading virus will result in a larger number of infected people travelling, reducing the variability across outbreak scenarios.

Sensitivity to observation time and control start date was conducted to illustrate the impact of these two implementation options available to policy makers. The model appears to be highly sensitive to observation time (see Section C.2), which is due to the exponential nature of outbreak growth; there is a small number of cases and affected cities in the early stages of the outbreaks (*e.g.* 25 days) compared to considerably higher infection rates at later time epochs (*e.g.* 100 days). This sensitivity analysis highlights the critical (short) timeline during which border control has the potential to play a substantial role, after which local control will be most impactful. The impact of delaying border control was also evaluated (Section C.3), and the results again highlight the robustness of the strategy rankings. The results also demonstrate the importance of implementing control in a timely fashion, with the best performing strategies revealed to be the most sensitive to delayed start times.

To address the assumption of perfect control, we evaluated the effectiveness of the proposed strategies under imperfect control conditions (Section C.4), limiting the maximum control rate to 80% and 90%, respectively. While the impact of control decreases with the control level, the relative performance and ranking across strategies remains constant, suggesting the best performing strategies remain effective and reliable under imperfect control. Similar results are observed in the case study, which utilizes an SEIR model, *i.e.*, asymptotic infected travelers are able to evade control and introduce infection into new cities.

The final sensitivity analysis evaluated the impact of outgoing passenger screening at the source of infection. This work assumes that outgoing screening at the source is an obvious decision; therefore the model addresses the more challenging problem of selecting which locations other than the source(s) should be prioritized for incoming passenger screening. The sensitivity analysis results reveal that when outgoing screening is implemented at the source (Section C.5) the proposed control strategies behave predictably, *i.e.*, the outbreak spreads faster but the strategy rankings remain consistent. Critically, the best performing strategies perform well even at low levels of outgoing screening, *i.e.*, ineffective screening, while the poorest performing strategies are more sensitive to the effectiveness of outgoing source screening.

Lastly, there are modeling assumptions and limitations of this study. First, the model only accounts for passenger air travel, and excludes mobility within and between cities via other modes of transport. Second, local disease spread (within a city) is modeled deterministically, and a uniform contact rate is used across all populations. Third, the model is currently limited to global control decisions through passenger screening, and does not evaluate local control mechanisms (prophylaxtics, vaccines, school closures, etc), with the exception of source exit screening. Fourth, airport (rather than route) screening rates are the control variable, and therefore assume passenger screening to be uniformly applied across all incoming routes at a given airport. Planned extensions of this work will address these limitations through i) integrating alternative modes of transport into the model, ii) adding additional decision variables to optimize local control decisions, iii) the development of a link-based modelling formulation to allow specific travel routes to be identified for screening (as opposed to airports), and iv) introducing a dynamic resource allocation formulation which relaxes the assumption of constant control across the entire planning period. These extensions would provide more degrees of freedom to improve the impact of control resources and further help in minimizing the risk posed by global outbreaks. The set of limitations and extensions listed currently lie outside the scope of this study, and provide the basis for future research.

## Data Availability

The air traffic data used in this study is available for purchase from IATA Passenger Intelligence Services (Pax-IS), https://www.iata.org/services/statistics/intelligence/paxis. The population data is publically available from ORNL's LandScan[55,58,61], https://landscan.ornl.gov/. H1N1 case data used for model calibration is publically available from CDC and WHO, and referenced in the supplementary materials.

## References


1. CDC. Outbreak of Swine-origin influenza A (H1N1) virus infection—Mexico, March-April 2009. *Morb. Mortal. Wkly. Rep* **58**, 467–470 (2009).
2. Smith, R. D. Responding to global infectious disease outbreaks: Lessons from SARS on the role of risk perception, communication and management. *Social Science & Medicine* **63**, 3113–3123, https://doi.org/10.1016/j.socscimed.2006.08.004 (2006).
3. WHO. WHO Situation Report: Zika Virus, Microcephaly and Guillain-Barre Syndrome (2016).
4. Massaro, E., Ganin, A., Perra, N., Linkov, I. & Vespignani, A. Resilience management during large-scale epidemic outbreaks. *Scientific Reports* **8**, 1859, https://doi.org/10.1038/s41598-018-19706-2 (2018).
5. Wang, L. & Wu, J. T. Characterizing the dynamics underlying global spread of epidemics. *Nature Communications* **9**, 218, https://doi.org/10.1038/s41467-017-02344-z (2018).
6. Nowzari, C., Preciado, V. M. & Pappas, G. J. Analysis and Control of Epidemics: A Survey of Spreading Processes on Complex Networks. *IEEE Control Systems* **36**, 26–46, https://doi.org/10.1109/MCS.2015.2495000 (2016).
7. Poletto, C. *et al.* Assessing the impact of travel restrictions on international spread of the 2014 West African Ebola epidemic. *Euro surveillance: bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin* **19**, 20936 (2014).
8. Ferguson, N. M. *et al.* Strategies for mitigating an influenza pandemic. *Nature* **442**, 448, https://doi.org/10.1038/nature04795, https://www.nature.com/articles/nature04795#supplementary-information (2006).
9. Mossong, J. *et al.* Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases. *PLOS Medicine* **5**, e74, https://doi.org/10.1371/journal.pmed.0050074 (2008).
10. Morens, D. M., Folkers, G. K. & Fauci, A. S. The challenge of emerging and re-emerging infectious diseases. *Nature* **430**, 242, https://doi.org/10.1038/nature02759 (2004).
11. Heesterbeek, H. *et al.* Modeling infectious disease dynamics in the complex landscape of global health. *Science* **347** (2015).
12. Wang, L. & Li, X. Spatial epidemiology of networked metapopulation: an overview. *Chinese Science Bulletin* **59**, 3511–3522, https://doi.org/10.1007/s11434-014-0499-8 (2014).
13. Pastor-Satorras, R., Castellano, C., Van Mieghem, P. & Vespignani, A. Epidemic processes in complex networks. *Reviews of Modern Physics* **87**, 925–979, https://doi.org/10.1103/RevModPhys.87.925 (2015).
14. Castellano, C. & Pastor-Satorras, R. Thresholds for Epidemic Spreading in Networks. *Physical Review Letters* **105**, 218701, https://doi.org/10.1103/PhysRevLett.105.218701 (2010).
15. Colizza, V., Pastor-Satorras, R. & Vespignani, A. Reaction–diffusion processes and metapopulation models in heterogeneous networks. *Nature Physics* **3**, 276, https://doi.org/10.1038/nphys560, https://www.nature.com/articles/nphys560#supplementary-information (2007).
16. Brockmann, D. & Helbing, D. The Hidden Geometry of Complex, Network-Driven Contagion Phenomena. *Science* **342**, 1337 (2013).
17. Sun, G.-Q., Jusup, M., Jin, Z., Wang, Y. & Wang, Z. Pattern transitions in spatial epidemics: Mechanisms and emergent properties. *Physics of Life Reviews* **19**, 43–73, https://doi.org/10.1016/j.plrev.2016.08.002 (2016).
18. Wang, Y., Ma, J., Cao, J. & Li, L. Edge-based epidemic spreading in degree-correlated complex networks. *Journal of Theoretical Biology* **454**, 164–181, https://doi.org/10.1016/j.jtbi.2018.06.006 (2018).
19. Colizza, V. & Vespignani, A. Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: Theory and simulations. *Journal of Theoretical Biology* **251**, 450–467, https://doi.org/10.1016/j.jtbi.2007.11.028 (2008).
20. Roberts, M., Andreasen, V., Lloyd, A. & Pellis, L. Nine challenges for deterministic epidemic models. *Epidemics* **10**, 49–53, https://doi.org/10.1016/j.epidem.2014.09.006 (2015).
21. Wang, Y. *et al.* Global analysis of an SIS model with an infective vector on complex networks. *Nonlinear Analysis: Real World Applications* **13**, 543–557, https://doi.org/10.1016/j.nonrwa.2011.07.033 (2012).
22. Britton, T. & Lindenstrand, D. Epidemic modelling: Aspects where stochasticity matters. *Mathematical Biosciences* **222**, 109–116, https://doi.org/10.1016/j.mbs.2009.10.001 (2009).
23. Nelson, M. I. *et al.* Stochastic Processes Are Key Determinants of Short-Term Evolution in Influenza A Virus. *PLOS Pathog* **2**, e125 (2006).
24. Charu, V. *et al.* Human mobility and the spatial transmission of influenza in the United States. *PLOS Computational Biology* **13**, e1005382, https://doi.org/10.1371/journal.pcbi.1005382 (2017).
25. Balcan, D. *et al.* Modeling the spatial spread of infectious diseases: The GLobal Epidemic and Mobility computational model. *Journal of computational science* **1**, 132–145 (2010).
26. Foundation, E. *Spatio-Temporal Epidemiological Modeler*, http://www.eclipse.org/stem/ (2018).
27. Tizzoni, M. *et al.* Real-time numerical forecast of global epidemic spreading: case study of 2009 A/H1N1pdm. *BMC Medicine* **10**, 165, https://doi.org/10.1186/1741-7015-10-165 (2012).
28. Grenfell, B. & Harwood, J. (Meta)population dynamics of infectious diseases. *Trends Ecol. Evol* **12**, 395–399 (1997).
29. Hanski, I. Metapopulation dynamics. *Nature* **396**, 41–49 (1998).
30. Riley, S. Large-scale spatial-transmission models of infectious disease. *Science* **316**, 1298–1301 (2007).

31. Daughton, A. R., Generous, N., Priedhorsky, R. & Deshpande, A. An approach to and web-based tool for infectious disease outbreak intervention analysis. *Scientific Reports* **7**, 46076, https://doi.org/10.1038/srep46076 (2017).
32. Gautreau, A., Barrat, A. & Barthelemy, M. Global disease spread: Statistics and estimation of arrival times. *J. Theor. Biol* **251**, 509–522 (2008).
33. Scalia Tomba, G. & Wallinga, J. A simple explanation for the low impact of border control as a countermeasure to the spread of an infectious disease. *Math. Biosci.* **214**, 70–72 (2008).
34. Flahault, A., Vergu, E., Coudeville, L. & Grais, R. F. Strategies for containing a global influenza pandemic. *Vaccine* **24**, 6751–6755, https://doi.org/10.1016/j.vaccine.2006.05.079 (2006).
35. Bajardi, P. *et al*. Human Mobility Networks, Travel Restrictions, and the Global Spread of 2009 H1N1 Pandemic. *PLOS ONE* **6**, e16591, https://doi.org/10.1371/journal.pone.0016591 (2011).
36. Colizza, V., Barrat, A., Barthelemy, M., Valleron, A.-J. & Vespignani, A. Modeling the Worldwide Spread of Pandemic Influenza: Baseline Case and Containment Interventions. *PLOS Medicine* **4**, e13, https://doi.org/10.1371/journal.pmed.0040013 (2007).
37. Weng, W. & Ni, S. Evaluation of containment and mitigation strategies for an influenza A pandemic in China. *SIMULATION* **91**, 407–416, https://doi.org/10.1177/0037549715581637 (2015).
38. Grais, R. F., Ellis, J. H., Kress, A. & Glass, G. E. Modeling the Spread of Annual Influenza Epidemics in the U.S.: The Potential Role of Air Travel. *Health Care Management Science* **7**, 127–134, https://doi.org/10.1023/B:HCMS.0000020652.38181.da (2004).
39. Rvachev, L. A. & Longini, I. M. A mathematical model for the global spread of influenza. *Math. Biosci* **75**, 3–22 (1985).
40. Colizza, V., Barrat, A., Barthélemy, M. & Vespignani, A. The role of the airline transportation network in the prediction and predictability of global epidemics. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 2015–2020 (2006).
41. Ruan, Z., Wang, C., Ming Hui, P. & Liu, Z. Integrated travel network model for studying epidemics: Interplay between journeys and epidemic. *Scientific Reports* **5**, 11401, https://doi.org/10.1038/srep11401, https://www.nature.com/articles/srep11401#supplementary-information (2015).
42. Hufnagel, L., Brockmann, D. & Geisel, T. Forecast and control of epidemics in a globalized world. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 15124 (2004).
43. Epstein, J. M. *et al*. Controlling Pandemic Flu: The Value of International Air Travel Restrictions. *PLoS ONE* **2**, e401, https://doi.org/10.1371/journal.pone.0000401 (2007).
44. Cooper, B. S., Pitman, R. J., Edmunds, W. J. & Gay, N. J. Delaying the International Spread of Pandemic Influenza. *PLOS Medicine* **3**, e212, https://doi.org/10.1371/journal.pmed.0030212 (2006).
45. Kenah, E., Chao, D., Matrajt, L., Halloran, M. & Longini, I. The global transmission and control of influenza. *PloS One* **6**, e19515 (2011).
46. Warren, A., Bell, M. & Budd, L. Airports, localities and disease: Representations of global travel during the H1N1 pandemic. *Health & Place* **16**, 727–735, https://doi.org/10.1016/j.healthplace.2010.03.004 (2010).
47. Cosford, P. Advantages of airport screening for Ebola. *BMJ: British Medical Journal* **349** (2014).
48. Read, J. M., Diggle, P. J., Chirombo, J., Solomon, T. & Baylis, M. Effectiveness of screening for Ebola at airports. *The Lancet* **385**, 23–24, https://doi.org/10.1016/S0140-6736(14)61894-8 (2015).
49. Evans, A. D. & Thibeault, C. Prevention of spread of communicable disease by air travel. *Aviation, space, and environmental medicine* **80**, 601–602 (2009).
50. Hollingsworth, T. D., Ferguson, N. M. & Anderson, R. M. Will travel restrictions control the international spread of pandemic influenza? *Nature Medicine* **12**, 497, https://doi.org/10.1038/nm0506-497, https://www.nature.com/articles/nm0506-497#supplementary-information (2006).
51. Longini, I. M. *et al*. Containing Pandemic Influenza at the Source. *Science* **309**, 1083 (2005).
52. Ferguson, N. M. *et al*. Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature* **437**, 209, https://doi.org/10.1038/nature04017, https://www.nature.com/articles/nature04017#supplementary-information (2005).
53. Brownstein, J., Wolfe, C. & Mandl, K. Empirical evidence for the effect of airline travel on inter-regional influenza spread in the United States. *PLoS Med* **3**, e401 (2006).
54. Balcan, D. *et al*. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proc Natl Acad Sci USA* **106**, 21484–21489 (2009).
55. Chen, N., Gardner, L. & Rey, D. Bilevel Optimization Model for the Development of Real-Time Strategies to Minimize Epidemic Spreading Risk in Air Traffic Networks. *Transportation Research Record: Journal of the Transportation Research Board* **2569**, 62–69, https://doi.org/10.3141/2569-07 (2016).
56. Chen, N., Rey, D. & Gardner, L. Multiscale Network Model for Evaluating Global Outbreak Control Strategies. *Transportation Research Record: Journal of the Transportation Research Board* **2626**, 42–50, https://doi.org/10.3141/2626-06 (2017).
57. Kermack, W. O. & McKendrick, A. G. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A* **115**, 700, https://doi.org/10.1098/rspa.1927.0118 (1927).
58. Allen, L. J. S. In *Mathematical Epidemiology* (eds Brauer, F., Driessche, P. V. D. & Wu, J.) 81–130 (Springer Berlin Heidelberg, 2008).
59. Selvey, L. A., Antão, C. & Hall, R. Evaluation of Border Entry Screening for Infectious Diseases in Humans. *Emerging Infectious Disease journal* **21**, 197, https://doi.org/10.3201/eid2102.131610 (2015).
60. IATA. *Passenger Intelligence Services (PaxIS)*, https://www.iata.org/services/statistics/intelligence/paxis (2016).
61. *LandScan, Geographic Information Science & Technology*, https://landscan.ornl.gov/ (2018).
62. Gardner, L. M., Chen, N. & Sarkar, S. Global risk of Zika virus depends critically on vector status of Aedes albopictus. *Lancet Infectious Diseases* **16**, 522–523, https://doi.org/10.1016/S1473-3099(16)00176-6 (2016).
63. Gardner, L., Chen, N. & Sarkar, S. Vector Status of Aedes Species Determines Geographical Risk of Autochthonous Zika Virus Establishment. *PLOS NTD*, https://doi.org/10.1371/journal.pntd.0005487 (2017).
64. Gillen, D. & Morrison, W. G. Aviation security: Costing, pricing, finance and performance. *Journal of Air Transport Management* **48**, 1–12, https://doi.org/10.1016/j.jairtraman.2014.12.005 (2015).


## Author Contributions

The authors contributed to (A) conceive and design the experiments, (B) perform the experiments, (C) write the paper, (D) develop the model, (E) perform the data driven simulations and (F) analyse the data. Aleksa Zlojutro B, C, D, E, F; David Rey A, C, D, F; Lauren Gardner A, C, D, F. All authors reviewed the results and approved the final version of the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at https://doi.org/10.1038/s41598-019-38665-w.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2019