Cox Survival Model based Optimal Screening Strategy for Tuberculosis

Kaiye Gao
School of Economics and Management
Beijing Information Science & Technology University
Beijing, China
kygao@foxmail.com

Rui Peng
School of Economics and Management
Beijing University of Technology
Beijing, China
pengruisubmit@163.com

Di Wu
School of Management
Xi'an Jiaotong University
Xi'an, China
wd_0824@stu.xjtu.edu.cn

Qingqing Zhai
School of Management
Shanghai University
Shanghai, China
zhaiqing59@163.com

***Abstract*—Tuberculosis is one of leading threats to human beings' health. Thus, the screening is essential as it can contribute to the more effective treatment of the patients infected with tuberculosis and the reduction of unintended infection from these patients to others. Adopting the Cox survival model to characterize the infection probability, an optimal screening strategy is proposed to minimize the corresponding total cost. Numerical examples are presented to illustrate the proposed model.**

Keywords-Screening interval; tuberculosis; infective diseases; cost; delay time

I. INTRODUCTION

Tuberculosis (TB) is the ninth leading cause of death worldwide and the leading cause from a single infectious agent, ranking above HIV/AIDS [1]. Active TB refers to disease that occurs in someone infected with mycobacterium tuberculosis and only these people can further infect others. However, since many people infected with active TB do not experience typical TB symptoms in the early stages of the disease, they are unlikely to seek medical care and to be correctly diagnosed on their own initiative [2]. Thus, the screening is essential to reduce the impact of TB on human being's health [3-5].

Ma et al. (2017) proposed a model to optimize the screening interval [6]. However, the infection probability is casually assumed. Differently, this paper adopts a commonly used cox survival model to characterize the infection probability. Using this cox model, the influences of different parameters on the optimal screening interval can be studied and such information will be useful for the management.

The total cost of screening strategy includes the following three components: 1) the cost of screening; 2) the cost associated with treatment if an individual is infected; and 3) the cost associated with onward transmission between the time of infection and the time of treatment.

Sponsors: The National Natural Science Foundation of China

The remainder of the paper is written as follows. Section 2 outlines the assumptions and notations used in our models. Section 3 presents the model for solving the optimal screening interval. Numerical examples are shown in Section 4. Finally, Section 5 concludes the study.

II. ASSUMPTIONS

In this paper, we study the transmission process and the screening scenarios of TB in a certain area. The problem considered here is described as following. Briefly, during an exposure period, a person may meet with others and some of them may be active TB carriers in their area. Then this person may get infected with TB through air. Herein, people generally can get infected with tuberculosis if they frequently meet with active TB carriers, especially in places with bad air circulation [7]. In order to limit the infection of TB, a periodic screening strategy is performed to detect TB. Based on the problem, the following assumptions are proposed for the model.

1) The exposure period L follows a distribution in a given interval [0, M].

2) A person meets others with rate τ per unit time. The number of people it meets is Poisson distributed with parameter $\mathcal{E}$. Both τ and $\mathcal{E}$ are functions of N.

3) The time of meeting follows exponential distribution with parameter μ. Its unit is hours, and it is negligible compared with screening interval.

4) The probability of infection is an increasing function of the number of infected people a person meets and the time span of meeting.

5) The proportion of people with active TB in the population is r.

978-1-7281-0860-5/19/$31.00 ©2019 IEEE

6) A screening occurs every T time units, costs C_s units and requires negligible time. In addition, the screening is perfect so that it can always detect the active TB.

7) A person can only be infected at most once throughout the exposure period.

8) Once a person is identified as a TB patient, the screening stops and the treatment starts. In this case, the person will not transmit others TB.

9) When the exposure period ends, if a person is infected, the individual will seek medical care initiatively. Finally, TB will be diagnosed and treated at that time.

III. THE MODEL

Cox proportional hazard model is used to calculate the expected probability that a person is infected during a meeting. In this study, a popular model based on Cox proportional hazard model is proposed to consider the specific hazard function of infection for a person during a meeting, which is

$$\lambda(t \mid i) = \lambda_0(t) e^{\beta(i-1)}, \ i \geq 1, \tag{1}$$

where $\lambda_0(t)$ is the baseline hazard that determines the shape of the survival function, i indicates the number of infected people meets, β indicates the regression coefficients and t indicates the time span of the meeting. In this equation, if β is positive, the hazard of infection is positively related with i; specifically, when i=1, it degenerates into the baseline hazard. It should be noted that the constant is absorbed in the baseline hazard: $\ln(\lambda_0(t))$ can be seen as a time-dependent intercept in the linear model of $\ln(\lambda(t \mid i))$.

Then the cumulative probability of non-infection can be expressed by Cox survival function, that is

$$S(t \mid i) = \exp(-\exp(\beta(i-1)\Lambda_0(t)) = \left[S_0(t)\right]^{\exp(\beta(i-1))}, \ i \geq 1, \tag{2}$$

where $\Lambda_0(t) = \int_0^t \lambda_0(s) ds$ is the cumulative baseline hazard, as well as $S_0(t) = \exp(-\Lambda_0(t))$ is the baseline survival function.

It is a common practice not to define a parametric model for the baseline hazard. Therefore, in this study, an exponential function is directly used for the baseline survival function, that is $S_0(t) = e^{\partial t}$. Then the probability of infection for a person during a meeting $p(i \mid t) = 1 - S(t,i) = 1 - \exp(\partial t)^{\exp(\beta i)}$.

Next, based on assumptions, it can be concluded that the infection rate for a healthy person is

$$\tau \left\{ \sum_{k=1}^{\infty} \frac{e^{-\varepsilon} \varepsilon^k}{k!} \sum_{i=1}^{k} \left[\binom{k}{i} r^i (1-r)^{k-i} \int_0^{\infty} \mu e^{-\mu t} p(t \mid i) dt \right] \right\}, \tag{3}$$

where $\tau = \omega N$, $\varepsilon = \eta e^{\theta(N-N_0)}$ and $p(i \mid t) = 1 - \exp(\partial t)^{\exp(\beta i)}$. Then, we can have the probability density function of the infection time x, as shown below

$$f(x) = \begin{cases} \tau \left\{ \sum_{k=1}^{\infty} \frac{e^{-\varepsilon} \varepsilon^k}{k!} \sum_{i=1}^{k} \left[\binom{k}{i} r^i (1-r)^{k-i} \int_0^{\infty} \mu e^{-\mu t} p(i \mid t) dt \right] \right\} e^{-\lambda \left\{ \sum_{k=1}^{\infty} \frac{e^{-\varepsilon} \varepsilon^k}{k!} \sum_{i=1}^{k} \left[\binom{k}{i} r^i (1-r)^{k-i} \int_0^{\infty} \mu e^{-\mu t} p(i \mid t) dt \right] \right\} x}, & x > 0 \\ 0, & x \leq 0 \end{cases}. \tag{4}$$

For the length of the exposure period L, a truncated distribution is used as shown in eq. (5):

$$h(l) = \frac{0.006 e^{0.3l}}{\int_0^M 0.006 e^{0.3l}}, \qquad l \in [0, M]. \tag{5}$$

Three scenarios in this model are considered as shown in Fig. 1, where $n = [l/T]$ is a variable representing the maximum number of screenings per person. The expected costs of these scenario are specified as follows.

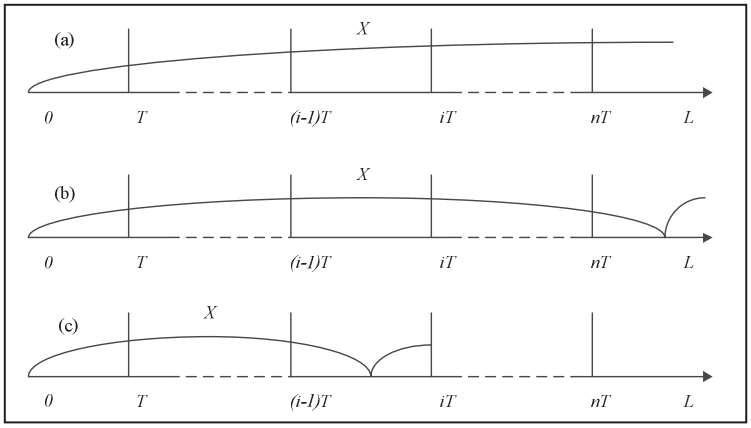

Figure 1. Three scenarios to consider

A. Scenario (1): never infected during the exposure period

As shown in Fig. 1 (a), the probability of this scenario is $P(X > L)$, indicating that an individual is never infected throughout the exposure period. In this case, the only cost item is the cost for screenings, as shown below

$$E_1(T) = \int_0^M h(l)(1-F(l))nC_s dl\,. \tag{6}$$

B. Scenario (2): infected after the last screening

In this scenario, a person is infected after the last screening. Therefore, the infection is detected by the self-seeking medical care rather than regular screenings, as shown in Fig. 1 (b). The probability of this scenario is $p(nT < X < L)$ and the cost incurred in this scenario includes the cost of screening, the cost of treatment as well as the cost of onward transmission from the time of infection to the end of the exposure period. Then, the expected cost can be expressed as

$$E_2(T) = \int_0^M h(l)\int_{nT}^{l} f(x)(nC_S + C_t(l-x) + C_o(l-x)\tau\left\{\sum_{k=1}^{\infty}\frac{e^{-\varepsilon}\varepsilon^k}{k!}\sum_{i=1}^{k}\left[\binom{k}{i}r^i(1-r)^{k-i}(k-i)\int_0^{\infty}\mu e^{-\mu t}(p(t\,|\,i+1)-p(t\,|\,i))dt\right]\right\}dxdl\,, \tag{7}$$

where $(p(t\,|\,i+1)-p(t\,|\,i))$ is the probability that one brings the burden to others if one more person is infected with active TB.

C. Scenario (3): infected before the last screening

In this case, an infection occurs before the last screening, as shown in Fig. 1 (c). According to the assumption of perfect screening, the active TB will be diagnosed at the next screening that follows the infection. Since screening stops after treatment, the number of screenings in this scenario is $[x/T]+1$. Similarly, as Scenario (2), the cost in this scenario includes three parts: the cost of screening, the cost of treatment as well as the cost of onward transmission that happens from the time of infection to the time of diagnosis. Then, we can have the expected cost, as shown below

$$\begin{aligned} E_3(T) = \int_T^M h(l)\int_0^{nT} f(x)(([x/T]+1)C_S + C_t(([x/T]+1)T - x) \\ + C_o(([x/T]+1)T - x)\tau\left\{\sum_{k=1}^{\infty}\frac{e^{-\varepsilon}\varepsilon^k}{k!}\sum_{i=1}^{k}\left[\binom{k}{i}r^i(1-r)^{k-i}(k-i)\int_0^{\infty}\mu e^{-\mu t}(p(t\,|\,i+1)-p(t\,|\,i))dt\right]\right\}dxdl\,. \end{aligned} \tag{8}$$

D. The expected cost considering all three scenarios

Overall, combining the costs of the three scenarios discussed above, we can have the expected total cost function, as shown below

$$E(T) = E_1(T) + E_2(T) + E_3(T). \tag{9}$$

Finally, the optimal screening interval can be determined by minimizing the Equation (9).

IV. Numerical Examples

To illustrate the proposed model, we present several numerical examples in this section. The parameter values are set as shown in Table 1.

Then, we assume the parameters and function of cost. The value of a single screening cost is set as $C_s = 1$. The cost functions of onward transmission and treatment are set as $C_o(t) = 100t$ and $C_t(t) = 100t$ respectively, where t is the delay time of diagnosis. Besides, the unit of N is assumed to be the number of people per square kilometer.

For N = 50, 100, 200, 500, 800, 1000, 1200, 1400 and 1600, using the model we proposed above, the expected costs versus different uniform inspection intervals have been evaluated, as shown in Fig. 2.

TABLE I. Values of Parameters

Parameter	Value
ω	2
η	4
N_0	800
θ	0.0005
r	0.00001
μ	1.5
∂	-0.5
β	0.1
M	20

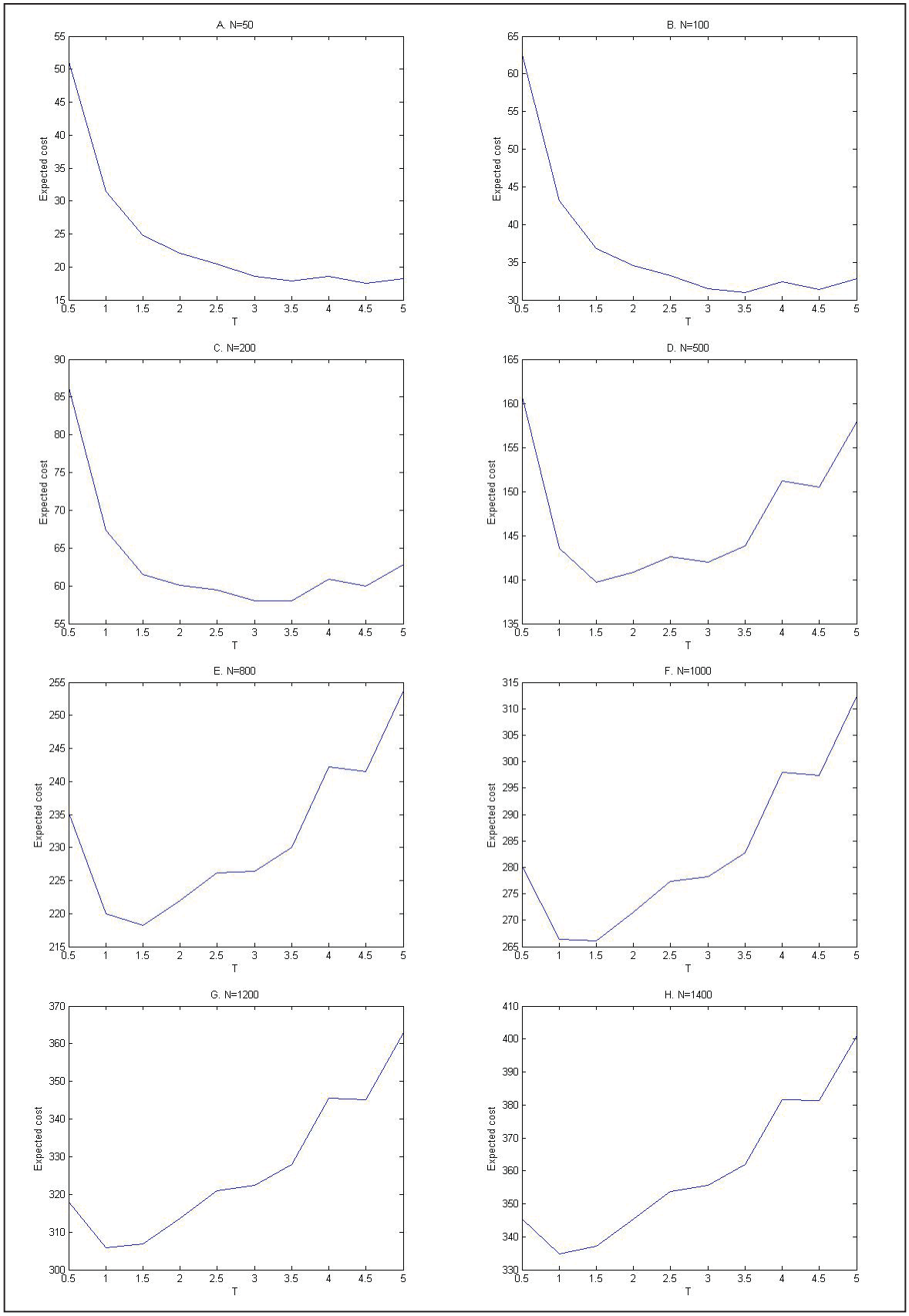

Figure 2. Expected cost by screening interval, T, from 0.5 to 5 years for different population density, *N*

It can be seen from Fig. 2 that when N is small, the expected cost decreases with the prolonging of screening interval. Actually, since the rate and the expected number of meeting people tend to be largely reduced in the sparsely populated region, the infecting channel is largely limited. Thus, excessive screenings are wasteful. However, with the increase of N, the value of T that minimizes the expected cost arises and gradually moves to the left. It means that when an area has a comparatively dense population, the optimal screening interval is short, as well as the expected cost first decreases until the optimal point and then turns to increase with screening interval.

V. CONCLUSIONS

In this study, we proposed a model to solve the optimal screening interval for TB. In the proposed model, the transmission process of infection and three different scenarios of screening outcomes are considered. The total cost of a TB screening strategy for an individual includes screening cost, treatment cost, and the burden if one more infected person that may further affect others. Numerical examples are presented to illustrate the proposed model. The results show that the optimal screening interval shortens with the increases of population density.

VI. ACKNOWLEDGMENT

The research reported here was partially supported by the NSFC under grant numbers 71531013 and 71671016.

REFERENCES

[1] Word Health Organization, “Global tuberculosis report 2017,” Australas. Med. JE., vol. 6, no. 2, 2017.

[2] Word Health Organization, “Systematic screening for active tuberculosis: principles and recommendations,” Word Health Organization, 2013.

[3] J. Kim, , K. S. Lee, E. B. Kim, S. Paik, C. L. Chang, and T. J. Park, et al., “Early detection of the growth of mycobacterium tuberculosis using magnetophoretic immunoassay in liquid culture,” Biosens. Bioelectron., vol. 96, no. 15, pp. 68-76, October 2017.

[4] M. Nakanishi, Y. Demura, S. Ameshima, N. Kosaka, Y. Chiba, and S. Nishikawa, et al., “Utility of high-resolution computed tomography for predicting risk of sputum smear-negative pulmonary tuberculosis,” Eur. J. Radiol., vol. 73, no. 3, pp. 554-550, March 2010.

[5] F. Feng, Y. X. Shi, G. L. Xia, Y. Zhu, and Z. Y. Zhang, “Computed tomography in predicting smear-negative pulmonary tuberculosis in aids patients,” Chinese Med. J., vol. 126, no. 17, pp. 3228-3233, 2013.

[6] X. Ma, X. Zhang, and R. Peng, “A framework for solving the optimal screening interval for tuberculosis,” HSOA J. Infect. Non-Infect. Dis., vol. 3, pp. 025, 2017.

[7] E. B. Famewo, A. M. Clarke, and A. J. Afolayan, “Ethno-medicinal documentation of polyherbal medicines used for the treatment of tuberculosis in Amathole District Municipality of the Eastern Cape Province, South Africa,” Pharm. Biol., vol. 55, no. 1, pp. 696-700, 2017.