as 'regions', each with a *repertoire* of actions that may be complex in and of itself. However, the number of diseases or regions would typically be small (fewer than 10), and the propagation of treatment effects across and among them could be very complex. This would require additional methodology to capture effectively. There are important 'art of the possible' considerations in this domain that must be addressed as well. The output of any system for the multimorbidity treatment problem can only be decision *support* rather than decision making, which in part means that uncertainty over different courses of action must be exposed in a way that end-users understand. This might be accomplished by using the authors' framework and somehow conveying the posterior distribution over optimal allocations. Furthermore, any recommended treatment may or may not be followed—an issue alluded to by the authors in their discussion of a 'compliance model'. I am interested to know the authors' opinions on their work's relevance to other such sequential decision-making problems with complex action spaces.

The following contributions were received in writing after the meeting.

Anna L. Choi (*Chinese University of Hong Kong in Shenzhen and Shenzhen Research Institute of Big Data*) **and Tze Leung Lai** (*Stanford University*)

Section 2 of this interesting paper gives a comprehensive review of white nose syndrome in north American bats. Section 3 formulates the authors' proposal to control the spread of white nose syndrome by using optimal treatment allocation. For each time point and each location, the decision is to 'apply a treatment or to do nothing'. It considers only one treatment with unknown treatment effects. However, there are already many treatments that are yet to be tested (Cornelison *et al.*, 2014; Hoyt *et al.*, 2015), and the (economic) cost of the treatment(s) should also be considered. In this connection, a relatively economical and ease-to-use new treatment has just emerged at the beginning of 2018. *Science Daily*, January 2nd, 2018, announced that

> 'scientists with the USDA Forest Service and the University of New Hampshire have found what may be an Achilles' heel in the fungus that causes white-nose syndrome: UV-light',

citing a study by Palmer *et al.* (2018). This adds new challenges to the optimal treatment allocation problem as new treatments can enter the treatment pool during the course of the study.

We next comment on the more general problem of optimal allocation of a (single) treatment 'over a countably infinite set of treatment periods and a finite number of locations'. The solution is described as 'estimating an optimal allocation strategy' in Section 4, following the framework of optimal dynamic regimes in the statistics literature. In control engineering, there is a counterpart called *stochastic adaptive control*; in computer science, the counterpart is *reinforcement learning*. The fascinating sequential treatment allocation problem over a dynamic network system, with covariate and outcome information from relevant nodes of the network, considered in this paper also arises in many other interdisciplinary applications on which we are writing a monograph (Choi *et al.*, 2019). In this connection, we want to point out that the authors' remarks near the end of Section 7 fall under the framework of *contextual bandits*, for which it is shown that Thompson sampling is not optimal because it is myopic whereas a much simpler ϵ-greedy strategy can be shown to be asymptotically optimal.

Dean Eckles (*Massachusetts Institute of Technology, Cambridge*) **and Maurits Kaptein** (*Tilburg University*)

Laber and his colleagues impressively model intervening to prevent epidemic spread as a multiarmed bandit problem, using Thompson sampling to add exploration to the choice of sites to treat. We wish to highlight how

(a) the method could be made more robust by use of a bootstrap and
(b) how less exploration will often be preferable with such a short horizon.

The authors approximate Thompson sampling by using a plug-in estimator of the sampling distribution of the maximum likelihood estimates (Section 5.1). One could further depart from trying to approximate the posterior of the posited parametric model by using a non-parametric bootstrap distribution (see Newton and Raftery (1994) and Efron (2012)). Bootstrap Thompson sampling (Eckles and Kaptein, 2014) samples from replicates formed by an on-line bootstrap, which can be easily parallelized or distributed, can account for dependent observations and is more robust to common forms of model misspecification. This method and related greedier variations have been used in multiple application areas and compared favourably with other methods (e.g. Agarwal *et al.* (2014), Osband *et al.* (2016), Lu and Van Roy (2017) and Bietti *et al.*

(2018)). In the context of the present application, this could enable the authors to avoid Markov chain Monte Carlo sampling altogether while being robust to, for example, heteroscedastic errors or within-region dependence.

We have already mentioned the popularity of greedier variations of bootstrap Thompson sampling, More generally, recent work has demonstrated that methods that explore less (i.e. are greedier) perform extremely well in contextual multiarmed bandit problems (Chapelle and Li, 2011; May *et al.*, 2012; Bastani *et al.*, 2018; Bietti *et al.*, 2018) with common characteristics. First, in many applications the horizon T is small; in the present application, actions are taken for eight periods only. With such a small horizon, it is not obvious that substantial exploration is preferable—at least for minimizing regret. In contrast, some other recent results in favour of exploration-free methods based on the distribution of context (e.g. Bastani *et al.* (2018)) do not obviously apply to the present setting with dependence due to contagion and with combinatorial actions. Nonetheless, we expect that a policy that is more greedy than the presented implementation of Thompson sampling would lead to an even smaller proportion of infections.

The best reason for introducing exploration may not be minimizing regret up to the horizon. Rather, introducing relatively low cost exploration might be motivated by the desire to make reuse of collected data possible for other purposes. In standard contextual multiarmed bandit and dynamic treatment regime settings, reuse of such data for evaluation of other polices is routine (Murphy *et al.*, 2001; Dudík *et al.*, 2014; Agarwal *et al.*, 2016). It would be interesting to see how readily data from a setting with dependence induced by contagion could be reused.

Seongho Kim (*Wayne State University, Detroit*) **and Weng Kee Wong** (*University of California at Los Angeles*)

We congratulate the authors on their detailed and interesting work in this important area of research.

Assumptions 1 and 2 would seem to imply that $\mathbf{A}^t \perp \mathbf{W}^t | \mathbf{H}^t$ for all $t \in T$ and so $\mathbf{A}^t$ is independent of $\mathbf{Y}^t$, $\mathbf{S}^{t+1}$ given $\mathbf{H}^t$ for all $t \in T$. Further, by the Markov homogeneity assumption,

(a) the observed data $\mathbf{S}^t$ and allocation $\mathbf{A}^t$ at time t are sufficient to predict the outcome at time t, $\mathbf{Y}^t$, and
(b) the data at time t, $\mathbf{S}^t$, depend only on the observed data, $\mathbf{S}^{t-1}$, and the allocation $\mathbf{A}^{t-1}$ at time $t-1$.

Fig. 12 depicts this time-dependent homogeneous Markov system but suggests that the conditional independence assumption between $\mathbf{A}^t$ and $(\mathbf{Y}^t, \mathbf{S}^{t+1})$ given $\mathbf{H}^t$ may be questionable because of the edges with asterisks and, if so, will appear to contradict assumptions 1 and 2. To avoid this discordance, it seems that either assumption 1 may be eliminated or needs to be modified to $\mathbf{Y}^{*t} \perp \mathbf{S}^{*t+1} | (\mathbf{A}^t, \mathbf{H}^t)$ for all $t \in T$.

At March 14th, 2018, white nose syndrome (WNS) had spread to Washington, but not to Florida (`https://www.whitenosesyndrome.org/sites/default/files/wnsspreadmap_3_14_2018.jpg`). Florida, near to the contaminated regions, has not reported WNS yet, whereas Washington has been infected without neighbouring infected regions. Can the spatial gravity model incorporate such information in equation (2) and, if so, how does the model proposed update the parameters by using a Bayesian framework with the latest WNS data? These are likely to be challenging tasks. It is also possible that symptoms from bats in Florida are temporary because of the relatively high average temperatures

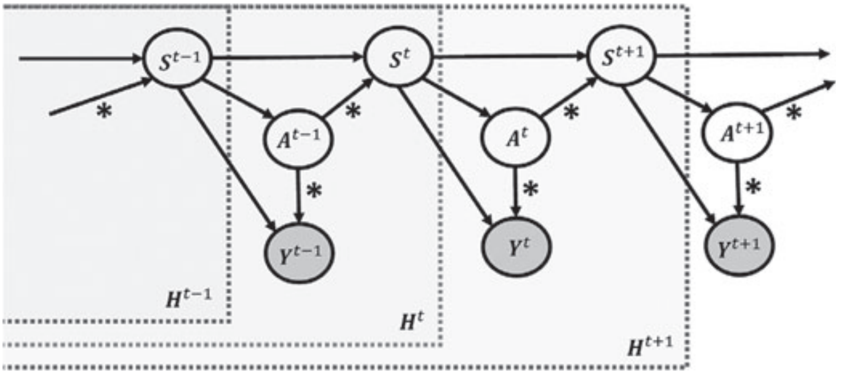

Fig. 12. Graphical representation of the dependence between $\mathbf{S}^t, \mathbf{A}^t, \mathbf{Y}^t$ and $\mathbf{H}^t$ where $t \geqslant 1$, $\mathbf{H}^t = (\mathbf{S}^1, \mathbf{A}^1, \mathbf{Y}^1, \ldots, \mathbf{S}^{t-1}, \mathbf{A}^{t-1}, \mathbf{Y}^{t-1}, \mathbf{S}^t)$ and $\mathbf{H}^1 = \mathbf{S}^1$