



OPEN

Optimizing hospital distribution across districts to reduce tuberculosis fatalities

Mi Jin Lee¹, Kanghun Kim², Junik Son³ & Deok-Sun Lee¹✉

The spatial distributions of diverse facilities are often understood in terms of the optimization of the commute distance or the economic profit. Incorporating more general objective functions into such optimization framework may be useful, helping the policy decisions to meet various social and economic demands. As an example, we consider how hospitals should be distributed to minimize the total fatalities of tuberculosis (TB). The empirical data of Korea shows that the fatality rate of TB in a district decreases with the areal density of hospitals, implying their correlation and the possibility of reducing the nationwide fatalities by adjusting the hospital distribution across districts. Approximating the fatality rate by the probability of a patient not to visit a hospital in her/his residential district for the duration period of TB and evaluating the latter probability in the random-walk framework, we obtain the fatality rate as an exponential function of the hospital density with a characteristic constant related to each district's effective lattice constant estimable empirically. This leads us to the optimal hospital distribution which finds the hospital density in a district to be a logarithmic function of the rescaled patient density. The total fatalities is reduced by 13% with this optimum. The current hospital density deviates from the optimized one in different manners from district to district, which is analyzed in the proposed model framework. The assumptions and limitations of our study are also discussed.

Complex systems are organized, by evolution or design, to satisfy the optimization conditions including the minimization of the traveling time in the transportation system¹ and the maximization of the stability of the airline networks², the resilience of the power-grid system^{3,4}, and the growth rate of cellular networks⁵. Likewise, the locations of facilities are expected to be subject to various optimization conditions^{6–8}. Despite the complexity of the facility location decisions^{9,10}, the empirically observed distributions of facilities often show simple and universal features, revealing the nature of the underlying optimization problem. Most remarkably, the spatial density of facilities scaling with the population density^{11–13} with exponent 2/3 or 1 implies that they are distributed to minimize the social opportunity cost such as the commute distance or to maximize the economic profit depending on the distribution of available customers¹⁴.

For coping with diverse social or economic demands in real-world applications, the objective function in the facility distribution optimization may need to be expanded beyond the commute distance or profit. Towards developing such a general theory, here we consider as an example the problem of distributing hospitals across districts to minimize the total fatalities of tuberculosis (TB) by using the empirical data of Korea. While the chemotherapy for TB is well established, showing a success rate as high as 85% on average¹⁵, TB spreads annually to about 10 million patients, being a major cause of death worldwide^{15,16}. In Korea, the incidence of TB is 77 per 100000 as of 2016, which is high compared with other developed countries, e.g., the member countries of the Organization for Economic Cooperation and Development¹⁷. Patients with TB can be cured if they are diagnosed and get treatment timely. Visiting a hospital and taking drugs for about 6 months are necessary for the full recovery from TB¹⁸, which may not be easy from the patients' perspective. Therefore, the accessibility of local hospitals and the well-trained attending staff providing consistent treatment and care should be crucial for the treatment of TB^{16,19}, which is recognized also in the reports of the World Health Organization¹⁵. The correlation between the hospital distribution and the fatality rate of TB in a district is indeed identified in the Korea TB data-sets which we will analyze in the present study; The fatality rate in a district tends to decrease as the areal density of hospitals

¹Department of Physics, Inha University, Incheon, 22212, Korea. ²Financial Engineering Team, Meritz Securities, Seoul, 07326, Korea. ³Department of Family Medicine, Daejeon Sun Medical Center, Daejeon, 34811, Korea. ✉e-mail: deoksun.lee@inha.ac.kr

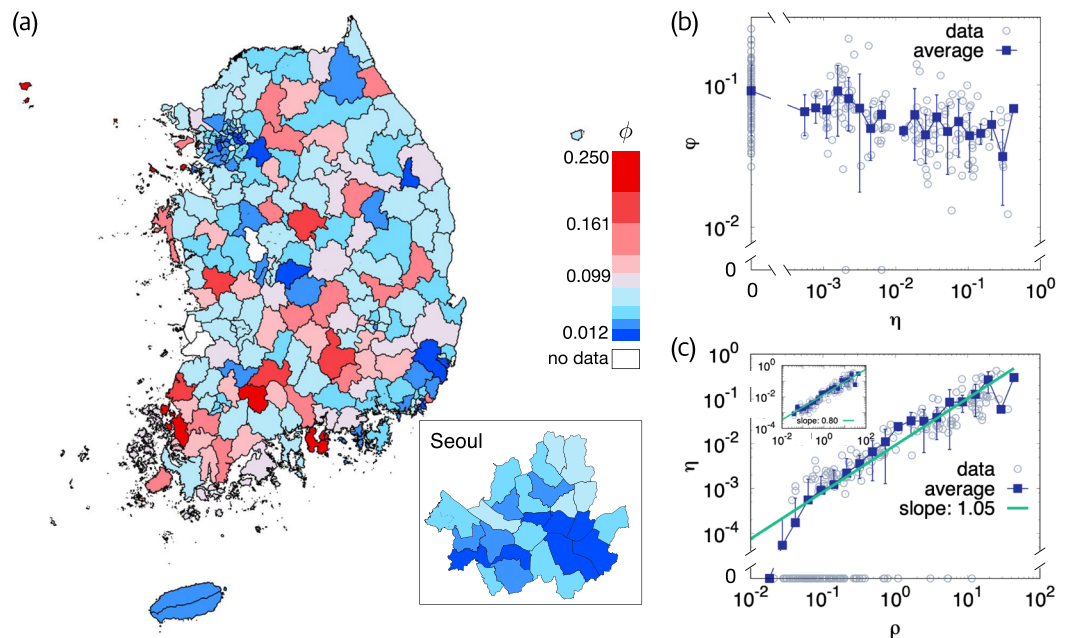


Figure 1. Distribution and relations of the TB fatality rate ϕ , the areal density of hospitals η , and the patient density ρ in Korean districts at the level of Gu, Gun, and Si. The unit of η and ρ is km^{-2} . **(a)** The TB fatality rate ϕ is represented by color in 228 districts. Seoul, the capital city, has 25 Gu's and is shown separately. **(b)** Fatality rate ϕ versus hospital density η . Open circles are the raw data for all districts and filled squares represent the average fatality rate for each given hospital density with the standard deviations as errorbars. The Pearson correlation coefficient is -0.26 with P-value 0.000070 for all districts and -0.25 with $P = 0.0025$ for the districts with both η and ϕ non-zero. **(c)** Hospital density η versus patient density ρ . As in **(b)**, open circles and filled squares represent the raw data and the average value, respectively. The solid line fits the averaged data and its slope is 1.05 ± 0.07 . Inset: The same plot for the districts with $\eta > 0$. The fitting line has slope 0.80 ± 0.045 .

therein increases, which is an important point demanding a quantitative explanation and leads us to expect that relocating hospitals across districts may reduce the total fatalities of TB nationwide.

The optimal distribution of hospitals across districts minimizing the total TB fatalities depends on the concrete form of the fatality rate as a function of the hospital density, which is, however, unknown; The empirically observed negative dependence cannot give this information, as districts are different not only in the hospital density but also in various other properties such as area or population. To address the district-dependent fatality rate, we take a modeling approach, in which the fatality rate is assumed to be identical to the probability of a patient *not* to visit a hospital and get the medical treatment for the duration period of TB. This is motivated by the expectation that a patient is very likely to be cured once she/he gets a proper treatment in a hospital, given the high success rate of the TB chemotherapy equally applicable to all districts. In this framework, the fatality rate turns out to be an exponentially decaying function of the hospital density, and we are able to derive the optimal hospital densities in all districts the collection of which decreases the total fatalities of TB by 13% from the current value. The predicted optimal hospital density is given by a logarithmic function of the rescaled patient density. Our results delineate an analytic approach to the facility optimization problem under an objective function from a public health perspective. The limitation and further generalization of the results will be discussed.

Results

TB fatality rate and hospital density: Empirical data. The incidence and mortality of TB are well recorded in Korea. In Statistics Korea²⁰, we obtain for district $i = 1, 2, \dots, I = 228$ in year 2014 the number of the newly reported TB patients N_i , the number of dead TB patients (fatalities) D_i , the number of private general hospitals H_i , and the area a_i . Here “district” includes three distinct units for administrative division, Gu, Gun, and Si, with the population ranging from 10^4 to 10^6 and smaller than the metropolitan cities.

We are interested in the fatality rate ϕ_i of TB, defined as the ratio of the number of dead TB patients to the number of new TB patients reported for one year in each district i ,

$$\phi_i \equiv \frac{D_i}{N_i}. \quad (1)$$

It is quite different from district to district, ranging between 0.01 and 0.25, as shown graphically in Fig. 1(a). What drives such difference in the TB fatality rate? Taking regularly medical treatments and examinations in hospitals may be the most important for curing TB, which is available in the easy-to-frequently-access medical environment established in the local community. Therefore a difference in the abundance and accessibility of

hospitals in the patients' residential districts will be a major factor giving rise to such variation of the fatality rate with district. In this light, we investigate the relation between the fatality rate ϕ_i and the areal hospital density

$$\eta_i \equiv \frac{H_i}{A_i}, \quad (2)$$

in unit of km^{-2} . In Fig. 1(b) ϕ_i tends to decrease with η_i ; The larger the hospital density is, the smaller the fatality rate is. This correlation is significant with $P < 10^{-4}$. Yet the dependence does not look so strong as expected. This will be shown to be due to that the fatality rate of a district may depend not only on the hospital density but also on other characteristics. In this work, all correlation values are measured in linear scales.

What principle underlies the current spatial distribution of hospitals? The scaling behavior with respect to the patient density has hinted at the answer¹⁴. The hospital density scales with the areal TB patient density $\rho_i \equiv \frac{N_i}{A_i}$ as

$$\eta_i \sim \rho_i^\alpha, \quad (3)$$

in which $\alpha = 1.05 \pm 0.07$ when all districts are included, and $\alpha = 0.80 \pm 0.045$ when the districts having no private general hospital are excluded [Fig. 1(c)]. Many other properties also scale with respect to the patient density. The patient density is almost linearly related to the population density $\rho'_i = P_i/A_i$ with P_i the number of people living in district i [Fig. S1]. The exponent α for the hospitals in United States is close to 1, rather than $2/3$ ¹⁴. These results suggest that the profit maximization affects the hospital distribution. For self-containment, let us sketch the corresponding optimization calculations. The sum of the economic profits of all hospitals distributed across I' districts in a country is given by

$$E_{\text{profit}} = \sum_{i=1}^{I'} H_i \omega \left(\frac{N_i}{H_i} \right) = \sum_i A_i \eta_i \omega \left(\frac{\rho_i}{\eta_i} \right) \quad (4)$$

with $\omega(x)$ the expected profit of a single hospital having x patients available. On the other hand, the sum of the social costs, such as the travel distances, of patients is given by

$$E_{\text{cost}} = \sum_{i=1}^{I'} N_i \psi \left(\frac{A_i}{H_i} \right) = \sum_i A_i \rho_i \psi \left(\frac{1}{\eta_i} \right) \quad (5)$$

with $\psi(x) = x^{1/2}$ being the expected travel distance of a patient residing in a district of x area per hospital. Then, for a fixed total number of hospitals

$$H_{\text{total}} = \sum_{i=1}^{I'} H_i = \sum_I A_i \eta_i, \quad (6)$$

one finds, by solving $\frac{\partial E_{\text{profit}}}{\partial \eta_i} = 0$, E_{profit} to be maximized when $\frac{\rho_i}{\eta_i} = \text{const.}$, corresponding to $\alpha = 1$, and by solving $\frac{\partial E_{\text{cost}}}{\partial \eta_i} = 0$, E_{cost} to be minimized when $\frac{\rho_i}{\eta_i^{3/2}} = \text{const.}$, corresponding to $\alpha = 2/3$ ¹⁴.

Our question is then whether the current hospital distribution, seemingly maximizing the economic profit, is the best also for minimizing the total fatalities of TB

$$E_{\text{fatalities}} = \sum_{i=1}^{I'} D_i = \sum_i N_i \phi_i. \quad (7)$$

Can $E_{\text{fatalities}}$ be reduced by the redistribution of hospitals across district, i.e., some change of $\{\eta_i\}$? To answer this, we should formulate the total fatalities in Eq. (7) as the objective function and minimize it with respect to the hospital density for the given total number of hospitals in Eq. (6). The fatality rate ϕ_i should be some function of the hospital density η_i . If the optimal hospital densities $\{\eta_i^{(\text{opt})}\}$ are obtained by this optimization computation, we will be able to evaluate the quality of the current spatial distribution of hospitals regarding its capacity of TB treatment. Also we will see immediately how to redistribute the hospitals to reduce the TB fatalities. In the present study we do not consider a variation in the numbers of TB patients $\{N_i\}$ but take them for given; The onset and spreading of the TB or a general epidemic disease depend strongly on the topology of human contact networks and the infection rate, which is another important research topic and has been studied extensively^{21,22}.

Fatality rate as a function of hospital density: Model. The empirical fatality rate ϕ_i in Eq. (1) can be considered as the probability of a TB patient to die, losing the opportunity to get proper medical treatment in time. Our idea is to approximate the latter by the probability that a patient does not visit any hospital in her/his residential district for a given period $t_{(\text{TB})} = 3$ years, the empirically reported period of TB duration from onset to either cure or death²³. In this model framework, it determines the fate of a TB patient whether she/he visits a hospital or not for the period of $t_{(\text{TB})}$. The patient will recover if yes, but will be dead otherwise. One can see that this is a trapping problem²⁴ from the viewpoint of a patient; Once a patient (walker) reaches a hospital (trap), she loses the status of a patient (absorbed at the trap). The probability of a walker to survive during a given number of steps corresponding to $t_{(\text{TB})}$ in this trapping problem is translated into the fatality rate of a TB patient in reality.

Suppose that H traps are uniformly and independently distributed in a two-dimensional Euclidean lattice of $L \times L$ sites and that a walker walks around the region, who disappears on reaching any one of the traps. Then the probability of the walker to survive (not to reach any of the traps) after τ steps is given by

$$\phi = \langle (1 - \lambda)^{S(\tau)} \rangle, \quad (8)$$

where $\lambda = \frac{H}{L^2}$ is the density of traps and $S(\tau)$ is the number of distinct sites visited up to τ steps. $\langle \dots \rangle$ represents the average over different realizations of walks. In the limit $\left| \log \left(1 - \lambda \right)^{\frac{\sigma_{S(\tau)}^2}{\langle S(\tau) \rangle}} \right| \ll 1$ reachable when the trap density is sufficiently low or the number of steps is small enough, the survival probability ϕ can be approximated in terms of the first cumulant of the probability distribution of S as²⁵

$$\phi = e^{-\lambda \langle S(\tau) \rangle}, \quad (9)$$

which is the exponential function of the trap density λ . It seems that Eq. (9) allows us to relate the hospital density and the fatality rate. However the dimensionless quantities λ and $\langle S(\tau) \rangle$ are not directly available. In random walks in two dimensions, the expected number of distinct visited sites is known to be²⁴

$$\langle S(\tau) \rangle \propto \frac{\tau}{\log \tau}, \quad (10)$$

which is inserted into Eq. (9) to give

$$\phi = \exp \left(-c \frac{\tau}{\log \tau} \lambda \right) \quad (11)$$

with the coefficient $c = 3.5/1.13^2$ known numerically²⁶. In the opposite limit $\left| \log \left(1 - \lambda \right)^{\frac{\sigma_{S(\tau)}^2}{\langle S(\tau) \rangle}} \right| \gg 1$, the survival of the walker is governed by the probability of a large trap-free region to be formed, which leads to a stretched exponential form $\log \phi \sim \sqrt{\lambda \tau}$ ^{26–28}.

The exponential decay of ϕ with λ in Eq. (9) holds when the hospital density is sufficiently low. The randomness of the mobility pattern is assumed in obtaining Eq. (11). We should remark that the human mobility pattern revealed by tracing the travel routes of bank notes²⁹ or the mobile phone records³⁰ displays deviation from random walk; The radius of gyration of individual trajectories grows logarithmically with time³⁰, in contrast to the square-root scaling in the conventional random walk, and such slow diffusion is known to arise under the memory effect^{31–33} or the spatial quenched disorder²⁴. The assumption we make about the human mobility pattern is that the *coarse-grained* trajectories of individuals on the time scale of $t_{(TB)} = 3$ years, much longer than the previous studies, show the survival probability given in Eq. (11) like random walks. The coarse-grained trajectory is obtained by neglecting the spots swiftly passed by and connecting the remaining notable places which an individual visits and stays for a while in, such as her/his house, workplace, parks, stores, banks, oil stations, and hospitals. In our model, we are interested in whether a hospital is included in the list of such notable places. We cannot check directly the validity of Eqs. (9) and (11), however, we will present indirect evidence that they are reasonable assumptions.

Lattice constant and dimensionless quantities. To relate the survival probability in the 2D trapping problem to the fatality rate of TB, we need to convert the empirical data into dimensionless ones of Eq. (11). To this end, we discretize the region of each district i by introducing the lattice constant a_i , corresponding to the typical length of one single step or the average distance between adjacent notable places appearing in the coarse-grained trajectories. Then the district is represented by the $L_i \times L_i$ Euclidean lattice with $L_i = \sqrt{\frac{A_i}{a_i^2}}$, for which the areal hospital density $\eta_i = H_i/A_i$ is converted to the dimensionless hospital density λ_i as

$$\lambda_i = \frac{H_i}{L_i^2} = \frac{H_i}{\frac{A_i}{a_i^2}} = a_i^2 \eta_i. \quad (12)$$

Let $\ell_{(TB)}$ be the typical travel distance of an individual for $t_{(TB)} = 3$ years. Then the number of steps taken in her/his coarse-grained trajectory for $t_{(TB)}$ in a district i will be given by

$$\tau_i = \frac{\ell_{(TB)}}{a_i}. \quad (13)$$

Plugging Eqs. (12) and (13) into Eq. (11), we find the fatality rate represented as

$$\phi_i = \exp \left(-\frac{\eta_i}{\tilde{\eta}_i} \right) \quad (14)$$

with the characteristic hospital density $\tilde{\eta}_i$ given by

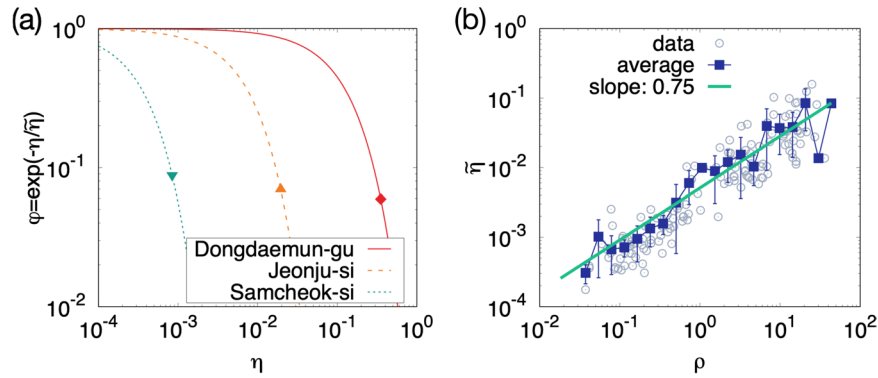


Figure 2. Theoretical prediction for the fatality rate and the estimated characteristic hospital density. **(a)** The theoretical prediction, Eq. (14), for the fatality rate ϕ as a function of the hospital density η for selected districts having $\tilde{\eta} = 1.2 \times 10^{-1}, 7.3 \times 10^{-3}$, and $3.5 \times 10^{-4} \text{ km}^{-2}$, respectively. Filled points represent the real data for each district. **(b)** Plot of $\tilde{\eta}$ versus the patient density ρ . Open circles and filled squares indicate the real data and the average, respectively. The solid line fits the average of $\tilde{\eta}$ as a function of ρ and the slope is 0.75 ± 0.056 .

$$\tilde{\eta}_i = \left(\frac{c \tau_i}{\log \tau_i} a_i^2 \right)^{-1} = \left(\frac{c \ell_{(TB)} a_i}{\log \left(\frac{\ell_{(TB)}}{a_i} \right)} \right)^{-1} \tag{15}$$

Assuming the validity of Eq. (14), one can estimate the characteristic hospital density $\tilde{\eta}_i$ by using the empirical data of the fatality rate ϕ_i and the hospital density η_i in Eq. (14) as

$$\tilde{\eta}_i = \frac{\eta_i}{|\log \phi_i|} \tag{16}$$

The exponential functions $\phi_i(\eta)$'s in Eq. (14) with the estimated coefficient $\tilde{\eta}_i$ for selected districts are shown in Fig. 2(a). $\tilde{\eta}_i$ is different from district to district, growing with the patient density [Fig. 2(b)], which underlies the weaker decay of the fatality rate with the hospital density [Fig. 1(b)] than would be expected if $\tilde{\eta}_i$ were identical for all districts. The estimated $\tilde{\eta}_i$ is the characteristic constant of each district and will be used throughout the optimization computation.

The lattice constant can be obtained by using the estimated $\tilde{\eta}_i$ in Eq. (15) and solving for a_i . Let us denote the solution by $a_i^{(F)}(\ell_{(TB)})$. To validate it, we compare it with another estimate independent of the empirical values of the fatality rate or the hospital density. We use the data of the number of business buildings B_i in each district²⁰. The business buildings, including hospitals, are the candidates for the notable places included in the coarse-grained trajectories. The typical distance between adjacent business buildings can be a candidate for the lattice constant, which is given by

$$a_i^{(B)} = \sqrt{\frac{A_i}{B_i}} \tag{17}$$

under the assumption that the business buildings are uniformly distributed in each district. For the comparison of $a^{(F)}(\ell_{(TB)})$ and $a_i^{(B)}$, we take the value of $\ell_{(TB)}$ minimizing the average logarithmic distance $\nu(a^{(F)}, a^{(B)}) = \sum_i (\log a_i^{(F)} - \log a_i^{(B)})^2 / \sum_i 1$, which is $\ell_{(TB)}^* = 10000 \pm 1000$ [Fig. 3(a)]. It corresponds to the annual traveling distance 3300 km which is reasonably close to the empirical value 8478 km of Korea³⁴. In Fig. 3(a), the two lattice constants $a^{(F)} = a^{(F)}(\ell_{(TB)}^*)$ and $a^{(B)}$ show good agreement in their magnitudes, supporting the validity of the assumptions of our model and its formulas, Eqs. (14) and (15). Due to this agreement and Eq. (17), we can see that a large or small value of $a_i^{(F)}$ originates from the sparse or dense business buildings in district i .

With $a_i^{(F)}$, the dimensionless hospital density λ_i and the number of steps τ_i taken for $t_{(TB)}$ can be evaluated by Eqs. (12) and (13), which are plotted versus the patient population density in Fig. 3(b,c), respectively. In contrast to the real hospital density η_i , λ_i is lower in a district with higher patient density [Fig. 3(b)]. It is attributed to the smaller lattice constants in the districts of higher patient densities, arising from the denser buildings. On the other hand, the number of steps taken for $t_{(TB)}$ increases as the patient density increases, increasing the chance to visit hospitals. To sum up, effectively less hospitals are distributed but the patients take more steps for the given period $t_{(TB)}$ in the higher-populated districts, which explains the slightly lower fatality rates therein than in lower-populated districts as shown in Fig. 1(b).

Optimal hospital density. The fatality rate formula in Eq. (14) is applicable to I_s districts having non-zero η and ϕ in the empirical data. Then one can minimize the total fatalities in those I_s districts

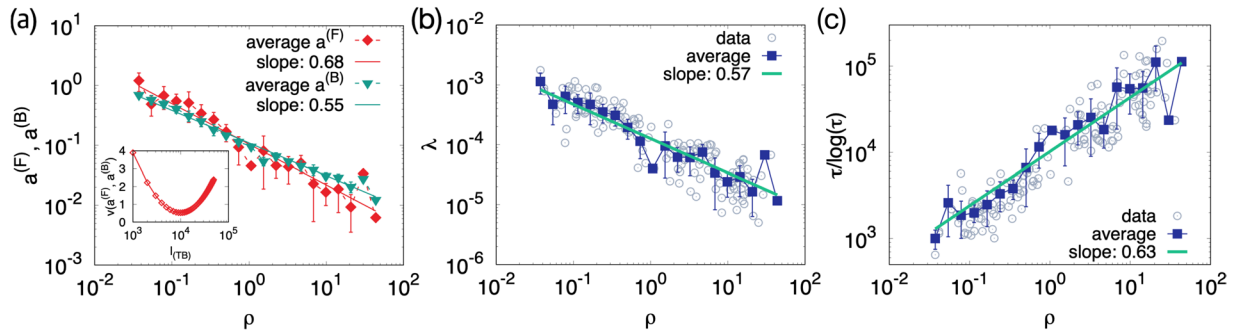


Figure 3. Lattice constant and dimensionless quantities. **(a)** Lattice constants $a^{(F)} = a^{(F)}(\ell_{(TB)}^*)$ and $a^{(B)}$ as functions of the patient density ρ in logarithmic scales. The errorbars are standard deviations. The fitting lines have slopes -0.68 ± 0.044 and -0.55 ± 0.019 , respectively. Inset: The average logarithmic distance ν is minimized at $\ell_{(TB)}^* = 10000$ km with errorbar 1000 km. **(b)** Dimensionless hospital density $\lambda = \eta^2$ [Eq. (12)] versus patient density. The dashed line with filled squares represents the average values with the errorbars being standard deviations. The solid line fits the average values and has slope -0.57 ± 0.050 . **(c)** Plot of $\frac{\tau}{\log \tau}$ versus patient density. The dashed line with filled squares and errorbars represent the average values and standard deviations. The slope of the solid line is 0.63 ± 0.046 .

$$E_{\text{fatalities}} = \sum_{i=1}^{I_s} N_i \phi_i = \sum_i A_i \rho_i \exp\left(-\frac{\eta_i}{\tilde{\eta}_i}\right), \tag{18}$$

with respect to the hospital density distribution $\{\eta_i\}$ for fixed $N_i, \tilde{\eta}_i$, and total number of hospitals H_{total} . In the data of year 2014, $I_s = 143, H_{\text{total}} = 328$, and $E_{\text{fatalities}} = 1718^{20}$. We allow H_i 's to be arbitrary real numbers, and the case of integer H_i 's will be discussed later. $E_{\text{fatalities}}$ in Eq. (18) is minimized when $\delta E = \sum_i A_i \delta \eta_i \left(-\frac{\rho_i}{\tilde{\eta}_i} e^{-\frac{\eta_i}{\tilde{\eta}_i}} + z\right) = 0$ and $\partial^2 E_{\text{fatalities}} / \partial \eta_i^2 > 0$ with z being the Lagrange multiplier. Consequently the optimal hospital density is found to be

$$\eta_i^{(\text{opt})} = \tilde{\eta}_i \log\left(\frac{\rho_i}{z \tilde{\eta}_i}\right), \tag{19}$$

and the optimal fatality rate is

$$\phi_i^{(\text{opt})} = z \frac{\tilde{\eta}_i}{\rho_i}. \tag{20}$$

The Lagrange multiplier z is computed by inserting Eq. (19) into Eq. (6) as $z = \exp\left[\frac{\sum_i A_i \tilde{\eta}_i \log\left(\frac{\rho_i}{\tilde{\eta}_i}\right) - H^{(\text{total})}}{\sum_i A_i \tilde{\eta}_i}\right] \simeq 12.9$

We call the hospital density in Eq. (19) *optimal* in a sense that it minimizes the objective function given in Eq. (18). If the objective function is changed, the optimal density may be changed, which is discussed in the Summary and Discussion section and in the Supplementary Information (SI).

Equations (19) and (20) are the main results of the present study. Remarkably the optimal hospital density and the patient density are rescaled commonly by $\tilde{\eta}_i$ and then related to each other logarithmically. In Fig. 4(a), the arrangement of the data points of the optimal hospital densities on a straight line is contrasted with the scattered distribution of the current (empirical) hospital densities in the $(\rho/\tilde{\eta}_i, \eta/\tilde{\eta}_i)$ plane in semi-logarithmic scale. The same phenomenon is observed for the fatality rate; the empirical fatality rates ϕ_i 's are scattered but the optimized fatality rates lie on a straight line in the $(\rho/\tilde{\eta}_i, \phi)$ plane in logarithmic scale as shown in Fig. 4(b). The rescaled patient density ranges between 30.80 (Yeonggwang-gun) and 2646 (Songpa-gu), and is larger than $z \simeq 12.9$ and thus guarantees $\eta_i^{(\text{opt})} > 0$ for all i in Eq. (19). More plots of the optimized hospital densities and fatality rates are given in Fig. S2.

The scattered distributions of the empirical data in Fig. 4 show clearly the deviation of the current distribution of hospitals from the optimum minimizing the total fatalities of TB. The minimized total fatalities $E_{\text{fatalities}}^{(\text{min})}$ obtained from the optimal hospital distribution is

$$E_{\text{fatalities}}^{(\text{min})} = \sum_i N_i \frac{z \tilde{\eta}_i}{\rho_i} = z \sum_i A_i \tilde{\eta}_i \simeq 1488.44, \tag{21}$$

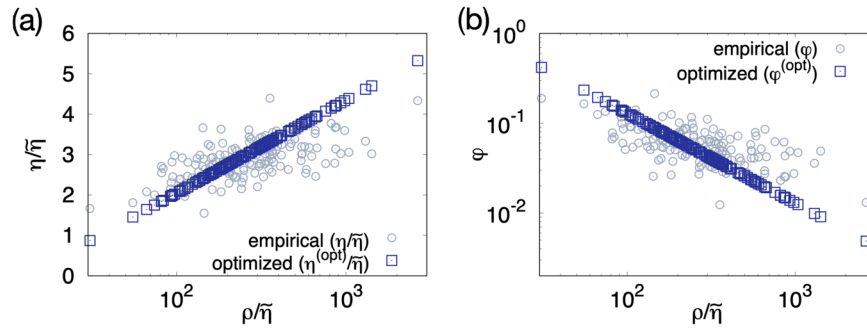


Figure 4. The rescaled hospital density and fatality rate before and after optimization as functions of the rescaled patient density. (a) Plots of the rescaled hospital density, $\frac{\eta}{\bar{\eta}}$ (circle) and $\frac{\eta^{(opt)}}{\bar{\eta}}$ (square) versus the rescaled patient density $\frac{\rho}{\bar{\eta}}$ in semi-logarithmic scale. $\bar{\eta}$ is the characteristic hospital density estimated empirically as Eq. (16). The data points for the optimized hospital density lie on the line corresponding to Eq. (19) with $z = 12.9$. (b) Plots of the fatality rate ϕ (circle) and $\phi^{(opt)}$ (square) versus the rescaled patient density $\frac{\rho}{\bar{\eta}}$. The data points for the optimal fatality rates are on the line corresponding to Eq. (20).

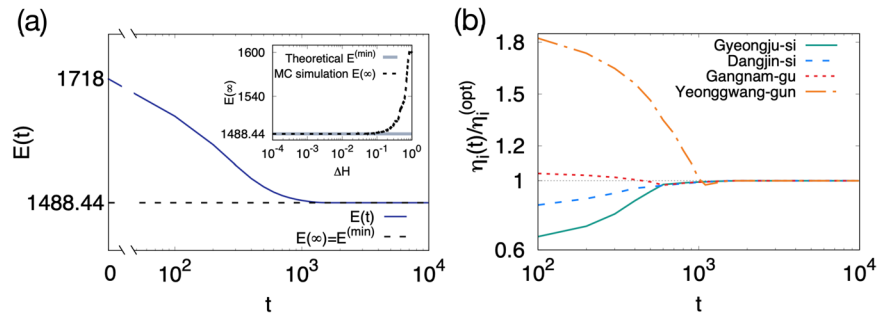


Figure 5. Monte-Carlo (MC) simulation for optimizing the hospital distribution. (a) The total energy $E(t) = E_{\text{fatalities}}$ as a function of the MC step t in the MC simulation with $\Delta H = 10^{-2}$. It becomes stationary at $E(\infty) = E^{(\min)} = 1488.44$ for $t \gtrsim 10^3$. Inset: The stationary-state value $E(\infty)$ depends on the increment ΔH . (b) The ratio $\frac{\eta_i(t)}{\eta_i^{(opt)}}$ is plotted as a function of the MC step t for selected districts with $\Delta H = 10^{-2}$. It converges to one for $t \gtrsim 10^3$.

which is smaller than the current value, 1718, by 13%. For this optimization, $\sum_i A_i |\eta_i^{(opt)} - \eta_i|/2 \simeq 24.4$ hospitals among a total of $H_{\text{total}} = 328$ are relocated. In the zero-temperature Monte Carlo (MC) simulation in which some small amount ΔH of hospitals are moved between randomly selected districts only when the attempted relocation reduces $E_{\text{fatalities}}$, the theoretical predictions in Eqs. (19) and (21) are realized in the steady state as long as ΔH is sufficiently small [Fig. 5(a)]. If H_i 's are restricted to be integers ($\Delta H = 1$), the total fatalities in the steady state is 1599.45. The stationary-state results remain unchanged in the simulations with different initial configurations or with gradually cooling down the temperature. For more details of the simulations, see Methods.

To achieve such reduction in the total fatalities, the hospital density should be increased in some districts and decreased in others [Fig. 5(b)]. For instance, Gyeongju-si should have its hospital density 1.61 times larger than the current hospital density but Yeonggwang-gun 0.534 times larger than the current one. In Fig. 6(a), 143 districts are colored blue (red) if the optimal hospital density is larger (smaller) than the current hospital density. Interestingly, the ratio $\frac{\eta_i^{(opt)}}{\eta_i}$ of the optimal to current hospital density turns out to depend strongly on the rescaled patient density $\frac{\rho}{\bar{\eta}}$ [Fig. 6(b)]. It implies that if the rescaled patient density is large in a district and small in another district, it is recommended to move some hospitals from the latter to the former district. The high and low values of $\frac{\eta_i^{(opt)}}{\eta_i}$ of Gyeongju-si and Yeonggwang-gun can be understood in this line, as they have quite different values of $\frac{\rho}{\bar{\eta}}$, 662 and 30.8, respectively. Such a significant correlation is absent between $\frac{\eta_i^{(opt)}}{\eta_i}$ and the raw patient density ρ [Fig. S3]. The change of the fatality rate shows the opposite trend to that of the hospital density; The districts with large (small) rescaled patient density $\frac{\rho}{\bar{\eta}}$ have their fatality rate decreased (increased) as they gain (lose) hospitals [Fig. 6(c)]. It is remarkable that the changes of the hospital density and the fatality rate are determined not by the patient density or the population density but the rescaled patient density.

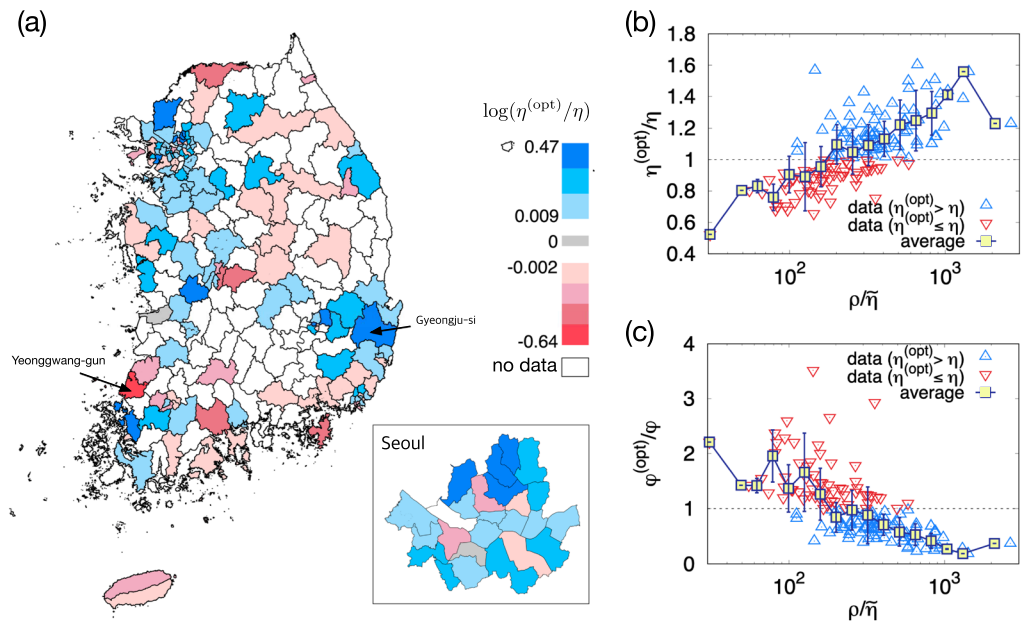


Figure 6. Changes of the hospital density by optimization. **(a)** The logarithmic ratio of the optimal to current hospital density $\log\left(\frac{\eta^{(opt)}}{\eta}\right)$ is encoded by color for each district of Korea. 85 districts are white, as they do not have both $\eta^{(opt)}$ and η available. **(b)** Plot of the ratio $\frac{\eta^{(opt)}}{\eta}$ versus the rescaled patient density $\frac{\rho}{\bar{\eta}}$. Upper and lower triangles are used for the data points with the optimal hospital density larger and smaller, respectively, than the current one. The filled square is the average and the errorbar is the standard deviation of the ratio $\frac{\eta^{(opt)}}{\eta}$. **(c)** Plot of the ratio of the optimal to current fatality rate $\frac{\phi^{(opt)}}{\phi}$ versus $\frac{\rho}{\bar{\eta}}$.

Summary and Discussion

We have here proposed a modeling framework which predicts the optimal distribution under a general objective function, going beyond the previous descriptive explanations for the facility distribution. In deriving the optimal hospital distribution over districts for minimizing the TB fatalities in the whole country, we have found that the characteristic hospital density of each district plays an important role in the optimization. The random-walk nature of the coarse-grained trajectories of patients has been assumed in establishing a theoretical model, and the lattice constant of each district has been introduced to connect the theoretical results and the empirical data. The incorporation of such heterogeneity of districts in the theoretical study of the facility optimization is done only in the present study and can be useful in future studies.

Examining the assumptions and limitations of the proposed model may help better understand and improve its predictive power. The exponential decay of the fatality rate with the hospital density given in Eq. (14) is valid when the dimensionless hospital density is low, $\frac{\sqrt{\lambda\tau}}{\log\tau} \ll 1^{26}$. The empirical data analyzed in the present study stay in this regime; $\frac{\sqrt{\lambda\tau}}{\log\tau}$ ranges between 0.23 and 0.37. If we were to extend to the case of high hospital density or the hospital locations being no more independent of one another, the fatality rate might behave differently from Eq. (14). For $\phi_i = f(\eta_i/\bar{\eta}_i)$ with $f(x)$ a decreasing convex function such as exponential, stretched exponential, or power law, the optimal hospital density will be given by $\frac{\eta_i}{\bar{\eta}_i} = (-f')^{-1}\left(\frac{\rho_i}{\bar{\rho}_i}\right)$ with $(f')^{-1}$ being the inverse of the derivative of $f(x)$. The relation between the two rescaled variables $\frac{\eta_i}{\bar{\eta}_i}$ and $\frac{\rho_i}{\bar{\rho}_i}$ depends on the specific form of $f(x)$ and reduces to Eq. (19) in case of $f(x) = e^{-x}$. The case of $f(x)$ being a power law is presented in the SI. Regarding the robustness of the functional form of the fatality rate, it will be of interest to investigate which model of walk with traps exhibits such a power-law survival probability.

We have counted only the private general hospitals, but there are mostly found one or two public health centers in a district, which can also provide the medical treatment to TB patients although its portion may not be large³⁵. One can optimize the private hospital distribution considering the contribution of the public health centers to the TB treatment, which is presented in the SI and Fig. S4. The results remain unchanged qualitatively. Extending the trajectories of patients to the nearby districts can be one way of making the model more realistic, which will address how the similarity or dissimilarity of adjacent districts may affect the fatality rate and the total fatalities. A correlation may be expected between the economic level of a district and the fatality rate of a disease, which is not significant in case of TB in Korea [Fig. S5], as the TB treatment is covered by the national medical insurance in Korea³⁶, but should be considered in the application to other diseases or other countries. When a given number of hospitals can be opened or should be shut down for financial or other reasons, our results will be helpful for the investigation of the optimal locations.

Methods

Data-sets. Under the control of the Ministry of Health and Welfare, several organizations such as the Korean National Tuberculosis Association and Korea Centers for Disease Control and Prevention cooperate to prevent and eradicate tuberculosis in Korea³⁶. The related information has been well recorded, which is accessible through Statistics Korea²⁰. The data-sets used in the present study have been collected district by district. As a result, 37347 new TB patients, 330 hospitals, and 2127 dead patients for 228 districts in year 2014 have been considered in the present study. The hospitals considered in our study are the private ones classified as general or superior general hospitals in the Korean Medical Service Act. The theoretical modeling for the fatality rate applies for $I_s = 143$ districts which have at least one hospital and at least one dead patient. The total number of new and dead patients, and hospitals in those 143 districts are 32322, 1718, and 328.

Monte carlo simulation. To illustrate the hospital relocation process, we perform the zero-temperature Monte Carlo (MC) simulation in which hospitals are relocated over I_s districts towards decreasing the energy, equal to the total fatalities given in Eq. (21), as follows:

- (i) Initially the number of hospitals in each district is set equal to the empirical data.
- (ii) For two randomly selected districts i and j , consider moving ΔH hospitals from i to j as long as $H_i - \Delta H > 0$.
- (iii) Accept this relocation if the energy change $\Delta E = N_i \left(e^{-\frac{H_i - \Delta H}{A_i \rho_i}} - e^{-\frac{H_i}{A_i \rho_i}} \right) + N_j \left(e^{-\frac{H_j + \Delta H}{A_j \rho_j}} - e^{-\frac{H_j}{A_j \rho_j}} \right)$ is zero or negative. Reject it otherwise.
- (iv) Repeat steps (ii) and (iii) I_s times to increase the MC step t by one.

We find that the energy becomes stationary around $t = 10^3$ MC steps and thus we run the simulations just up to 10^4 MC steps [Fig. 5(b)]. ΔH represents the amount of hospitals moved by one relocation. For $\Delta H \lesssim 0.05$, the hospital configuration and the energy in the stationary state coincide with the theoretical predictions in Eqs. (19) and (21), respectively. Replacing the initial hospital configuration by a random one, the energy and the hospital configuration in the stationary-state are not changed but remain the same as the theoretical prediction. Since a hospital relocation is accepted only when the corresponding energy change is not positive, this simulation corresponds to zero temperature $T = 0$. We have also run the MC simulation with lowering temperature gradually from $T = 20$ to $T = 3 \times 10^{-9}$ but the hospital configuration and the energy in the stationary state are found to be the same as those of the zero-temperature MC simulation.

Data availability

The datasets generated during the current study are available from the corresponding author on reasonable request.

Received: 3 March 2020; Accepted: 1 May 2020;

Published online: 25 May 2020

References

1. Youn, H., Gastner, M. T. & Jeong, H. Price of anarchy in transportation networks: Efficiency and optimality control. *Phys. Rev. Lett.* **101**, 128701 (2008).
2. Wuellner, D. R., Roy, S. & D'Souza, R. M. Resilience and rewiring of the passenger airline networks in the United States. *Phys. Rev. E* **82**, 056101 (2010).
3. Withaut, D. & Timme, M. Braess's paradox in oscillator networks, desynchronization and power outage. *New Journal of Physics* **14**, 083036 (2012).
4. Lee, M. J. & Kim, B. J. Spatial uniformity in the power-grid system. *Phys. Rev. E* **95**, 042316 (2017).
5. Motter, A. E., Gulbahce, N., Almaas, E. & Barabási, A.-L. Predicting synthetic rescues in metabolic networks. *Molecular Systems Biology* **4**, 168 (2008).
6. Owen, S. H. & Daskin, M. S. Strategic facility location: A review. *European Journal of Operational Research* **111**, 423–447 (1998).
7. Hodgart, R. L. Optimizing access to public services: a review of problems, models and methods of locating central facilities. *Progress in Human Geography* **2**, 17–48 (1978).
8. Calvo, A. B. & Marks, D. H. Location of health care facilities: An analytical approach. *Socio-Economic Planning Sciences* **7**, 407–422 (1973).
9. Megiddo, N. & Supowit, K. J. On the complexity of some common geometric location problems. *SIAM Journal on Computing* **13**, 182–196 (1984).
10. Current, J., Min, H. & Schilling, D. Multiobjective analysis of facility location decisions. *European Journal of Operational Research* **49**, 295–307 (1990).
11. Stephan, G. E. Territorial division: The least-time constraint behind the formation of subnational boundaries. *Science* **196**, 523–524 (1977).
12. Gusein-Zade, S. M. Bunge's problem in central place theory and its generalizations. *Geographical Analysis* **14**, 246–252 (1982).
13. Gastner, M. T. & Newman, M. E. J. Optimal design of spatial distribution networks. *Phys. Rev. E* **74**, 016117 (2006).
14. Um, J., Son, S.-W., Lee, S.-I., Jeong, H. & Kim, B. J. Scaling laws between population and facility densities. *Proceedings of the National Academy of Sciences* **106**, 14236–14240 (2009).
15. World Health Organization. *Global tuberculosis report 2017* (World Health Organization, 2017).
16. Glaziou, P., Sismanidis, C., Floyd, K. & Raviglione, M. Global epidemiology of tuberculosis. *Cold Spring Harbor perspectives in Medicine* (2014).
17. Sook, C. K. Tuberculosis control in the republic of Korea. *Epidemiol Health* **40**, e2018036–0 (2018).
18. Centers for Disease Control and Prevention. *Core Curriculum on Tuberculosis: What the Clinician Should Know* (Centers for Disease Control and Prevention, 2013).

19. Hopewell, P. C., Pai, M., Maher, D., Uplekar, M. & Raviglione, M. C. International standards for tuberculosis care. *The Lancet Infectious Diseases* **6**, 710–725 (2006).
20. Korean Statistical Information Service. Available: <http://kosis.kr> (Accessed 1.20.2018).
21. Pastor-Satorras, R. & Vespignani, A. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* **86**, 3200–3203 (2001).
22. Lee, M. J. & Lee, D.-S. Understanding the temporal pattern of spreading in heterogeneous networks: Theory of the mean infection time. *Phys. Rev. E* **99**, 032309 (2019).
23. Tiemersma, E. W., van der Werf, M. J., Borgdorff, M. W., Williams, B. G. & Nagelkerke, N. J. D. Natural history of tuberculosis: Duration and fatality of untreated pulmonary tuberculosis in HIV negative patients: A systematic review. *PLoS One* **6**, 1–13 (2011).
24. Hughes, B. *Random Walks and Random Environments*. (Oxford Univ. Press, Clarendon, 1995).
25. Rosenstock, H. B. Random walks on lattices with traps. *Journal of Mathematical Physics* **11**, 487–490 (1970).
26. Barkema, G. T., Biswas, P. & van Beijeren, H. Diffusion with random distribution of static traps. *Phys. Rev. Lett.* **87**, 170601 (2001).
27. Donsker, M. & Varadhan, S. Asymptotic evaluation of certain markov process expectations for large time, i. *Communications on Pure and Applied Mathematics* **28**, 1–47 (1975).
28. Grassberger, P. & Procaccia, I. The long time properties of diffusion in a medium with static traps. *The Journal of Chemical Physics* **77**, 6281–6284 (1982).
29. Brockmann, D., Hufnagel, L. & Geisel, T. The scaling laws of human travel. *Nature* **439**, 462–465 (2006).
30. González, M. C., Hidalgo, C. A. & Barabási, A.-L. Understanding individual human mobility patterns. *Nature* **453**, 779–782 (2008).
31. Song, C., Qu, Z., Blumm, N. & Barabási, A.-L. Limits of predictability in human mobility. *Science* **327**, 1018–1021 (2010).
32. Choi, J., Sohn, J.-I., Goh, K.-I. & Kim, I.-M. Modeling the mobility with memory. *EPL (Europhysics Letters)* **99**, 50001 (2012).
33. Kim, K., Kyoung, J. & Lee, D.-S. Self-attracting walk on heterogeneous networks. *Phys. Rev. E* **93**, 052310 (2016).
34. Korea Transport Database. *Korea Transport Database Newsletter* **12** (Korea Transport Database, 2013).
35. Lee, Y., Kwon, Y., Lee, S., Sohn, H. & Koh, Y. Overview of tuberculosis control and prevention policies in Korea. *Public Health Wkly Rep* **8**, 651–656 (2015).
36. Ministry of Health and Welfare. Available: <http://www.mohw.go.kr>.

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grants funded by the Korean Government [No. 2019R1A2C1003486 (DSL) and No. 2018R1A6A3A01010866 (MJL)].

Author contributions

M.J.L. analyzed the data-sets, developed the theory, performed the simulations, and wrote the manuscript. K.K. and J.S. analyzed the data-sets. D.-S.L. designed and supervised the research, and wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-65337-x>.

Correspondence and requests for materials should be addressed to D.-S.L.

Reprints and permissions information is available at www.nature.com/reprints.

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) 2020