MASS SCREENING MODELS FOR CONTAGIOUS DISEASES WITH NO LATENT PERIOD

HAU L. LEE
Stanford University, Stanford, California

WILLIAM P. PIERSKALLA
University of Pennsylvania, Philadelphia, Pennsylvania

(Received August 1985; revision received December 1986; accepted December 1987)

In this paper, a simplified model describing the stochastic process underlying the etiology of contagious and noncontagious diseases with mass screening is developed. Typical examples might include screening of tuberculosis in urban ghetto areas, venereal diseases in the sexually active, or AIDS in high risk population groups. The model is addressed to diseases which have zero or negligible latent periods. In the model, it is assumed that the reliabilities of the screening tests are constant, and independent of how long the population unit has the disease. Both tests with perfect and imperfect reliabilities are considered. It is shown that most of the results of a 1978 study by W. P. Pierskalla and J. A. Voelker for noncontagious diseases can be generalized for contagious diseases. A mathematical program for computing the optimal test choice and screening periods is presented. It is shown that the optimal screening schedule is equally spaced for tests with perfect reliability. Other properties relating to the managerial problems of screening frequencies, test selection, and resource allocation are also presented.

A population of human beings, plants or livestock is constantly subject to randomly occurring diseases. Early detection of such diseases is usually desirable as this may increase the chance of curing the disease, as well as reduce the chance that the asymptomatic population will develop the disease in the case of contagious diseases. Modern technological advancement in medical diagnosis has resulted in the introduction of various test procedures to detect different diseases. Administration of these test procedures to groups of the population, i.e., mass screening programs, may thus be advisable. For example, immigrants are tested for hepatitis B, school children are tested for tuberculosis, and military servicemen and women are tested for diseases such as AIDS and others. Some HMOs are implementing annual check-ups of patients for different diseases.

Monitoring mass screening programs, however, is an expensive task. The cost of mass screening includes easily quantifiable economic costs such as those of the labor and materials needed to administer the testing. Other cost components may be more difficult to quantify. For example, the cost may include the inconvenience and possible discomfort necessitated by the test; the cost of false positives which entails both emotional distress and the need to do unnecessary follow-up testing; and even the risk of physical harm to the testee, e.g., the cumulative effect of X-ray exposure or unnecessary surgery.

To the policy maker, mass screening programs have to be designed in light of the tradeoff of the expenses of testing, which increases with the frequency of test applications and with the cost of the type of test used, against the benefits to be achieved from detecting the defect in an earlier stage of development. Such a design must determine which kind of testing technology is to be used, as different technologies may have different reliability characteristics and costs. In addition, the frequency of testing must be decided. Moreover, because different subpopulations may have different susceptibility to the disease, the problem of optimal allocation of a fixed testing budget among subpopulations must be considered.

As will be discussed in the next section, while there exists a literature concerning the theory of screening, there has not been a comprehensive study on the...
analysis of mass screening to consider the above design issues. Most of the past research efforts utilize an approach to develop screening models based on a single individual’s decision for screening rather than a large group(s) of individuals. An individual through a lifetime is subject to different probabilities of incurring the disease and screening schedules are consequently evaluated for an individual. In cases when a sector of society (such as a health department or a prepaid group practice [HMO]) will bear the cost of administering screening programs and when a health agency, or the society at large, has only a fixed and limited amount of resources to be used for mass screening purposes, screening tests may not be provided continuously throughout the year, but rather periodically. The questions of which tests to use and when, which personnel to use to administer the tests and which utility functions of the group, subgroups or society should be maximized have not been analytically addressed when one considers the different etiology of diseases and different test capabilities and effects.

Most previous works have focused on noncontagious diseases such as cancer. In view of the potential benefit of detecting contagious diseases in reducing the risk of asymptomatic but susceptible groups in the population, models of screening for contagious diseases and their analyses will be of value to those decision makers responsible for administering such programs. The objective of the current study is to construct models of mass screening programs for both contagious and noncontagious diseases, and to conduct comprehensive analyses of the models to the screening objectives.

For contagious diseases, there are, in general, two distinct stages of the disease that can be identified once a population unit has contracted the disease: the latent and infectious periods. After a unit has contracted a contagious disease, a certain amount of biological development frequently is necessary before the disease can be passed on to others. This interval is usually termed the “latent period.” At the end of the latent period, the infected unit becomes contagious for a period of time, called the “infectious period.” The infectious period ends when some of the disease are recognizable and the unit is isolated or removed from the population, or the unit leaves the system by death or other causes. The sum of the latent and infectious periods is called the “incubation period” (see Bailey 1975).

Unfortunately, the model with both the latent and infectious periods is very complex and analytically intractable. A review of the literature has shown that, in most past studies, a zero latent period has been assumed (see, for example, McKendrick 1926; Bartlett 1949; Bailey; and Sanders 1971). Moreover, there are diseases in which the latent period is small and almost negligible, for example, scarlet fever, diphtheria, and possibly, hepatitis B. In this paper, we consider the model of mass screening for diseases with negligible or no latent period. The reliability (sensitivity and specificity) of the screening test is assumed to be constant, i.e., independent of how long the population unit has the disease. Both the cases of tests with perfect and imperfect reliabilities are considered. The managerial problems of screening frequencies, test selection and resource allocation are all examined.

1. Literature Review

Pierskalla and Voelker (1976), and McCall (1965) have provided comprehensive reviews of maintenance models for the control and surveillance of deteriorating systems in general. Rather than detail the very large literature on maintenance, in this section we focus on those past efforts which have a bearing on the problem of mass screening.

Kirch and Klein (1974a) address explicitly a mass screening application that seeks an inspection schedule to minimize expected detection delay (the time from disease incidence until its detection). The methodology is then applied to determining examination schedules for breast cancer detection (Kirch and Klein 1974b). McCall (1969) considers the problem of scheduling dental examinations, where cavities are assumed to occur according to a Poisson process.

Lincoln and Weiss (1964) study the statistical characteristics of detection delay under the assumption that the times of examinations form a renewal process and that the probability of detecting the defect is a function of the defect’s age. Zelen and Feinleib (1969) and Feinleib and Zelen (1969) have studied the statistical characteristics of the lead time (the time a disease is detected by screening to the time it is self-detected or symptomatic) provided by a screening program.

Much research has also been done in the etiology and progress of a disease and its relationship to screening effectiveness. Both the reliability of the test and the lead time gained from detection can be modeled as a function of the state of the disease rather than the time since the defect’s incidence (for example: Prorok 1976a, b; Thompson and Doyle 1976; Shwartz and Gallieher 1975; Thompson and Disney 1976; and Voelker 1976). Blumenson (1977) develops a mathematical model to evaluate a screening strategy, termed “compromise screening strategy,” which consists of
two stages of screening examinations with different harmful effects as well as accuracies. The model has also been applied to evaluate different screening intervals for breast cancer detection (Bross and Blumenson 1976). More recently, Shwartz (1978b) has developed a mathematical model of breast cancer and used it to evaluate the benefits of screening (Shwartz 1978a). Again the rate of disease progression is explicitly included in affecting the probability of the disease detection.

Eddy (1980), well aware of the complexity of relationships among test reliabilities, disease development, and prognosis of the disease, has constructed a breast cancer screening model by focusing on two attributes that carry information about the effectiveness of the screening tests: the mammogram interval and the patient interval. By modeling these two intervals as random variables, Eddy is able to derive analytical expressions for the sensitivity (true-positive rate) and specificity (true-negative rate) of test procedures, utilizing repeatedly the Bayesian statistical approach. The design of screening strategies to optimally allocate fixed resources, however, has only been briefly discussed. Eddy's work is important as it is one that has been implemented by health agencies.

The above mentioned studies concentrate on incorporating the process of disease progress in their models to evaluate screening programs. Almost all of these studies, however, take a longitudinal view of an individual. An individual through his/her lifetime is subject to different probabilities of incurring the disease and screening schedules are evaluated (these also with respect to an individual). For many diseases an important alternative approach is to take the society's or a group practice point of view. Hence, instead of screening program schedules for an individual, mass screening programs must be considered.

The only research on mass screening are the works by Pierskalla and Voelker (1978) and Voelker and Pierskalla (1980). Analytical models of a mass screening program are developed and analyzed for both the cases where the test procedures are perfect or imperfect. Optimal allocation of a fixed budget to different subpopulations are given in the perfect test case, whereas optimal decision rules concerning the best choice and frequency test are derived for the imperfect case.

Finally, to date, the theory of mass screening for contagious diseases has been generally overlooked. From a societal viewpoint, screening for such contagious diseases as tuberculosis of persons in urban ghetto areas and venereal diseases in the sexually active may have very significant benefits for the individuals, but also for their disease-free contacts and for newborns. However, such mass screening programs are costly, so models must be developed to tradeoff the costs and benefits for their most cost effective implementations.

2. The Mass Screening Model with No Latent Period

The following assumptions will be adopted:

1. The latent period of the disease is negligible.
2. The infectious period of the disease is exponentially distributed. Note that the termination of the infectious period can be a result of natural deaths.
3. The size of the population is large enough so that the number of susceptibles is relatively constant.
4. The incidence rates and the rates of transmission of the disease are stationary over time, and are independent of each other.
5. The probability of an increase in the number of infected units at time t is directly proportional to the number of units in the infectious period of the disease at time t.
6. Once an infected unit is identified as having the disease after a mass screening test, it will be removed from the population and subsequently treated. As a result, it is assumed to be no longer infectious.

The assumption that the infectious periods are exponentially distributed is a common one in most models (for example, see McKendrick; Bartlett 1949; Bartholomew 1967; and Bailey). Assumption 5 is also one that is generally used in most stochastic models in epidemics, dating as far back as McKendrick (see also, the reviews of Dietz (1967) and Bailey). Assumption 6 is relaxed in Section 6.

The following notation will be used:

\[ X(t): \] number of infected units at time t;
\[ \lambda: \] natural incidence rate of the disease;
\[ \gamma: \] rate of transmission of the disease from a contagious unit to a susceptible;
\[ \mu: \] rate of infected units ending the infectious period;
\[ \eta: \] probability that an infected unit will not be detected by a screening test.

Let \( \Delta X \) be the change in the number of infected units in \( (t, t + \Delta t) \). Then the stochastic process describing the etiology of the disease can be represented by
the following equations:

$$\Pr(\Delta X = +1) = \lambda \Delta t + \gamma X(t) \Delta t,$$

(1)

$$\Pr(\Delta X = -1) = \mu X(t) \Delta t.$$

(2)

Let $X_0 = X(0)$. The conditional probability generating function of $X(t)$ as defined by (1) and (2) is given in Bartlett (1980):

$$M_{X|X_0}(z, t | X(0) = X_0) = E[z^{X(t)} | X(0) = X_0] = \frac{(\gamma - \mu)^{\lambda U(t) - 1}}{[\gamma U(t) - \mu - \gamma(U(t) - 1)/\lambda]^{X_0+\lambda U(t)}}$$

(3)

where $U(t) = \exp(\gamma - \mu)$. is used to simplify the notation.

The conditional mean number of infected units in the population at time $t$ is given by evaluating $(\partial/\partial z)M_{X|X_0}(z, t | X(0) = X_0)$ at $z = 1$, which gives

$$E[X(t) | X(0) = X_0] = U(t)X_0 + \lambda[U(t) - 1]/(\gamma - \mu).$$

(4)

Define

$$X(t^-) = \lim_{\varepsilon \to 0} X(t - \varepsilon),$$

and

$$X(t^+) = \lim_{\varepsilon \to 0} X(t + \varepsilon).$$

Let $T_0 = 0$. Suppose that mass screening is performed at times $T_0, T_1, \ldots$. Consider the mass screening that takes place at time $T_m$. We note that $X(T^-_m)$ is the number of infected units in the population right after the last screening. Assume that the number of screenings taking place in a finite interval of time is finite, so that $X(T_m^-)$ and $X(T_m^+)$ are well defined. Now, at the moment just before the $m$th mass screening, the number of infected units in the population, $X(T^-_m)$, has a distribution whose generating function is given by $M_{X|X_0}(z, t | X(T^-_m))$ as in (3), where $X_0$ is replaced by $X(T^-_m)$, and whose conditional mean is:

$$E[X(T_m^-) | X(T^-_m)] = U(T_m - T_{m-1})X(T^-_m)$$

$$+ \lambda[U(T_m - T_{m-1}) - 1]/(\gamma - \mu).$$

(5)

Thus, theoretically, if the distribution of $X(T^-_m)$ is known, then the distribution of $X(T^-_m)$ can be obtained.

Given $X(T^-_{m-1})$, the number of infected units undetected in the $(m - 1)^{th}$ screening test is binomially distributed with parameters $X(T^-_{m-1})$ and $\eta$. As a result, we can write

$$E[X(T^-_m) | X(T^-_{m-1})] = \eta X(T^-_{m-1}),$$

(6)

From (5) and (6), we also have for $m \geq 1$,

$$E[X(T_m^-)] = \eta U(T_m - T_{m-1})E[X(T^-_{m-1})]$$

$$+ \lambda[U(T_m - T_{m-1}) - 1]/(\gamma - \mu).$$

(7)

Hence, (7) gives a recursive relationship between $E[X(T_m^-)]$ and $E[X(T_{m-1}^-)]$, the respective expected numbers of infected units before two consecutive mass screenings.

Summarizing, once the distribution of $X(T^-_{m-1})$ is known, the distribution of $X(T^-_m)$, and consequently, that of $X(T^-_m)$ can also be obtained. We can thus start with the distribution of $X(T^-_0)$ to obtain $X(T^-_1)$, then $X(T^-_2)$, ..., and so on.

Applying (7) recursively, we obtain the following relationship, which can be proven by induction:

$$E[X(T_m^-)] = \eta U(T_m - T_{m-1})E[X(T^-_{m-1})]$$

$$+ \lambda[U(T_m - T_{m-1}) - 1]/(\gamma - \mu).$$

(8)

3. Performance Measures for Evaluating Mass Screening Programs

To conduct a cost-benefit analysis so as to evaluate a mass screening program, we must be able to measure the benefits of the screening program. Voelker and Pierskalla have discussed various disutility functions, the reduction of which constitutes the benefit of mass screening. Two commonly used measures applicable here are the average detection delay for an individual, and the average number of infected units per unit time period in the population. Detection delay is the gap between the time of detection and the time of the incidence of the disease.

The total detection delay and the average number of infected units per period in the population are intimately related. Consider a time interval $[0, T]$ where $T$ is the point in time when mass screening takes place. Let $t_i$ be the point in time, $0 \leq t_i < T$, that the incidence of the disease occurs for the $i$th unit
in the population infected with the disease in \([0, T]\). Let \(t_{2i}\) be the end of the incubation period for the corresponding population unit. Define \(l_i(t) = 1\) if \(t_{1i} \leq t < \min(t_{2i}, T)\), and 0 otherwise. Hence, \(\sum l_i(t)\) is the number of infected units at time \(t\). Then:

aggregate detection delay for the population

\[
\sum l_i(t) dt \quad (t_{1i} \leq t < \min(t_{2i}, T))
\]

\[
= \sum \left[ \int_{t_{1i}}^{t_{2i}} l_i(t) dt \right]
\]

\[
= T \left[ \sum l_i(t) dt/T \right]
\]

\[
= T \cdot \text{average number of infected units per period.}
\]

As a result, defining the disutility function on the aggregate detection delay or the average number of infected units per period in the population are equivalent. Of course, this does not imply that these measures are equivalent to the average detection delay of an individual. In what follows, we will concentrate on the average number of infected units as our performance measure for evaluating mass screening programs.

For notational convenience, define \(\delta_m = U(T_m) - U(T_{m-1})\), \(m = 1, 2, \ldots\).

Let \(E[X(T_{m-1})] = X_{m-1}\). Using (4) and (6), the expected average number of infected units in the population in the interval \([T_{m-1}, T_m]\) is given by

\[
E\left\{ \frac{1}{T_m - T_{m-1}} \sum_{t=T_{m-1}}^{T_m} U(t - T_{m-1})X(T_{m-1}^{+}) \right. \]

\[
+ \lambda \left[ U(t - T_{m-1}) - 1 \right]/(\gamma - \mu) \left. \right\} dt \right\}
\]

\[
eq E\left\{ \frac{1}{T_m - T_{m-1}} \sum_{t=T_{m-1}}^{T_m} U(t)X(T_{m-1}^{+}) \right. \]

\[
+ \lambda \left[ U(t) - 1 \right]/(\gamma - \mu) \left. \right\} dt \right\}
\]

\[
= \frac{1}{(T_m - T_{m-1})(\gamma - \mu)} \left( \eta X_{m-1} + \frac{\lambda}{\gamma - \mu} \right)
\]

\[
\cdot [U(T_m - T_{m-1}) - 1] - \frac{\lambda}{\gamma - \mu}.
\]

**Theorem 1.** Consider a time horizon \(T\), where mass screening occurs at \(T_0(=0), T_1, T_2, \ldots, \text{and } T_{M-1} < T\), with \(T_0 < T_1 < \ldots < T_{M-1} < T\). The average number of infected units in the population over \((0, T)\) is given by:

\[
\frac{1}{T(\gamma - \mu)} \cdot \left\{ X_0^* + \frac{\lambda}{\gamma - \mu} \sum_{m=1}^{M} \eta^m \delta_m 
\right.
\]

\[
+ \frac{\lambda(1 - \eta)}{\gamma - \mu} \sum_{m=1}^{M} U(T_m - T_{m-1})
\]

\[
+ \frac{\lambda(1 - \eta)}{\gamma - \mu} \sum_{m=2}^{M} \delta_m \sum_{i=0}^{m-2} \eta^{m-1-i} U(-T_i)
\]

\[
- \frac{M \lambda(1 - \eta)}{\gamma - \mu} - \lambda T
\}
\]

where \(X_0^* = E[X(T_0^*)]\).

The proof of Theorem 1 and the subsequent theorems are given in the Appendix.

In the case where the reliability of the screening test is perfect, i.e., \(\eta = 0\), then the expected average number of infected units in the population over \((0, T)\) can be simplified to:

\[
\frac{\lambda}{T(\gamma - \mu)} \cdot \left\{ \sum_{m=1}^{M} U(T_m - T_{m-1}) - M - T(\gamma - \mu) \right\}. \quad (10)
\]

**4. The Model with Perfect Test Reliability**

The special case of perfect screening test reliability (10) provides some interesting results that we shall describe in this section. We shall first explore the answer to the question: should mass screening be given at equal intervals? We note that such an assumption has often been used in past mass screening models. Pierskalla and Voelker (1978) have shown that, for noncontagious diseases, schedules with equally spaced test intervals are optimal in terms of the detection delay, when the reliability of the screening test is perfect. We shall prove an analogous result in the following theorem. We shall use the term "M-periodic schedule" in a planning horizon \(T\) to denote mass screening programs that consist of \(M\) screening tests performed at \(0 = T_0 < T_1 < \ldots < T_{M-1} < T\) over the time horizon, where \(M = 1, 2, \ldots\).

**Theorem 2.** For screening tests with perfect reliability, screening schedules with equally spaced test intervals minimize the expected average number of infected units in the population, within the class of M-periodic schedules for \(M = 1, 2, \ldots\).
Theorem 2 thus implies that, in the case when screening tests are perfectly reliable, we need only to consider screening schedules that are equally-spaced apart. Suppose policy makers are concerned with a disutility function \( D(\cdot) \) defined on the average number of infected units in the population over a time horizon. Then, as long as \( D \) is increasing in its argument, equally-spaced screening schedules would minimize the disutility function. This is because any other schedule would give rise to a higher average number of infected units in the population, which would thus lead to a higher disutility function value, as long as \( D(\cdot) \) is increasing.

For screening schedules with tests with perfect reliability (hence, equal-interval tests), the average number of infected units in the population over a planning horizon of \( T \) time units can be written, using (10), as
\[
[\lambda/(\gamma - \mu)](M/T)[\exp((\gamma - \mu)(T/M)) - 1] - \frac{\lambda}{\gamma - \mu}.
\]  

Evidently, we would expect that, given \( T \), (11) would be decreasing in \( M \). We formally state this in the following theorem.

**Theorem 3.** Given \( T \) and a screening schedule of tests with perfect reliability, the average number of infected units in the population is a decreasing and convex function in the number of screening tests held in \( T \).

The consequence of Theorem 3 is that, if the disutility function \( D(\cdot) \) defined on the average number of infected units in the population is increasing in its argument, then \( D(\cdot) \) is decreasing in \( M \), the frequency of screening.

Now, suppose it costs \( C \) for each mass screening, and \( B \) is the amount that the society can afford or is willing to pay per unit time for mass screening. Then Theorem 3 shows that, given a budget, the society would try to increase \( M \) as much as possible. For argument's sake, suppose \( M \) is increased so that \( M/T = B/C \). The average number of infected units in the population under the optimal screening schedule becomes:
\[
\psi(B) = \frac{\lambda}{\gamma - \mu}[(B/C)][\exp((\gamma - \mu)(C/B)) - 1] - \frac{\lambda}{\gamma - \mu}.
\]  

As \( T \to \infty \), (12) can be viewed as the long range average number of infected units in the population under the optimal mass screening schedule.

Suppose there are \( J \) subgroups in the population, where there is little or no interaction among the subgroups as far as the transmission of the disease is concerned. Thus, each subgroup may be viewed as an independent population in terms of the etiology of the disease under screening. Let \( D_j(\cdot) \) be the respective disutility function for subgroup \( j \). Suppose \( B_j \) and \( C_j \) are the amount per unit time allotted to mass screening and the cost of each mass screening for subgroup \( j \), respectively. Let \( B \) be the amount per unit time available to all the groups. So far our results show that the problem facing the policy maker is to determine the \( B_j's \) so as to (see (12)):

minimize
\[
\sum_{j=1}^{J} D_j \left( \frac{\lambda}{(\gamma_j - \mu_j)^2} \frac{B_j}{C_j} \right) \cdot \left[ \exp \left( (\gamma_j - \mu_j)(\frac{C_j}{B_j}) \right) - 1 \right] - \frac{\lambda}{\gamma_j - \mu_j}
\]

subject to
\[
\sum_{j=1}^{J} B_j = B,
\]
\[
B_j \geq 0; \quad j = 1, \ldots, J.
\]

The parameters in the mathematical program above are subscripted according to the respective subgroups.

From Theorem 2, it is clear that if \( D_j(\cdot) \) is convex and increasing in the number of infected units for all \( j \), then the objective function (13) is decreasing and convex in \( B_j \). As a result, the resource allocation problem as described in (13) is a knapsack problem with separable convex functions as its objectives. We can thus use a marginal algorithm, which is guaranteed to be optimal because of the convexity property, to solve the resource allocation problem (see Fox 1966). Furthermore, the convexity of the objective functions ensures that the optimum solution is unique.

Suppose that the disutility function is defined on the average total number of infected units in the population so that the objective is to minimize \( D(\sum_{j=1}^{J} \psi_j(B_j)) \), where \( \psi_j(\cdot) \) is the average number of infected units in subgroup \( j \). Since \( D(\cdot) \) is increasing in its argument, it is thus sufficient to consider:

minimize
\[
\sum_{j=1}^{J} \psi_j(B_j)
\]

subject to
\[
\sum_{j=1}^{J} B_j = B,
\]
\[
B_j \geq 0; \quad j = 1, \ldots, J.
\]

Here we observe some interesting parallel results to that of Pierskalla and Voelker (1978). First, if
\((B_j^*, B_j^*, \ldots, B_J^*)\) is an optimal solution to (14), then 
\(B_j^* > 0\) for \(j = 1, \ldots, J\).

**Theorem 4.** Suppose \(B_j^*\)'s are the optimal solutions to (14), and that
\[
\lambda_i / [(\gamma_i - \mu_i)C_i] < \lambda_i / [(\gamma_i - \mu_i)C_i]
\]
(15)
(i) if \(\gamma_i > \mu_i\) and \(\gamma_i > \mu_i\), then
\[
(\gamma_i - \mu_i)B_j^* > (\gamma_i - \mu_i)C_j / B_j^*
\]
(16)
(ii) if \(\gamma_i < \mu_i\) and \(\gamma_i < \mu_i\), then
\[
(\gamma_i - \mu_i)C_j / B_j^* < (\gamma_i - \mu_i)C_j / B_j^*
\]
(17)

There are some interesting implications of Theorem 4. First, we define,
\[
r_j = B_j / C_j = M_j / T \quad \text{as} \quad T \to \infty, \quad j = 1, \ldots, J;
\]
(18)
as the long run frequency of mass screening for population subgroup \(j\). Then (16) and (17) can be rewritten, respectively, as
\[
(\gamma_i - \mu_i)/r_j > (\gamma_i - \mu_i)/r_j;
\]
(19)
and
\[
(\gamma_i - \mu_i)/r_j < (\gamma_i - \mu_i)/r_j.
\]
(20)

Suppose we hold \(\gamma_i - \mu_i\) and \(\gamma_i - \mu_i\) to be roughly the same to see the impact of \(\lambda_i\) and \(C_i\) on screening frequencies. Then Theorem 4 states that, if \(C_i/\lambda_i > C_i/\lambda_i\), then \(r_i < r_i\). Note that this mathematical result is true for both cases (i) and (ii) in Theorem 4. Hence, if the incidence rate for subgroup \(j\) is low and/or the cost of screening subgroup \(j\) is high, the group should be screened less frequently. This result is exactly analogous to the one observed in Pierskalla and Voelker (1978) for noncontagious diseases.

To see the relationship between \(\gamma - \mu\) and \(r\), suppose that \(\lambda_i/C_i = \lambda_i/C_i\). Suppose further that \(\gamma_i > \mu_i\) and \(\gamma_i > \mu_i\), or \(\gamma_i < \mu_i\) and \(\gamma_i < \mu_i\), and
\[
|\gamma_i - \mu_i| > |\gamma_i - \mu_i|.
\]
As a result (15) holds, implying (19) or (20), by Theorem 5. It is interesting to note that both (19) and (20) lead to
\[
|\gamma_i - \mu_i| / |\gamma_i - \mu_i| > r_i / r_i.
\]
(21)

The implication of (21) is that the effect of \(\gamma - \mu\) on \(r\) is not linear. For example, if \(|\gamma_i - \mu_i| = 2|\gamma_i - \mu_i|\), then the ratio of \(r_i\) to \(r_i\) is not necessarily 2. In fact, by (21), it is less than 2. The frequency of screening is thus not linearly proportional to the difference between the transmission and leaving rates of the disease.

### 5. Equal-Interval Tests with Imperfect Reliability

In this section, we investigate the model with \(\eta > 0\), i.e., tests with imperfect reliability. We shall assume that \(X_0 = 0\). We shall also confine our attention to tests that are equally spaced apart.

Define \(r\) to be the frequency of mass screening. Hence \(1/r\) is the interval between two consecutive screenings, and the screening schedule is \(T_i = 1/r, T_2 = 2/r, \ldots, T_m = m/r, \ldots\), and so on.

Define also \(a = \gamma - \mu\), so that \(U(x) = \exp(ax)\).

**Theorem 5.** For tests with imperfect reliability that are an interval \(1/r\), the average number of infected units in the population over a period \(T = M/r\) is given by
\[
\frac{\lambda r(1 - \eta)[e^{a/r} - 1]}{a^2[1 - \eta e^{a/r}]^2} - \frac{- \lambda}{a - \frac{\lambda}{\eta}}
\]
\[
+ \frac{\lambda \mu}{T} \left( \frac{e^{a/r} - 1}{a(1 - \eta e^{a/r})} \right)^2 (\eta e^{a/r})M - 1].
\]
(22)

Note that when \(\eta = 0\), (22) is reduced to (11), as expected. Evidently, the average number of infected units as given in Theorem 5 is still very complicated. We shall give an approximation for this expression.

We observe that the last term in (22) involves a term
\[
(\eta e^{a/r})^M - 1)/T = [\eta M e^{a/r} - 1]/T,
\]
as \(r = M/T\). Hence, if \(a = \gamma - \mu\) is negative, then by the fact that \(\eta < 1\), the above expression approaches zero as \(T \to \infty\) in the long run. When the test frequency \(r\) is such that \(\eta e^{a/r}\) is smaller than 1, \((\eta e^{a/r})^M\) will also be finite so that the above expression again approaches zero as \(T \to \infty\) in the long run. Obviously, if \(\gamma = \mu\) so that \(a = 0\), the above expression also vanishes as \(T \to \infty\). We can then approximate the average number of infected units as a function of test frequency as:
\[
\psi = \frac{\lambda r(1 - \eta)[e^{a/r} - 1]}{(\gamma - \mu)(1 - \eta e^{a/r})} - \frac{- \lambda}{\gamma - \mu}.
\]
(23)

Such an approximation is reasonable, of course, only if the above-mentioned conditions are met. We shall assume, here onward, that
\[
\eta e^{a/r} < 1, \quad \text{or} \quad r > -a/(\ln \eta).
\]
(24)

It is interesting to note that (23) differs from (11), the case with perfect test reliabilities, by the term \((1 - \eta)/(1 - \eta e^{a/r})\). By partially differentiating \(\psi\) w.r.t. \(\eta\), we can see that \(\psi\) is increasing in \(\eta\), as expected.
Condition 24 serves as a rough guideline for the lower bound of $r$. If (24) does not hold, then it can be seen that the last term of (22) would become very large as $T \to \infty$. Hence as a control of the contagious disease, we would like to set $r$ to be at least such that (24) holds.

Equation 23 can also be useful for selecting among different tests for mass screening. Suppose there are 2 tests, 1 and 2, with costs per screening $C_1$ and $C_2$, and test reliabilities $\eta_1$ and $\eta_2$, respectively. Also suppose that $B$ is the amount per unit time that the society can afford or is willing to pay for screening. Suppose also that the disutility function is increasing in $\psi$ so that it is sufficient for us to consider $\psi$. Then, from (18), we see that test 1 is preferred to test 2 if

$$\frac{(1 - \eta_1)e^{-\alpha B/t} - 1}{C_1[1 - \eta_1 e^{-\alpha B/t}]} < \frac{(1 - \eta_2)e^{-\alpha B/t} - 1}{C_2[1 - \eta_2 e^{-\alpha B/t}]}.$$  

It is also interesting to note that the above condition does not depend on the natural incidence rate $\lambda$.

6. Noncompliance of Identified Units

In this section, we relax the assumption that a population is no longer infectious once detected by mass screening. In the medical literature, there has been increasing evidence that patients may not comply or continue with the medical treatment program prescribed to them (for example, see Rogers and Curtis 1980; Shortell 1976; Eriksson and Mattsson 1983; and Eijlertsson and Berg 1984). The degree of patient compliance to prescribed treatments will evidently affect the effectiveness of mass screening programs. Hence, the evaluation and management of mass screening programs require an analysis of the compliance behaviors of these patients.

Compliance of a unit identified at a mass screening can be measured as the probability of the unit’s complying with treatment (which could be simple in prescription but difficult in practice). Such a probability can be expected to decrease over the elapsed time from when the infected unit has been identified at a mass screening. In this paper, we attempt to capture the effect of noncompliance of units by a simplified assumption. Suppose $\alpha$ is the probability that an infected unit identified at a mass screening test will not comply with prescribed treatments and therefore remain infectious. For those who comply, we assume that after $r$ unit times, there is a probability $\alpha_1$ that they will not continue to comply. Such a two-stage compliance model is a first attempt to analyze the effect of compliance on the effectiveness and management of mass screening programs. We assume that $r$ is usually of short duration, so that it would be smaller than the interscreening times.

Define

$$X(t) = \text{number of infected units who may communicate the disease to susceptibles at time } t;$$

they are the ones who are termed “infectious.”

$$V(t) = \text{number of infected units who are complying to prescribed treatment and are, thus, not infectious in the sense that they would not communicate the disease to others.}$$

$$W(t) = \text{number of surviving and complying infected units who stop complying at time } t \text{ (and hence, become infectious again).}$$

$$V(t') = \text{lim}_{t \to \infty} V(t - \epsilon).$$

$$V(t') = \text{lim}_{t \to \infty} V(t + \epsilon).$$

Suppose mass screenings are conducted at times $T_0, T_1, \ldots, T_m, \ldots$. Consider the mass screening at time $T_{m-1}$. We observe that the number of infectious units after the mass screening, given $X(T_{m-1})$, i.e., $X(T_{m-1}) | X(T_{m-1}^*)$, is binomially distributed with parameters $X(T_{m-1})$ and $1 - (1 - \alpha)(1 - \eta)$. Furthermore,

$$V(T_{m-1}) = X(T_{m-1}) - X(T_{m-1}^*),$$

and

$$E[V(T_{m-1}) | X(T_{m-1}^*)] = (1 - \alpha)(1 - \eta) X(T_{m-1}^*). \quad (25)$$

Define $\mu_0$ as the rate of leaving the system for an infected unit complying to prescribed treatments. We expect that $\mu_0 < \mu$. At the time just before $T_{m-1} + \tau$, the number of complying units remaining, given $V(T_{m-1}^*)$, or $V(T_{m-1}^* + \tau) | V(T_{m-1}^*)$, is thus binomially distributed with parameters $V(T_{m-1}^*)$ and $q$, where $q = \exp(-\mu_0 \tau)$.

The number of infected units who discontinue compliance of treatment at time $\tau$, $W(T_{m-1} + \tau)$, given $V(T_{m-1}^* + \tau)$, is binomially distributed with parameters $V(T_{m-1}^* + \tau)$ and $\alpha$. Hence, $W(T_{m-1} + \tau) | V(T_{m-1}^*)$ is binomially distributed with parameters $V(T_{m-1}^*)$ and $\alpha_q$. Therefore,

$$E[W(T_{m-1} + \tau) | V(T_{m-1}^*)] = \alpha V(T_{m-1}^*). \quad (26)$$

During $(T_{m-1}, T_{m-1} + \tau)$, the etiology of the disease follows (1) and (2). Hence, using (5) we can write:

$$E[X(T_{m-1}^* + \tau) | X(T_{m-1}^*)] = U(\tau)X(T_{m-1}^*) + \lambda(U(\tau) - 1)/(\gamma - \mu). \quad (27)$$

Now, the number of infected units right after $T_{m-1} + \tau$ is the sum of the existing infected units and
those surviving units who discontinue compliance of treatment at time $T_{m-1} + \tau$. Thus,

$$X[(T_{m-1} + \tau)^r] = X[(T_{m-1} + \tau)^r] + W(T_{m-1} + \tau). \quad (28)$$

Combining (25), (26), (27), and (28), we get:

$$E[X(T_{m-1})] = \beta U(T_{m-1}) X(T_{m-1}) + \lambda [U(T_{m-1}) - 1]/(\gamma - \mu), \quad (29)$$

where $\beta = 1 - (1 - \alpha)(1 - \eta)(1 - \alpha_1 q U(-\tau))$.

During $(T_{m-1} + \tau, T_m)$, the etiology of the disease again follows (1) and (2), so that from (5) we again have:

$$E[X(T_m) | X(T_{m-1})] = U(T_m - T_{m-1} - \tau) X[(T_{m-1} + \tau)^r] + \lambda [U(T_m - T_{m-1} - \tau) - 1]/(\gamma - \mu). \quad (30)$$

Combining (29) and (30) we get

$$E[X(T_m) | X(T_{m-1})] = \beta U(T_m - T_{m-1}) X(T_{m-1}) + \lambda [U(T_m - T_{m-1}) - 1]/(\gamma - \mu). \quad (31)$$

Equation 31 is thus the analogue of (7), giving a recursive relationship between $E[X(T_{m-1})]$ and $E[X(T_m)]$.

Define $X_0 = E[X(T_0)]$. By replacing $\eta$ with $\beta$ in (7) and using Theorem 1 we get for $m \geq 1$,

$$E[X(T_m)] = \beta^n U(T_m) X_0 + \frac{\lambda}{(\gamma - \mu)} \left\{ 1 - \beta \sum_{i=0}^{m-1} \beta^{m-i-1} U(T_m - T_i) + \beta^n U(T_m) - 1^{-} \right\}. \quad \text{Theorem 6.}$$

Consider a time horizon $T$, where mass screening occurs at $T_0, T_1, \ldots, T_{M-1}$, with $0 = T_0 < T_1 < \ldots < T_{M-1} < T$. The average number of infected units in the population over $(0, T)$ with the existence of noncompliance, is given by

$$\frac{1}{T(\gamma - \mu)} \left\{ \sum_{m=1}^{M} \left[ (1 - \alpha)(1 - \eta) \right. \right.$$

$$\left. - 1 + \beta U(T_m - T_{m-1}) \right] E[X(T_{m-1})] + \frac{1}{\gamma - \mu} \sum_{m=1}^{M} \left[ U(T_m - T_{m-1}) - 1 \right] + \frac{\lambda}{\gamma - \mu} \cdot \left\{ U(T_m - T_{m-1}) - 1 \right\}.$$
Now using (8), the expected average number becomes:

\[
\frac{1}{(\gamma - \mu)^T} \left\{ \eta \lambda \sum_{m=2}^{M} [\eta^{m-1}U(T_{m-1}) - 1][U(T_m - T_{m-1}) - 1] \\
+ X_0 \sum_{m=1}^{M} [\eta^{m-1}U(T_m) - U(T_{m-1})] \\
+ \frac{\lambda(1 - \eta)}{\gamma - \mu} \sum_{m=2}^{M} [U(T_m - T_{m-1}) - 1] \\
- \frac{\lambda}{\gamma - \mu} \sum_{m=1}^{M} U(T_m - T_{m-1}) - \frac{M\lambda}{\gamma - \mu} - \lambda T \right\},
\]

which, upon simplification and rearranging terms appropriately, yields the desired result.

**Proof of Theorem 2**

Suppose we have a planning horizon \(T\), and \(M\) is the number of screenings over \((0, T]\). When \(\eta = 0\), the average number of infected units in the population, given \(M\), is given by (10):

\[
\frac{\lambda}{T(\gamma - \mu)^2} \sum_{m=1}^{M} U(T_m - T_{m-1}) - \frac{\lambda M}{T(\gamma - \mu)^2} - \frac{\lambda}{\gamma - \mu}.
\]

(Eq. A1)

Evidently, the last two terms in (A1) are not affected by the screening schedule (i.e., \(T_0, T_1, \ldots, T_{M-1}\)). Hence, we need only to concentrate on the first term of (A1). Since \(U(t)\) is convex in \(t\), then the first term in (A1) is convex because the sum of convex functions is convex.

Suppose that, in the screening schedule \(T_0, T_1, \ldots, T_{M-1}\), there exist \(T_{m-1}, T_m, T_{m+1}\) such that:

\(T_m - T_{m-1} \neq T_{m+1} - T_m,\)

i.e., they are not equally spaced. Then, we note that

\[
\begin{align*}
[\lambda/T(\gamma - \mu)^2] & [U(T_{m+1} - T_m) + U(T_m - T_{m-1})] \\
\geq & 2[\lambda/T(\gamma - \mu)^2]U[(T_{m+1} - T_m) + (T_m - T_{m-1})]/2, \\
\text{(since } U \text{ is convex)} \\
= & [\lambda/T(\gamma - \mu)^2][2U[(T_{m+1} - T_m)/2]].
\end{align*}
\]

This implies that we can reduce the average number by rescheduling \(T_m\) to be at \((T_{m+1} - T_{m-1})/2\), the midpoint between \(T_{m-1}\) and \(T_{m+1}\). By the same argument, we can always improve by equal-spacing the screening intervals. This proves the theorem.

**Proof of Theorem 3**

Denote the expression (11) by \(\phi(M)\). Then:

\[
\phi'(M) = (\lambda/T)[\exp((\gamma - \mu)T/M)[1 - (\gamma - \mu)T/M] - 1]/(\gamma - \mu)^2
\]

\[
\phi''(M) = \lambda T \exp[(\gamma - \mu)T/M]/M^2.
\]

It is clear that \(\phi''(M) > 0\), implying that \(\phi'(M)\) is increasing and \(\phi(M)\) is convex in positive values of \(M\). But \(\phi'(M) \to 0\) as \(M \to \infty\). Hence, \(\phi'(M)\) is negative for all positive values of \(M\). This proves that \(\phi(M)\) is decreasing in \(M\).

**Proof of Theorem 4**

Since \(B^*_j > 0\), for all \(j\), the Kuhn-Tucker condition for the mathematical program (14) states that \(\psi_j(B^*_j) = \psi_j(B^*_j),\) or

\[
\frac{\lambda_j}{(\gamma_j - \mu_j)^2C_j} \left\{ \exp\left[\frac{(\gamma_j - \mu_j)C_j}{B_j} \right] 1 - \frac{(\gamma_j - \mu_j)C_j}{B_j} \right\} - 1
\]

\[
= \frac{\lambda_j}{(\gamma_j - \mu_j)^2C_j} \times \left\{ \exp\left[\frac{(\gamma_j - \mu_j)C_j}{B_j} \right] 1 - \frac{(\gamma_j - \mu_j)C_j}{B_j} \right\} - 1. \tag{A2}
\]

Define \(\phi(x) = \exp(x)(1 - x) - 1\).

(i) Now \(\phi'(x) = -x \exp(x) < 0\) for all \(x > 0\). Also \(\phi(0) = 0\), so that \(\phi(x) < 0\) for all \(x > 0\). Hence (15) implies that

\[
\phi((\gamma_j - \mu_j)C_j/B_j) < \phi((\gamma_j - \mu_j)C_j/B_j). \tag{A3}
\]

But \(\phi(x)\) is a decreasing function in \(x\), for all \(x > 0\). Hence, if \(\gamma_j > \mu_j\) and \(\gamma_j > \mu_j\), then (A3) implies (16).

(ii) \(\phi'(x) = -x \exp(x) > 0\) for all \(x < 0\). But \(\phi(0) = 0\). Therefore, \(\phi(x) < 0\) for all \(x < 0\). Hence (15) implies that (A3) holds. Now \(\phi(x)\) is an increasing function in \(x\), for all \(x < 0\). Hence, if \(\gamma_j < \mu_j\) and \(\gamma_j < \mu_j\), then (A3) implies (17).
Proof of Theorem 5

For tests equally spaced at 1/r time units apart, then

\[ U(T_m - T_{m-1}) = e^{\alpha/r}; \quad m = 1, \ldots, M \]

\[ U(- T_i) = e^{\alpha/r}; \quad i = 1, \ldots, M \]

\[ \delta_m = e^{\alpha/r} - e^{\alpha(r-1)/r}; \quad m = 1, \ldots, M. \]

Using Theorem 1 with \( X_0 = 0 \), the average number of infected units over \((0, T)\) is:

\[ \frac{\lambda}{T a^2} \left\{ \sum_{m=1}^{M} \frac{\eta e^{\alpha/r} - e^{\alpha(r-1)/r}}{\lambda} - aT \right. \]

\[ + (1 - \eta) \left[ -M + \sum_{m=1}^{M} e^{\alpha/r} \right. \]

\[ \left. + \sum_{m=2}^{M} \left[ e^{\alpha(r-1)/r} - e^{\alpha(r-1-m)/r} \right] \sum_{i=0}^{m-2} \frac{\eta^{m-1-i} e^{-\alpha/i}}{\lambda} \right]. \]

\[ = \frac{\lambda}{T a^2} \left\{ -aT + M(1 - \eta)(e^{\alpha/r} - 1) \right. \]

\[ + (e^{\alpha/r} - 1) \eta((\eta e^{\alpha/r})^M - 1)/(\eta e^{\alpha/r} - 1) \]

\[ - \eta(1 - \eta) e^{\alpha/r}(e^{\alpha/r} - 1) \frac{(M - 1)(\eta e^{\alpha/r} - 1)}{(\eta e^{\alpha/r} - 1)^2} \]

\[ - e^{\alpha/r} \left[ (\eta e^{\alpha/r})^{M-1} - 1 \right]. \}

Appropriately rearranging terms yields (22).

Proof of Theorem 6

Similar to the proof of Theorem 1, the cumulative number of infected units in the time interval \((T_{m-1}, T_{m-1} + \tau)\) can be found as:

\[ \frac{1}{\gamma - \mu} \left\{ (1 - (1 - \alpha)(1 - \eta))E[X(T^m_{m-1})] \right. \]

\[ + \frac{\lambda}{\gamma - \mu} \left[ U(\tau) - 1 \right] - \frac{\lambda \tau}{\gamma - \mu}, \] \quad (A4)

The cumulative number of infected units in the time interval \((T_{m-1} + \tau, T_m)\) can be found, in a similar way, as:

\[ \frac{1}{\gamma - \mu} \left\{ E[X(T_{m-1} + \tau)] + \frac{\lambda}{\gamma - \mu} \right\} \]

\[ \cdot [U(T_m - T_{m-1} - \tau) - 1] \]

\[ - \frac{\lambda}{\gamma - \mu} (T_m - T_{m-1} - \tau). \] \quad (A5)

Using (29) in (A5) and summing (A4) and (A5), we obtain the cumulative number of infected units in

\[ (T_{m-1}, T_m): \]

\[ \frac{1}{\gamma - \mu} \left\{ (1 - (1 - \alpha)(1 - \eta)(1 - \alpha,q) - 1 \right. \]

\[ + \beta U(T_m - T_{m-1})E[X(T^m_{m-1})] \]

\[ + \frac{\lambda}{(\gamma - \mu)^2} [U(T_m - T_{m-1}) - 1] \]

\[ + \frac{\lambda}{\gamma - \mu} (T_m - T_{m-1}). \] \quad (A6)

Summing (A6) over \( m = 1, \ldots, M \), and dividing by \( T \) gives the desired result.

Acknowledgment

This research was supported by the National Science Foundation under grant SES83-13446.

References


