Who Do You Know?
A Simulation Study of Infectious Disease Control Through Contact Tracing

Benjamin Armbruster and Margaret L. Brandeau
Department of Management Science and Engineering
Stanford University
Stanford, California 94305-4026, USA
armbruster@stanford.edu, brandeau@stanford.edu

Keywords: contact tracing, infectious disease, network, cost-effectiveness

Abstract

Contact tracing (also known as partner notification) is a primary means of controlling infectious diseases such as tuberculosis (TB), human immunodeficiency virus (HIV), and sexually transmitted diseases (STDs). The effectiveness of contact tracing depends on a number of factors including the contact tracing policy used, the social mixing network, and characteristics of the disease (e.g., the transmission mechanism, variability in infectiousness over time and across individuals, and the rate at which symptoms develop). We develop a simulation model for contact tracing and use it to explore the effectiveness of different contact tracing policies in a budget constrained setting. We evaluate several alternative contact tracing policies. We then introduce a cost-effectiveness framework and show how it can be used to determine the optimal level of investment in contact tracing. We first assume that the only incremental disease control is contact tracing. We then extend the analysis to consider the optimal allocation of a budget between contact tracing and screening for exogenous infections.

INTRODUCTION

Contact tracing (also known as partner notification) is a primary means of disease control for infectious diseases with low prevalence. In the US, contact tracing is required for TB [CDC 2000], recommended for HIV [CDC 2002], and not uncommon for other STDs [Cowan et al. 1996; Clarke 1998]. Contact tracing has also been used (and modeled) for SARS [Lipsitch et al. 2003], foot-and-mouth-disease [Kiss et al. 2005], smallpox [Porco et al. 2004; Kretzschmar et al. 2004], and avian influenza [Wu et al. 2006].

Hyman et al. [2003] and Armbruster and Brandeau [2006] studied contact tracing using differential equation models that assume homogeneous mixing of the population. Kretzschmar [2000] reviewed STD models on networks. Müller et al. [2000] introduced one of the first models of contact tracing on a network and analyzed a stochastic branching process that approximates it. Subsequent work [Huerta and Tsimring 2002; Eames and Keeling 2003; Kiss et al. 2005] analyzed similar models using both stochastic simulations and moment closures (also called mean-field approximations) which lead to ordinary differential equations. Most of these papers study the effectiveness of contact tracing but do not consider the costs. Armbruster and Brandeau [2006] and Wu et al. [2006] incorporated costs in their analyses, but considered contact tracing as an all-or-nothing decision, with a fixed level of intensity.

Empirical studies of the cost effectiveness of contact tracing programs have been carried out for diseases such as TB [Dasgupta et al. 2000; Macintyre et al. 2000], HIV [Varghese et al. 1999; Cohen et al. 2004], chlamydia [Howell et al. 1997], syphilis [Oxman and Doyle 1996], and gonorrhea [Welte et al. 2000]. These studies all evaluate a single fixed level of contact tracing.

In this paper, we develop and apply a simulation model to explore the cost effectiveness of different levels of contact tracing. Using such results, along with a cost-effectiveness threshold, we can determine the optimal level of contact tracing. The following section describes our simulation model for contact tracing on a network. We use the simulation model to compare the effectiveness of several different contact tracing policies. We then simulate the most effective policy using different budgets to determine how much should be spent on contact tracing, assuming that the only incremental disease control is contact tracing. We then extend the analysis to consider the optimal allocation of a budget between contact tracing and screening for exogenous cases of infection. We conclude with discussion of results and directions for future research.

SIMULATION MODEL

We consider a population of n individuals. We model individuals as nodes on an undirected graph where an edge between nodes i and j indicates that i and j can infect each other (we say they are *contacts* of each other).

We assume an SIRS epidemic model with exponential waiting times: susceptible individuals become infectious, become removed when they are treated, and finally become susceptible after treatment. Figure 1 illustrates the state transi-

tions. We assume that the rate of infection (transition from $S \to I$) of node i is proportional to d_i, the number of infected neighbors of node i: specifically, the transition rate is d_i/t_1, where t_1 is a time constant. This stochastic process on a network is called a *contact process*.

To model contact tracing, Eames and Keeling [2003] and Kiss et al. [2005] extend the contact process so that infected nodes are found and cured at a rate proportional to the number of index case neighbors a node has (in our model, this would be individuals in states R), analogous to the infection process. This model of contact tracing does not allow us to compare different contact tracing policies or budgets. Thus, we use a discrete-event simulation.

When an infected individual seeks treatment for symptoms of the disease (and thus becomes known to the public health system), he or she becomes an *index case*. This corresponds to a transition from $I \to R$. We assume that this transition happens at rate $1/t_2$ where t_2 is a time constant. When a new index case occurs, we apply our contact tracing policy to decide (based on only the graph structure and the removed nodes) which nodes to trace. Nodes selected for tracing transition to state ST or state IT, depending on whether the individual is susceptible or infected, respectively. Tracing requires a fixed amount of time, t_4 for state ST and t_5 for state IT.

After tracing is completed, a node in state ST returns to state S while a node in state IT transitions to state R and becomes a new index case. We assume a budget B for contact tracing, expressed in terms of the maximum allowable contact tracing rate: at any point in time, at most B individuals in total can be in states ST or IT.

Individuals can also become infected from exogenous sources. This could be through international travel or by healthy people leaving the system and being replaced by infected immigrants. We assume that the rate at which exogenous infection occurs is given by a constant, η. In our simulation, the sojourn time in each state was exponentially distributed for all states except for states ST and IT, where the sojourn time was a constant.

We used random small-world graphs for our simulation. To generate these graphs we started with a cyclic regular graph of n nodes with degree 4 where node i connects to nodes $i \pm 1, \pm 2 \pmod n$. For every other pair of nodes (i, j) we created a link independently with probability $1/n$. Figure 2 shows a small example of such a network with its nodes in various states.

In subsequent sections we will frequently make statements about the steady-state prevalence of the disease under certain conditions. To measure the steady state we performed hundreds of runs (ranging from 400 to 1600 for different analyses). For each run, we generated a random small-world graph and infected a single random node. Then we simulated the network for five years, taking the daily average prevalence

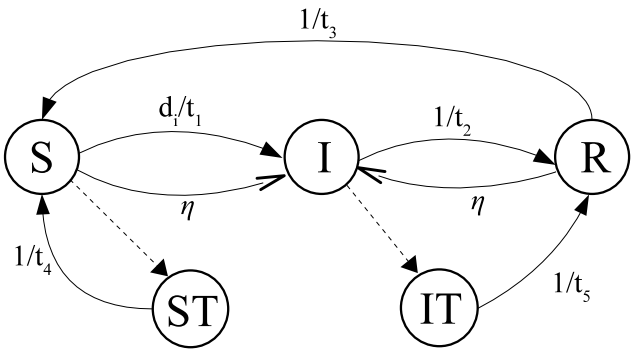

Figure 1. Possible states of an individual and the transition times between them. The dashed arrows mark the instantaneous transitions $S \to ST$ and $I \to IT$ that occur when we decide to trace this individual.

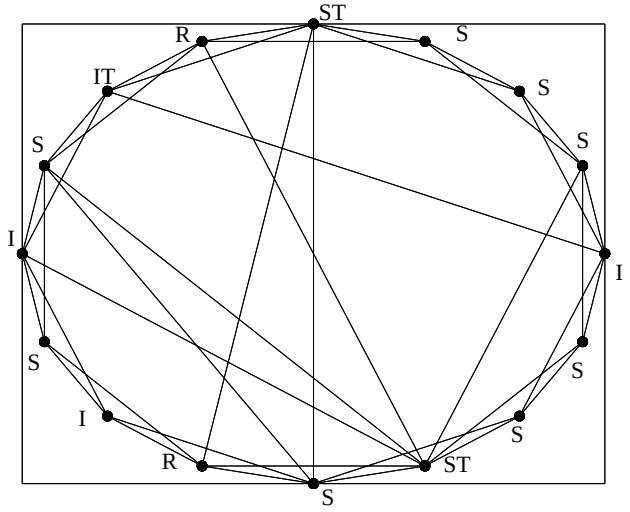

Figure 2. A small-world graph with nodes in various states.

(per capita frequency of states I and IT) starting with day 181 (hoping that the system is in steady state at that time). We averaged over all the runs and set our error bars to the 95% confidence intervals.

Table 1 shows the values of all parameters we used in our simulations.

Table 1. Simulation parameters

n	500	individuals
t_1	90	days
t_2	30	days
t_3	90	days
t_4, t_5	5	days
η	$1/9000$	new cases/day/person

WHO TO TRACE?

In contact tracing, every index case is asked to name his or her contacts (graph neighbors who may be infected). Then public health officials seek out these contacts (as time and resources permit) to test whether they are infected and treat them if so. Who to trace is an important tactical decision since the contact tracing budget limits the number of individuals who can be traced at any point in time.

In our simulation we keep a prioritized list of contacts who have not yet been traced (nodes in state S or I that are neighbors of removed nodes). Every time a new index case is identified, we update this list and decide on additional nodes to trace. We let k be the number of contacts we would like to trace each time a new index case arrives. Since the list is prioritized, we trace the k nodes of highest priority, provided we have not exhausted the budget.

In our simulation, contact tracing policies are parameterized by k, the ranking scheme, and the budget. In this section we focus on the first two parameters while the next section focuses on the budget. Fixing $k = 5$ and $B = 8$, we compared the steady-state disease prevalence achieved with no contact tracing and with contact tracing under three different ranking schemes.

The first policy, Random, is a random ordering. The second and third policies assign each contact a score intended to reflect the likelihood that the contact is infected (the higher the score, the more likely that a contact is infected) and then rank the contacts from highest to lowest score (using a random ordering for ties). In the Most Named policy, a contact's score is the number of index cases who name that person. The List Length policy is motivated by an assumption that each index case only has one infected neighbor. Thus, if an index case has m contacts, it contributes $1/m$ to the score of each of its contacts.

Figure 3 shows that the performance of the three ranking schemes is very similar, with the Most Named policy performing best and the Random policy worst. Simulating under different scenarios with different parameters, the Most Named policy always performed slightly better than the List Length policy, which performed slightly better than the Random policy. The remainder of the paper uses the Most Named policy to prioritize contacts.

Figure 4 examines the effect of varying k, the number of contacts we trace each time a new index case arrives (assuming we still have resources). If k is too large, all of our resources may be utilized when the next index case arrives. If k is too small, our resources are not fully utilized. Figure 4 shows (for a budget of 8) that the steady-state prevalence decreases as we increase k to 5 and then becomes insensitive to further increases in k. In the remainder of this paper we set $k = 5$.

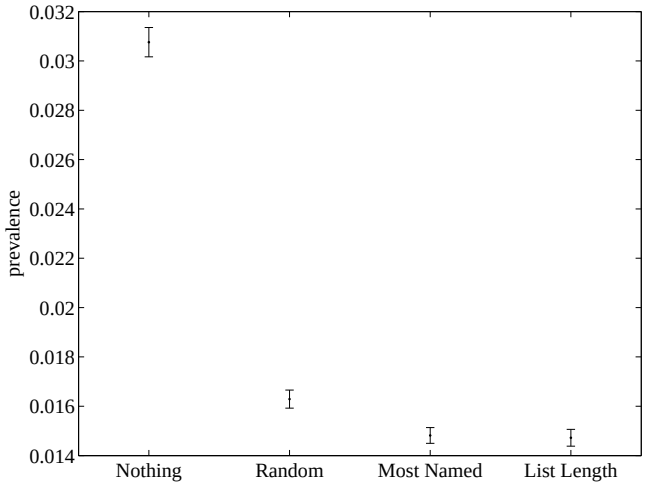

Figure 3. Steady-state prevalence with no contact tracing and contact tracing using policies Random, Most Named, and List Length. The simulations assumed $k = 5$, $B = 8$.

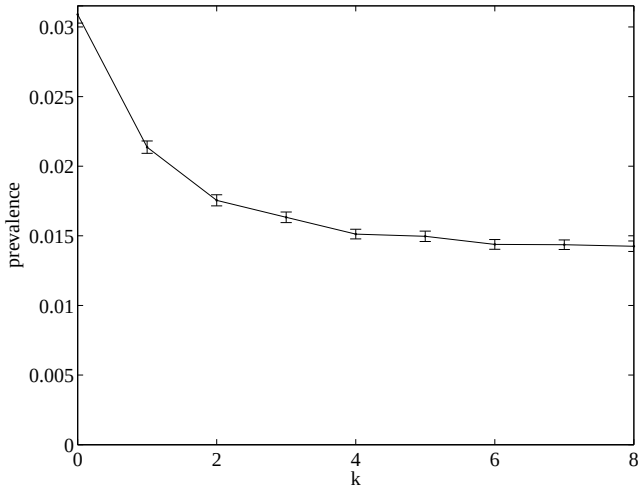

Figure 4. Steady-state prevalence as a function of k, the number of contacts we trace each time a new index case arrives. The simulations used the Most Named policy with $B = 8$.

HOW MANY TO TRACE?

Choosing the budget for contact tracing is an important strategic decision. Funds not spent to trace a particular disease could be used for tracing other diseases, for other disease control efforts, or even for other public health efforts. Thus, we would like to determine the most "cost effective" level of investment in contact tracing for a particular disease.

Cost-effectiveness analysis is a tool that can help policy makers allocate money across different interventions for the same or different diseases. Suppose we have m different programs, and allocating b_i to program i produces a benefit of $f_i(b_i)$. The cost effectiveness of program i (at budget level

b_i) is the incremental cost divided by the incremental benefit $1/f_i'(b_i)$ of increasing its budget above b_i. Preferred programs are those with a small cost-effectiveness ratio (that is, a small cost per benefit achieved). We want to select investments so as to maximize the total benefit subject to a budget of B:

$$\begin{aligned} &\max_{(b_i)} \ f_1(b_1) + \cdots + f_m(b_m) \\ &s.t. \ b_1 + \cdots + b_m \leq B. \end{aligned} \tag{1}$$

Assuming that the functions f_i are concave, this is equivalent to minimizing the cost effectiveness of the least cost effective programs:

$$\begin{aligned} \alpha = \min_{(b_i)} \max_i \ & 1/f_i'(b_i) \\ s.t. \ & b_1 + \cdots + b_m \leq B. \end{aligned} \tag{2}$$

The way to allocate the budget is to choose a cost-effectiveness threshold α for all the programs and then increase it until the money is spent. The term α is expressed in units of cost per health outcome achieved (e.g., cost per life year gained or cost per infection averted).

We will suppose in our analyses below that a value for the cost-effectiveness threshold α is known. This value could be determined from analysis similar to that suggested from the above optimization problem, or could be determined as an implicit value given by accepted public health/medical practice [Owens 1998].

In our model, the cost is equivalent to the budget, B, and thus is the maximum number of nodes we can trace (i.e., have in states ST and IT) at any time. The benefit or effectiveness is the decrease in disease prevalence achieved by contact tracing. Figure 5 shows the steady-state disease prevalence as we vary the budget. The convexity of the curve shows that the cost effectiveness of contact tracing decreases with its budget (i.e., it has diminishing returns to scale): for each incremental increase in the budget, the corresponding reduction in steady-state prevalence diminishes. This makes intuitive sense because as the budget increases and prevalence decreases, we trace more contacts, fewer of whom will be infected. Thus, the probability that the contacts we trace are infected decreases as the budget increases.

Such a figure can help us decide how large a budget we should allocate to contact tracing and what level of disease prevalence we will tolerate. Suppose that it costs \$6,000 per year for each additional node we can trace at a time (i.e., for each increase in B), and suppose that our cost-effectiveness threshold α is \$10,000 per infected person per year. (This is equivalent to a threshold of \$50,000 per quality adjusted life year, QALY, with the assumption that the quality of life with the disease is 0.8 on a scale of 0 to 1 where 0 is death and 1 is a healthy life.)

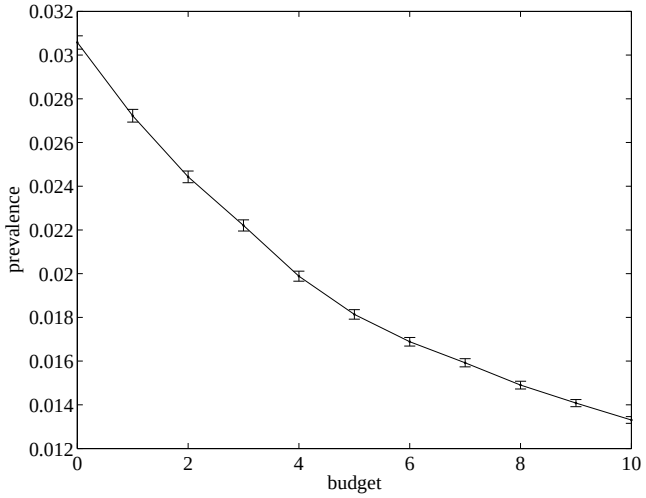

Figure 5. Steady-state prevalence under various budgets.

Figure 6 is a transformation of figure 5 with the yearly budget expressed in dollars and the prevalence expressed as average number of infected persons in the population (in a population of $n = 500$). The optimal budget is given by the point on the curve where the tangent line has a slope of $-1/\alpha$. In this picture, the optimal budget is about \$36,000 per year (or the ability to trace 6 people at a time) with about 8.5 people infected in steady state (a prevalence of 1.7% in a population of $n = 500$). At this point, the incremental cost per reduction in prevalence equals the maximum level we will tolerate, α.

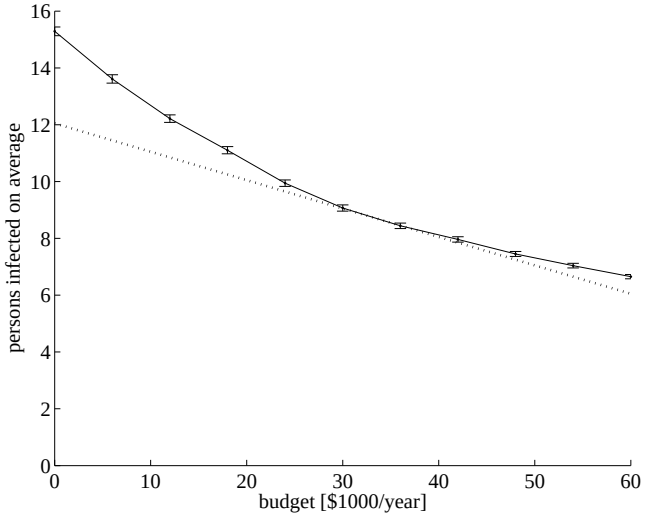

Figure 6. The number of infected persons per year in steady state under various budgets. At the optimal budget (the dotted line tangent to) this curve has a slope of $-1/\alpha$. Here $\alpha =$ \$10,000/infected person/year.

EXOGENOUS INFECTION

Thus far, the only form of disease control we have considered is contact tracing. Disease prevalence can also be decreased by screening for cases of exogenous infection (e.g., among immigrants, visitors from other countries, and travelers returning from vacation). Exogenous infection can be a major source of new infection for many diseases: for example, many TB index cases in the US are individuals who have brought the infection from another country. In the US and Canada, long term immigrants are screened for active TB and HIV as part of the visa process [U.S. Citizenship and Immigration Services 2006; Citizenship and Immigration Canada 2002].

In this section, we address the problem of allocating a combined budget $B_{total} = B + \lambda$ among contact tracing and screening for exogenous infection and the problem of determining the optimal size of the combined budget B_{total}. Here B is the budget allocated to contact tracing (as in the previous section) and λ is the budget allocated to screening. Without any screening, 0.056 exogenous infections occur in our population each day (calculated as 500/9000). We assume that with each budget unit we can screen 25 people per year, of whom 6% are infected on average; thus, the rate of exogenous infection as a function of λ is $\eta = \frac{1}{n}(5/90 - 0.06\lambda 25/356)$.

The framework from equation (2) does not help us allocate our budget because the benefits of contact tracing and screening are larger than the sum of the benefits of doing them separately: the cost-effectiveness of contact tracing varies with the amount of screening performed and vice versa. We instead use simulation to determine the optimal mix: we simulate different budget allocations to determine the effectiveness of each combination.

Figure 7 shows the steady-state prevalence achieved as we vary λ/B_{total} for different total budgets B_{total}. (In previous sections we set $\lambda = 0$.) We see from this figure that allocating a small fraction of the total budget to screening is optimal for smaller total budgets (no screening, $\lambda = 0$, is optimal up until $B_{total} = 5$), and for larger total budgets it is optimal to allocate a larger fraction to screening.

With this information about how to optimally split any given budget among contact tracing and screening, we can revisit the decision of how large to make the combined budget B_{total}. Figure 8 shows the steady-state prevalence as a function of the total budget (where the total budget is optimally allocated between screening and tracing). It clearly shows how for budgets $B_{total} > 6$, using an optimal mix of screening and contact tracing achieves lower disease prevalence than does contact tracing alone.

The way to use this information is still the same: given a cost-effectiveness threshold α we choose the point on the optimal-mix curve whose tangent has slope $-1/\alpha$. If we again choose $\alpha = \$10,000$/infected person/year then there will be will be little change compared to figure 6. In figure 6 the optimal budget was 6, a value for which there is little screening in the optimal mix and the two curves in figure 8 are very close for a total budget of 6.

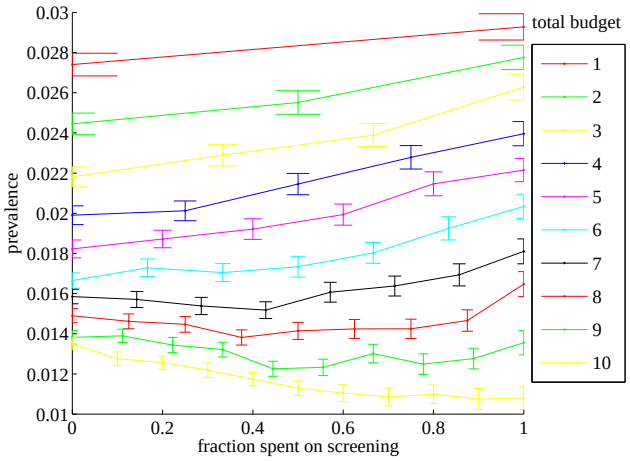

Figure 7. Steady-state prevalence as a function of the fraction, λ/B_{total}, spent on screening for exogenous infection and the total budget, $B_{total} = B + \lambda$.

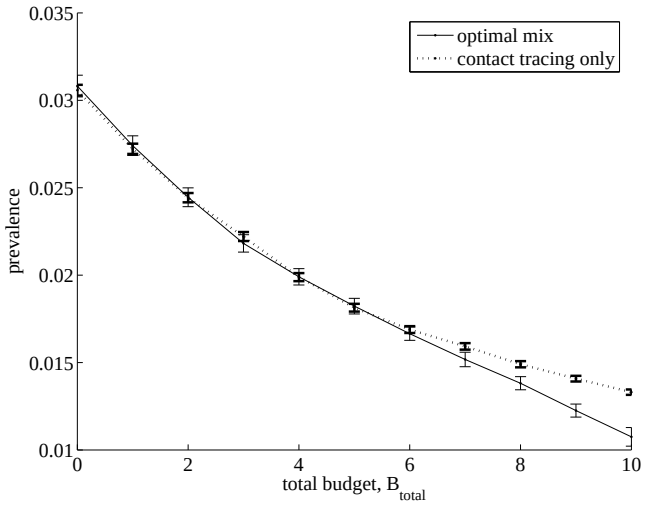

Figure 8. Steady-state prevalence achieved as a function of the combined budget B_{total} for screening and contact tracing. The solid line allocates the budget optimally while the dotted line is from figure 5 where we used no screening ($\lambda = 0$).

DISCUSSION

Our simulations show how a very simple policy for ranking contacts can improve significantly upon random selection of contacts. Our simulation results suggest that contact tracing is likely to have diminishing returns to scale: incremental increases in the budget for contact tracing yield diminishing decreases in the disease prevalence. Use of a cost-effectiveness

framework allows one to determine the appropriate level of investment in contact tracing. We show how such an investment decision is best made simultaneously with that for other interventions for the same disease (such as screening).

Our results are based on a limited set of simulations. Further simulations could explore the robustness of our findings under different conditions (e.g., for different networks, disease parameters, etc.).

Our simulations are based on a fairly stylized model of contact tracing and disease transmission and progression. For example, our current model stylizes the screening of infections from exogenous sources. A more realistic model could break out the various sources of exogenous infection (e.g., holiday travelers, visitors from certain countries, legal immigrants, and illegal immigrants) and the opportunities to screen them (e.g., when they request a visa or at clinics in immigrant neighborhoods). Our model of contact tracing also does not include the genotype information available to investigators which allows them to distinguish between new and continuing outbreaks. In practice, when a new outbreak of a disease is detected, the intensity of contact tracing is often increased until a significant level of epidemic control has been achieved. A natural extension of our work is to consider the case of dynamically changing levels of contact tracing.

Another useful avenue for further research would be to tailor the analysis to specific diseases of interest. A tailored model could be used to determine the appropriate level of contact tracing for a specific disease in a specific region.

For TB, the disease model should include latent and active infection stages, with disease progression times set appropriately. Further, the contact network needs to allow contacts of greater and lesser strength (e.g., family members versus coworkers in a well-ventilated office). In our simulation, a contact's priority is an indicator of the likelihood that this contact is infected. To better model TB contact tracing, it would be useful to distinguish individuals by their potential danger of acquiring infection, as is done in practice. For example, TB contact tracing in the US gives priority to contacts who are children or who have AIDS.

To more accurately model STDs and HIV, the disease model should distinguish between males and females and should include the asymptomatic and symptomatic disease stages. In addition, use of a dynamic contact network would reflect the pair formation and dissolution that occurs in social networks of such diseases (see, for example, Kretzschmar [2000]).

Contact tracing can be an effective means of disease control, but it is only useful up to a point because incremental increases in the level of contact tracing yield diminishing benefits. Simulation can be used to estimate the benefits of contact tracing as a function of its intensity. Then, cost-effectiveness analysis can be used to determine the optimal level of investment in contact tracing and the optimal level of investment in contact tracing and screening. Such analysis can help public health departments allocate funds for disease control.

ACKNOWLEDGMENT

B.A. gratefully acknowledges support of an NSF Graduate Research Fellowship. M.L.B. acknowledges support of the National Institute on Drug Abuse (Grant Number DA-R01-15612). The authors thank Dr. Neil Abernethy from the National Tuberculosis Control Center for helpful input.

REFERENCES

Armbruster, Benjamin, and Margaret L. Brandeau. 2006. Optimal mix of screening and contact tracing for endemic diseases. Working paper. Stanford University. submitted.

CDC. 2000. Chapter 10 - community TB control: Identification of persons who have clinically active TB.

———. 2002. Sexually transmitted diseases treatment guidelines 2002.

Citizenship and Immigration Canada. 2002. Fact sheet — medical testing and surveillance. `http://www.cic.gc.ca/english/pub/fs-medical.html`.

Clarke, Janette. 1998. Contact tracing for chlamydia: data on effectiveness. *International Journal of STD & AIDS* 187–191.

Cohen, Deborah A, Shin-Yi Wu, and Thomas A Farley. 2004. Comparing the cost-effectiveness of HIV prevention interventions. *Journal of Acquired Immune Deficiency Syndromes* 37(3).

Cowan, Frances M., Rebecca French, and Anne M. Johnson. 1996. The role and effectiveness of partner notification in STD control: a review. *Genitourinary Medicine* 72:247–252.

Dasgupta, Kaberi, Kevin Schwartzman, Robert Marchand, Terry Nan Tennenbaum, Paul Brassard, and Dick Menzies. 2000. Comparison of cost-effectiveness of Tuberculosis screening of close contacts and foreign-born populations. *American J. Respir. Crit. Care Med.* 162(6):2079–2086. `http://ajrccm.atsjournals.org/cgi/reprint/162/6/2079.pdf`.

Eames, Ken T. D., and Matt J. Keeling. 2003. Contact tracing and disease control. *Proceedings of the Royal Society B: Biological Sciences* 270:2565–2571.

Howell, M Rene, William J Kassler, and Anne Haddix. 1997. Partner notification to prevent pelvic inflammatory disease in women: Cost-effectiveness of two strategies. *Sexually Transmitted Diseases* 24(5):287–292.

Huerta, Ramon, and Lev S. Tsimring. 2002. Contact tracing and epidemics control in social networks. *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)* 66(5): 056115.

Hyman, James M., Jia Li, and E. Ann Stanley. 2003. Modeling the impact of random screening and contact tracing in reducing the spread of HIV. *Mathematical Biosciences* 181:17–54.

Kiss, Istvan Z., Darren M. Green, and Rowland R. Kao. 2005. Disease contact tracing in random and clustered networks. *Proceedings of the Royal Society B: Biological Sciences* 272:1407–1414.

Kretzschmar, Mirjam. 2000. Sexual network structure and sexually transmitted disease prevention: A modeling perspective. *Sexually Transmitted Diseases* 27:627–635.

Kretzschmar, Mirjam, Susan van den Hof, Jacco Wallinga, and Jan van Wijngaarden. 2004. Ring vaccination and smallpox control. *Emerging Infectious Diseases* 10(5).

Lipsitch, Marc, Ted Cohen, Ben Cooper, James M. Robins, Stefan Ma, Lyn James, Gowri Gopalakrishna, Suok Kai Chew, Chorh Chuan Tan, Matthew H. Samore, David Fisman, and Megan Murray. 2003. Transmission dynamics and control of severe acute respiratory syndrome. *Science* 300(5627):1966–1970. `http://www.sciencemag.org/cgi/reprint/300/5627/1966.pdf`.

Macintyre, C. R., A. J. Plant, and D. Hendrie. 2000. The cost-effectiveness of evidence-based guidelines and practice for screening and prevention of tuberculosis. *Health Economics* 9(5):411–421.

Müller, Johannes, Mirjam Kretzschmar, and Klaus Dietz. 2000. Contact tracing in stochastic and deterministic epidemic models. *Mathematical Biosciences* 164.

Owens, D. K. 1998. Interpretation of cost-effectiveness analyses. *J Gen Intern Med* 13:716–717.

Oxman, Gary L., and Linda Doyle. 1996. A comparison of the case-finding effectiveness and average costs of screening and partner notification. *Sexually Transmitted Diseases* .

Porco, Travis C, Karen A Holbrook, Susan E Fernyak, Diane L Portnoy, Randy Reiter, and Tomás J Aragón. 2004. Logistics of community smallpox control through contact tracing and ring vaccination: a stochastic network model. *BMC Public Health* 4(34).

U.S. Citizenship and Immigration Services. 2006. Medical examinations. `http://www.uscis.gov/propub/ProPubVAP.jsp?dockey=1c56d78b831659735c9fd1c1f7dbd68f`.

Varghese, Beena, Thomas A Peterman, and David R Holtgrave. 1999. Cost-effectiveness of counseling and testing and partner notification: a decision analysis. *AIDS* 13(13): 1745–1751.

Welte, R, M Kretzschmar, MJ Postma, R Leidl, JAR van den Hoek, JC Jager, and Maarten J. Postma. 2000. Cost-effectiveness of screening programs for Chlamydia trachomatis: a population based dynamic approach. *Sexually Transmitted Diseases* 27:518–529.

Wu, Joseph T., Steven Riley, Christophe Fraser, and Gabriel M. Leung. 2006. Reducing the impact of the next influenza pandemic using household-based public health interventions. *PLoS Medicine* 3:1532–1540.