RR–75–42

AN INVESTIGATION OF A HYPOTHETICAL MEDICAL SCREENING PROGRAM

J. H. Bigelow
H. V. Lee

December 1975

An Investigation of a Hypothetical Medical Screening Program

J. H. Bigelow and H. V. Lee

Introduction

This paper addresses the question what are the key parameters that determine the cost and effectiveness of a medical screening program intended to discover occult cases of a disease. The cost of the program will simply be the dollar value of the resources it uses annually. The effectiveness of the program will be measured by the mortality rate owing to a disease. The lower the mortality, the more effective we will consider the program.

To help answer this question, we have developed a model which relates the cost and effectiveness of a screening program to the parameters in question, and we have established the following relations. First, we express cost and mortality in terms of quantities more fundamental to a screening program. These are the number of screening tests per capita per year, N; the annual number of tests per capita that erroneously indicate the presence of the disease, F (the false positives); the annual number of cases per capita detected by screening, D; and the annual number of cases per capita missed by screening, and discovered instead in the clinic, K. Second, we establish relations between these quantities and a description of the disease process, the interval between successive tests performed on the same individual, and various other parameters. Among these parameters are the incidence of the disease, the costs and the effectivenesses of both early and late treatment, and the cost and capabilities (false positive and false negative error rates) of the screening test.

We will try to answer our question by varying each parameter individually, each time adjusting the interval between successive screening tests so that a constant effectiveness (i.e. mortality) is maintained. The effect of this change on the cost of the program will be a measure of the importance of that parameter. In addition, we will select pairs of parameters, which we will adjust along with the screening interval to yield a continuum of programs with constant cost and effectiveness. This will yield curves showing parameter changes about which one is indifferent.

Finally, we will investigate certain sensitivities of our results. Thus, we will ask whether the parameters which most heavily influence the cost are different for a program with

low effectiveness from one whose effectiveness is high. Another quantity we will examine is the transit time of the disease--that is, the length of time during which a case is occult but detectable by screening. We ask what difference does it make if there is a great deal of variation in the transit times of different cases?

A problem this paper gives insight into is the guiding of biomedical research. Such guidance is attempted by anyone who must allocate limited resources to competing research proposals. It is widely agreed that resources should be allocated to the different proposed projects on the basis of the relation of benefits to be expected and costs to be borne. The results of this paper are directly relevant to one-half of this question: how to assign a value to the expected results of a research project. The answer we give, of course, is that the value of a result that changes one or more of our parameters is indicated by the money that change will save in a constant-effectiveness screening program. For example, one of the parameters describing the test will be the false negative error rate, which might be reduced by research. Another parameter will be the incidence of the disease. A technique for preventing the disease could be modelled by reducing the incidence. We will also be able to model an improvement in the test that permits detection of the disease at an earlier stage. Of course, it will be easy to deal with reductions in the cost of the screening test or the cost of treating the disease in its early or late stage, and it will be equally easy to deal with improvements in the survival rates of treated patients.

Steady-State

The measures of cost and effectiveness are annual quantities. That is, they may change from year to year. For example, the screening program may be growing, increasing N from one year to the next. Or, the age structure, habits of hygiene, or nutritional status of the population may be changing, thus altering the incidence of the disease. If the screening program is new, there may be a backlog of cases, which allows temporarily for cases to be discovered either by screening or clinically, at a rate greater than the incidence.

To consider such time dependencies would make the comparison of benefits from different research results intolerably complicated. The cost and mortality would have to be calculated for each of many future years, as they change owing to the implementation of a result. Their values in each year would depend on the implementation strategy: should we make the change rapidly, at high cost? Should we move slowly, at little expense? Benefits, and/or costs incurred at different times would have to be compared, for which no uncontroversial method exists.

Therefore, we choose to consider only the steady-state values of the impacts. Thus, we assume that during each year

the same number of screening tests are done, that the backlog of old cases has been exhausted, and that the incidence of the disease does not change from one year to the next. This will reduce a multitude of impacts arising from a research result--namely, the cost and mortality in each of many years--to only two impacts. These two, of course, are the steady-state annual cost and the steady-state annual mortality.

Homogeneity

Another simplification we will make is that we will assume the population we are screening is effectively homogeneous. Thus, either all identifiable groups within the target population have the same incidence, or any group that does not cannot be induced to be screened either more or less frequently than any other group. This assumption implies that all cases of the disease are treated equally by the screening program, whether or not these cases are concentrated among certain individuals. By equal treatment, we mean that each case has the same chance of being screened any particular number of times.

This assumption of homogeneity is clearly unrealistic. The incidence of most diseases depends on age. Cancer, for example, is largely a disease of the old. Further, in most screening programs frequency of participation is also age dependent.

However, the assumption of homogeneity can easily be relaxed should this appear desirable in any specific application. Such a relaxation would be accomplished by partitioning the population into homogeneous groups, building a model of the kind we will describe for each group separately, and, finally, adding the results of the various models together. But since this paper describes a theoretical exercise--based to be sure on a study of the particular disease, cervical cancer--it seems unnecessary to complicate the model by denying homogeneity.

No Spontaneous Regression

We will also ignore in our model the problem of regressive cases. These are cases which, if left alone, would heal themselves rather than progressing to a clinical stage. However, when detected by screening they are indistinguishable in their early stage from progressive cases.

If we chose to include such cases, we could simply provide a second model for regressive cases, similar to the progressive case model we will describe. In the regressive model, however, "clinically discovered cases, K," would become "spontaneously recovered cases, S."

But to include such cases would require that we define a new consequence of the screening program, namely, the number of regressive cases detected and unnecessarily treated. If some risk of death were associated with the treatment, then the screening program would cause some to lose their lives unnecessarily, even while others were prevented from dying. The medical profession, whose primary ethic is "first, do no harm," might well value the two sources of mortality differently.

Cost and Mortality

Our model expresses cost and mortality as linear functions of the four fundamental variables defined above. To remind the reader, these are (all taken as annual per capita figures) the number of screening tests, N; the number of false positive tests, F; the number of cases detected by screening, D; and the number of cases discovered in the clinic, K.

The annual per capita cost C of the program is:

$$C = \alpha_N \cdot N + \alpha_F \cdot F + \alpha_D \cdot D + \alpha_K \cdot K \quad . \tag{1}$$

Here, α_N is the cost of carrying out one screening test. The cost of following up a falsely positive test with the procedures that prove the individual to be free of the disease is α_F. This procedure may be different from the one used for screening--e.g. a biopsy following a Pap smear in screening for cervical cancer--or it may simply be a repeat of the screening procedure. The cost of confirming the presence of disease, given that the screening test has indicated that it is present, plus the cost of treating the disease in its early stages is α_D. The cost of diagnosing and treating a clinical case--that is, a case of the disease in its later stages--is α_K.

Mortality owing to the disease is a linear combination only of cases detected by screening and cases diagnosed in the clinic. That is,

$$M = \phi_D \cdot D + \phi_K \cdot K \quad . \tag{2}$$

Here, the coefficients ϕ_D and ϕ_K are probabilities that a patient will die if he has the disease. The probability if the case was detected by screening, and therefore in an early stage, is ϕ_D, and ϕ_K is the probability if the case was discovered clinically, and hence in a late stage.

The Disease

We wish to determine how the number of screening tests, N, will influence the number of cases detected by screening, D. For this, we must have a precise description of the disease and the manner in which it can be detected in an early stage.

We conceive of the disease as passing from an initial point, through a development phase, to a terminal point. The initial point represents the earliest point at which the screening test can detect the disease. The terminal point represents the instant at which the disease would be detected in the clinic in the absence of any screening.

The time between these points is called the transit time. This is the time during which screening is expected to uncover some cases, to the benefit of the patient or the doctor. Not all cases will have the same transit time. Some will pass very rapidly through the development phase, and others may linger. Mathematically, we express this by means of a distribution G(T), which gives the fraction of cases with transit time no greater than T. The associated density function is $g(T) = \frac{dG}{dT}$.

In addition, we define an incidence I, which is the fraction of the target population we can expect to contract the disease each year. Ordinarily, we would have to specify whether we meant the number of cases that became detectable during a year, or the number of cases actually discovered. But because we are considering a steady state situation, these numbers are the same.

The Screening Test

Our model supposes there is a screening test which can detect the disease during its development phase. The test might measure the size of a tumor (e.g. by x-ray), the concentration of some substance (e.g. sugar concentration in urine), or the existence of abnormal cells on a microscope slide (e.g. from a Pap smear). There are, of course, additional possibilities.

The test cannot be perfect. On some occasions, it will fail to detect the disease when the disease is present. These are the false negative tests. On other occasions, it will indicate the presence of the disease in an individual without it. These are the false positive tests. Both sorts of errors could be owing to contamination of samples or laboratory chemicals, observer error, or failures of administration and record keeping.

We can partition all tests into four categories according to two criteria. First, we may consider whether a test is positive, indicating the presence of the disease, or negative, indicating its absence. In addition, we may note whether the test was carried out on an individual with the disease, or without. Table 1 below pictures this categorization.

Table 1. Table of test results versus true outcomes.

Result of Test	True State of Individual: Diseased	Healthy
Positive	a_{PD}	a_{PH}
Negative	a_{ND}	a_{NH}

The entries in the table symbolize the numbers of tests in each category. For example, a_{PD} is the number of tests that had positive results that were done on diseased individuals. Similarly, a_{NH} is the number of tests with negative results that were done in healthy individuals. According to our definitions above, then, a_{ND} is the number of false negatives, and a_{PH} the number of false positives.

It seems reasonable to suppose that the number of errors of a given kind will be proportional to the number of opportunities to make that error. That is, the number of false negatives, a_{ND}, will be proportional to the number of tests done on diseased individuals. Similarly, the number of false positives will be proportional to the number of tests done on healthy individuals. Mathematically, then, we can define a false negative error rate "p," a false positive error rate "q," and write

$$a_{ND} = p(a_{ND} + a_{PD}) \tag{3}$$

$$a_{PH} = q(a_{PH} + a_{NH}) \quad . \tag{4}$$

Participation in the Screening Program

We will assume that the interval between successive screening tests is a constant, σ. Clearly, this assumption is unrealistic. Not everyone will attend a real screening program with the same frequency. Nor will most individuals attend with perfect regularity. Instead, attendance will be irregular and nonuniform. Moreover, this lack of regularity may have a considerable impact upon the effectiveness of the screening program. For example, if all cases of the disease had a transit time of one year, and if everyone were tested at constant intervals (as we have assumed), we would have to screen the population an average of once each year in order to be certain of subjecting each case to at least one test. If some people were tested with no more than half the average frequency, we would have to test people at an average frequency of twice a year in order to be sure of testing each case at least once. However, this study is intended to be illustrative, not definitive, and we feel that this simplifying assumption will not dilute that goal.

Detecting Cases

A case will be detected if at least one screening test is done during the development period of the case, and if not all of those tests are falsely negative. Conversely, the case will be missed by screening, and therefore detected in the clinic, if every test done on that case is a false negative.

Consider, then, a case with transit time T. Suppose that the first test on this case occurs at time τ after its inception, and that further tests occur at time $\tau + \sigma$, $\tau + 2\sigma$,..., $\tau + n\sigma$. Since we are assuming a constant interval σ between successive screening tests, we must have $o \leq \tau \leq \sigma$. Clearly, the probability of missing such a case is

$$M(T,\tau,\sigma) = \begin{cases} 1 & \text{if } T < \tau \\ p & \text{if } \tau \leq T < \tau + \sigma \\ p^2 & \text{if } \tau + \sigma \leq T < \tau + 2\sigma \\ p^n & \text{if } \tau + (n-1)\sigma \leq T < \tau + n\sigma \end{cases} .$$

Integrating over the distribution of transit times $g(T)$, we obtain the probability of missing a case, given that the first test occurs at time τ after its inception. Integrating the result over all admissible values of $\tau (0 \leq \tau \leq \sigma)$, we obtain the unconditional probability of missing a case, as a function of the screening interval σ, and the false negative rate p. This, after some manipulations is:

$$\lambda(\tau,p) = \frac{(1 - p)}{\sigma} \sum_{n=o}^{\infty} p^n \int_{no}^{(n + 1)\sigma} G(T)dT \qquad (5)$$

where G(T) is the cumulative transit time distribution, defined by

$$G(T) = \int_{o}^{T} g(s)ds \quad .$$

To find the number of cases missed per capita, we merely multiply $\lambda(\sigma,p)$ by the incidence I. That is

$$K = I \cdot \lambda(\sigma,p) \quad . \qquad (6)$$

In our model, cases are either discovered in the clinic, or detected by screening. There is no other possibility. Thus,

$$D = I - K \quad . \qquad (7)$$

False Positives

The number of false positive test results is equal to the false positive rate, q, multiplied by the number of tests done on healthy individuals. The number of tests done on healthy individuals is the difference between the total number of tests, and the number done on diseased individuals. Thus, to calculate the number of false positive tests, we must first calculate the number of tests done on diseased individuals, U.

To calculate U, we note that a fraction of p of these cases must be falsely negative, and the complementary fraction 1 - p must be truly positive. But each true positive test will detect a case. Hence

$$D = (1 - p) \cdot U \quad .$$

Therefore, we can calculate the number of false positive tests as

$$F = q\left(N - \frac{D}{1 - p}\right) \quad . \qquad (8)$$

The Model

What we have done so far permits us to express the cost and effectiveness of a screening program in terms of a variable that research might influence. These are such things as the false negative and false positive rates of the screening test, the interval between successive tests on the same individual, the incidence of the disease, etc. These relations are obtained by substituting equations (6), (7), and (8), into equations (1) and (2). The result is

$$C = \left(\alpha_N + \alpha_F \cdot q\right) \cdot \frac{1}{\sigma T} + \left(\alpha_D - \frac{\alpha_F \cdot q}{1 - p}\right) \cdot I \cdot \left(1 - \lambda(\sigma,p)\right) + \alpha_K \cdot I \cdot \lambda(\sigma,p) \quad . \tag{9}$$

$$M = \phi_D \cdot I \cdot \left(1 - \lambda(\sigma,p)\right) + \phi_k \cdot I \cdot \lambda(\sigma,p) \quad . \tag{10}$$

In these equations cost and mortality are expressed as annual quantities in per capita terms. Further, we have taken our unit of time to be $\bar{T}$, the average transit time. Thus σ is the number of average transit times between successive screenings; $\sigma\bar{T}$ is the years between screenings. Accordingly, the factor $1/\sigma\bar{T}$ in the first term of equation (9) is the number of screenings per year per capita.

An Example: Screening for Cervical Cancer

In using this model we chose as an example the screening program described by the parameters shown in Table 2. These values were obtained from the literature, and are appropriate for a hypothetical program for screening for cancer of the uterine cervix. Schneider and Twiggs [9] provide most of the cost factors, including the cost of a test[1] (α_N = \$5), the cost of

---

[1] In screening for cervical cancer, the test is called the Pap smear. It involves scraping cells from the surface of the cervix, placing them on a glass slide, and staining and reading them. The costs of all these operations, plus administration costs, are included.

Table 2. Nominal values of parameters.

Description of Parameter	Symbol	Value
Cost of screening test	α_N	\$5
Cost of false positive result	α_F	\$460
Cost of case detected early (by screening)	α_D	\$460
Cost of case discovered late (in clinic)	α_k	\$2,670
False negative rate	p	0.25
False positive rate	q	3.0×10^{-5}
Mortality among cases detected early	ϕ_D	0.0
Mortality among cases discovered late	ϕ_K	0.5
Incidence (new cases per capita per year)	I	2.7×10^{-4}
Average transit time (years)	$\bar{T}$	10

a false positive test (α_F = \$460), and the cost of a case detected early[2] (α_D = \$460). The cost of treating a clinically discovered case was obtained by weighting the costs of treating the disease at different stages (from Schneider Twiggs [9]) by the probabilities of discovery at those stages (from Campbell [4]). The result is α_k = \$2,670.

The false positive rate (p = 0.25) was obtained using a method in Bigelow and Ellis [2] applied to data of Dunn et al. [5]. The false positive rate ($q = 3.0 \times 10^{-5}$) was taken so that 10% of the positive smears would be falsely positive, this being the figure quoted by Schneider and Twiggs [9].

Mortality among cases detected and treated early is very low, as shown by Boyes et al. [3]. Thus we have used a mortality rate among early cases of zero (ϕ_o = 0.0). Campbell [4] on the other hand, reports a mortality among late cases of about 50% (ϕ_k = 0.5).

The incidence we have chosen (I = 27 cases per 100,000 population per year) is taken to be the incidence of clinical cases among the unscreened population of British Columbia (Fidla et al. [6]). This figure seems typical of many areas of the world.

There is considerable controversy concerning the average transit time of cervical cancer. Figures as low as 4.5 years and as high as twenty years have been proposed (see Green [7]). We have chosen $\bar{T}$ = ten years, although we will later investigate the effect of different choices.

Choosing Base Cases

We will examine screening programs, using the values in Table 2 for the parameters, under a variety of circumstances. Thus, we will consider screening programs which aim at only a moderate 40% reduction in mortality from the uncontrolled value, yielding eighty-one deaths annually per one million population; and other, more ambitious programs, which reduce mortality by 90% to 13.5 deaths annually per million people.

---

[2]The cost of determining that a test was falsely positive is the same as the cost of treating a case detected by screening. This is so because the procedure is the same in both instances. A cone is cut around the opening of the cervical canal, and then examined. If no cancer is found, the test was falsely positive. If cancer is found, but is confined in the epithelial layer of tissue, the case is deemed to have been treated. More advanced cases we include with clinically detected cases, since it is likely that in the absence of screening most such cases would be discovered almost immediately.

A final quantity that we require is the transit time distribution. We assume this distribution to be log-normal, which implies that only two parameters are needed to specify it. We have already taken the mean $\bar{T}$ of the distribution to be ten years. As our one remaining parameter we choose the distribution's coefficient of variation.

To determine whether the coefficient of variation is truly important, we have calculated the cost of a program with each of our target mortalities as a function of that parameter. The results appear in Figure 1. From these results we can conclude, first, that the cost of the program rises as the target mortality falls, and second, that the cost (at constant mortality) rises as the coefficient of variation rises. These two results should be expected, and that they are predicted merely increases our confidence in the model.

A third conclusion is that the more one wishes to reduce mortality, the more important is the coefficient of variation. Illustrating this conclusion is the fact that if the target mortality is eighty-one deaths per million people, the cost of the program rises only 6% as the coefficient of variation goes from 0.1 to 1.0. If, however, the target mortality is 13.5 deaths per million people, the same change in the coefficient of variation doubles the cost of the program.

This conclusion is still further buttressed by Figure 2. There we have shown the change in the cost of two programs as a function of the mean transit time. The upper curve gives the cost of a program whose target mortality is 13.5 deaths per million people. The low curve is for a less effective program, with a target mortality of eighty-one deaths per million people. Not only is the cost of the low-mortality program uniformly greater, but it also varies more startingly over the range of uncertainty of mean transit time (approximately 5 to 20 years). Thus, the cost of the high-mortality program over that range in $\bar{T}$ varies from 0.616 to 1.02, a ratio of 1.66 while the cost of the low-mortality program ranges between 0.638 and 2.0, giving a ratio of 3.13.

In general, then, we can support the conclusion of Knox [8]. This conclusion is that if one wishes a program of only moderate effectiveness, then ignorance of the precise nature of the disease is not too important. But, "...the cost of ignorance rises with the investment, and justification for research to resolve uncertainties must also mount as the investment mounts"[8].

Accordingly, for each target mortality, we will carry out our analysis for two coefficients of variation, one low (CV = 0.1), and one high (CV = 1.0). These two distributions are shown in Figure 3. This gives rise to four base cases, with different combinations of target mortality (high and low) and coefficient of variation (high and low). These four cases we described in Table 3. All figures in this table are per capita per year. Thus the "number of tests" is the annual number of tests per person. All costs are given in US dollars.

Table 3. Four base cases.

	Case #1: $M = 13.5/10^6$, CV = 0.1	Case #2: $M = 13.5/10^6$, CV = 1.0	Case #3: $M = 81/10^6$, CV = 0.1	Case #4: $M = 81/10^6$, CV = 1.0
PERFORMANCE				
No. tests	0.1812	0.3923	0.0534	0.0623
No. false positives	5.43×10^{-6}	11.76×10^{-6}	1.60×10^{-6}	1.86×10^{-6}
No. false negatives	8.10×10^{-5}	8.10×10^{-5}	3.60×10^{-5}	3.60×10^{-5}
No. cases found early	2.43×10^{-4}	2.43×10^{-4}	1.08×10^{-4}	1.08×10^{-4}
No. cases missed	0.27×10^{-4}	0.27×10^{-4}	1.62×10^{-4}	1.62×10^{-4}
COST, BY ELEMENT				
Taking tests	0.9060	1.9615	0.2670	0.3115
False positives	0.0025	0.0054	0.0007	0.0009
Treating early cases	0.1118	0.1118	0.0497	0.0497
Treating late cases	0.0721	0.0721	0.4325	0.4325
TOTAL COST	1.0924	2.1508	0.7499	0.7946

Discussion of Base Cases

Looking at Table 3, we can make the following observations. First, regardless of which case we consider, there are never very many false positive tests, and their cost is never significant. They would have to be much more likely (increase q) or much more costly (increase α_F) to change this.

Second, if the target mortality is low (Cases 1 and 2), the major portion of the cost is for taking tests. This suggests that the parameters of greatest importance in determining the total program cost will be those which primarily influence the cost or number of tests. This conclusion is even stronger if the coefficient of variation is high (Case 2) than if it is low (Case 1).

On the other hand, if the target mortality is high (Cases 3 and 4), the major part of the program cost is incurred in treating late (clinical) cases. Thus one would expect that measures reducing the cost or increasing the effectiveness of late treatment would most effectively reduce the program cost.

Finally, we note that when the target mortality is high (Cases 3 and 4), the coefficient of variation makes no important difference in any of the cost elements, and, alone (as we know already) it makes no difference in the total cost. It would seem, therefore, to be unnecessary to further analyze the influence of the coefficient of variation upon low-mortality screening programs. Therefore, we will drop Case 4, and do our further analyses on Cases 1, 2 and 3 only.

Sensitivity of Cost to Selected Parameters

The total cost of a screening program will, of course, depend upon the various parameters describing the program. In this section we explore some of these dependencies. We shall also examine the sensitivity of these dependencies to changes in the target mortality and changes in the coefficient of variation of the transit time distribution.

The first parameter we look at is α_K, the cost of treating the disease in its late, clinical stage. In Figure 4a, we see that when the target mortality is high, total program cost is very sensitive to α_K. This should not surprise us, as such a large percentage of cases are treated late (see Table 3). Conversely, when target mortality is low, α_K has relatively little influence on total program cost. Figure 4b shows that the sensitivity of total cost to α_K is not affected by changes in the coefficient of variation.

The opposite is true of the sensitivity of total cost to α_N, the cost of a screening test (see Figure 5). In Figure 5a we see that the cost of a program with a high target mortality

is very great indeed. Again, this should not come as a surprise, since a program with a low target mortality requires that many more screening tests be performed than a program with a high target mortality. Figure 5b demonstrates that changes in the coefficient of variation affect the amount of money that a reduction in α_N might save, but not the percentage of total program cost.

In Figure 6a, we see that if target mortality is high, total program cost is insensitive to the false negative rate of the screening test. The explanation is that since a program of this kind uses relatively few screening tests, the precise description of the test--including its cost and error rates--should not strongly influence program cost. On the other hand, we would expect, and we do in fact observe, that the cost of a program with low target mortality depends markedly on the false negative rate. Figure 6b shows that changes in the coefficient of variation influence the amount of money that a reduction in "p" might save, but not the percentage of total program cost.

The results shown in Figure 7 are somewhat counterintuitive. Figure 7a assents that if the target mortality of a program is high, then changing ϕ_K, the proportion of people with late-stage disease who die, should have virtually no affect on total program cost. But because the treatment of late-stage disease forms such a large part of a program with high target mortality (see Table 3), one would naively expect ϕ_K to strongly influence program cost--more strongly, in fact, than if target mortality were low.

The explanation of this phenomenon is that we are considering programs which maintain the same mortality rate even as ϕ_K is changed. Thus, changes in ϕ_K must be counterbalanced by an increase in the fraction of cases treated in their late stage, and hence by a reduction in the number of screening tests given. Those two factors--the increase in late-stage cases, and the decrease in screening tests--have opposite effects on total program costs. When target mortality is high, the two factors almost exactly offset each other. When target mortality is low, the cost of screening tests is a much larger part of the total cost, and hence the reduction in this factor outweighs the increase in the cost of treating the late-stage cases.

We account for the result shown in Figure 7b by carrying this explanation still farther. From Table 3 we find that as the coefficient of variation rises, so does the importance of the cost of screening tests. Thus changes in this cost will outweigh changes in late-stage treatment cost more heavily, the larger the coefficient of variation.

Perhaps the most satisfying result we have obtained is illustrated by Figure 8. This shows the effect of changing the incidence of the disease on total program cost. In this instance, as throughout this section, we are considering programs whose mortality is constant. Thus, the changes in incidence contemplated in Figure 8 are counterbalanced by changes in the number of screening tests, and hence changes in the fraction of cases that are treated in their late, clinical stage. What Figure 8 shows is that regardless of the coefficient of variation, a given reduction in incidence will permit the same percentage reduction in total program cost.

Trade-Offs Between Parameters

It should be said that changing the incidence of a disease is not possible. For example, the incidence of cervical cancer is correlated with the rate of vaginal infections (e.g. see Beral [1]). If the rate of infections rises, so will the rate--some years later--of cervical cancer. It would seem, therefore, that an educational program that taught principles and practice of good hygiene might well reduce the incidence of cervical cancer.

Indeed, all of the parameters considered in the last section might be changed by appropriate action. The case α_K of late-stage treatment might rise, or the mortality ϕ_K among the late cases fall, if a new treatment technique were perfected. A new test might have a different cost α_N or a false negative rate "p."

However, it seems more likely that such action would result in changes in two parameters simultaneously, rather than only one. Thus a different test would have both a different cost α_N and a different false negative rate p. A different late-stage treatment would have both a different cost α_K and a different mortality rate ϕ_K.

In this section, therefore, we will examine trade-offs between selected pairs of parameters. The trade-off curves we obtain map out the different values that the pairs of parameters may assume, and--with proper choice of an interval between screening tests--still yield a program with the same cost and mortality.

The curves presented in Figure 9 give the trade-off between the false negative rate and the cost per test. Each curve depicts the trade-off between the two parameters for a program with different target mortality and coefficient of variation. Thus, the left hand graph shows the sensitivity of the trade-off curve to changes in target mortality, while the right hand curve shows the effect of changing the coefficient of variation.

A new screening test is represented on this graph by a point, whose coordinates are that test's false negative rate p and cost α_N. If this point lies below a particular trade-off curve, then the new test will permit one to improve the associated program, either by lowering the cost while maintaining the same mortality, or by lowering the mortality without raising the cost.

From Figure 9 (left) we note that the lower the target mortality, the more one is willing to pay for a test with a smaller false negative rate. This is especially true for large reductions in "p". From Figure 9 (right), we conclude that the trade-off between α_N and p is unaffected by changes in the coefficient of variation, unless very small values of "p" are considered. Such values of "p" are worth more if the coefficient of variation is low than if it is high.

To motivate our next trade-off curves, between the false positive rate "q" and the mean transit time $\bar{T}$, we have included Figure 10. Here, the progression of the disease is indicated by the rise in a quantity to be measured by the screening test. This is shown by the solid heavy line. The lowest level at which the quantity in question is deemed abnormal is termed the threshold for early detection. The level at which this quantity becomes clinically obvious--or the level of this quantity when the disease manifests itself some other way--is called the threshold for clinical discovery. The time required for this quantity to rise from the first threshold to the second, is the transit time.

The region we have denoted as the normal range is not, of course, the range into which all healthy people will fall. Some people will have unusually high values of the measured quantity even though they are free of the disease. Thus some healthy people will yield a falsely positive result on the screening test. Lowering the threshold for early detection will increase errors of this kind.

On the other hand, if this threshold were reduced, then the transit time would increase. Thus there will be a trade-off between the mean transit time $\bar{T}$ and the false positive rate q. Figure 11 shows that the contemplated change in threshold would be worth making if the point described by the new values of q and $\bar{T}$ lay above the trade-off curve.

An interesting feature of this trade-off is its independence from changes in target mortality and coefficient of variation. Another interesting feature is the relatively large increase in "q" one is willing to suffer for only a moderate increase in $\bar{T}$. Thus one is willing that "q" should increase from its nominal value of 3×10^{-5}, to 1.2×10^{-3}, a fourty-fold increase, merely to increase $\bar{T}$ from ten to eleven years.

We should point out that at least in programs that screen for cervical cancer, distinctions are made among degrees of positiveness of screening results. There are not only the two results "healthy" and "cancerous." Rather there are many intermediate results such as "mild dysplasia," "severe dysplasia," and "borderline carcinoma in situ." (Different screening programs give these results different names.) And for intermediate results there are intermediate responses, the usual one being to take the next screening test sooner than ordinarily. Thus one should accept our trade-off between $\bar{T}$ and q with a grain of salt.

Figure 12 trades off the mortality among late cases, ϕ_K, and the cost of late treatment α_K. In Figure 12 (left), we see that if target mortality is high, we are indifferent to changes in ϕ_K (as we found earlier, see Figure 7), and only count a late stage treatment as preferable if it costs less. If target mortality is low, however, we are willing to pay more for a more effective treatment, up to a point.

Beyond that point, one is willing to pay progressively less and less for a better and better treatment. This seemingly contradictory result occurs because, when the cost of late treatment, α_K, is sufficiently high, and when the number of tests done annually is sufficiently low, total program cost is a locally decreasing function of the number of tests per year. That is, if the number of tests is increased by a small amount, the savings owing to treating a smaller number of cases in their late stage is greater than the money required to detect and treat those cases early.

Of course, this holds only for small increases in the number of tests. If the number of tests is increased enough, total cost will eventually rise.

How this phenomenon explains the anomalous portion of the ϕ_K versus α_K trade-off curve can best be seen by considering the scheme used for calculating that curve. This is illustrated in Figure 13. Thus suppose we are now at the initial point in the curve, with $\phi_K = 0.5$ and $\alpha_K = 2{,}670$, corresponding to $\phi_K = 0.45$. The first step is to reduce ϕ_K from 0.5 to 0.45, represented in Figure 13 (left) by the horizontal, left-printing line labeled "1, reduce ϕ_K." The result of this, as shown in the figure, is to leave total program cost constant while reducing mortality.

Because we are looking for a program with the same mortality as the base-case program, we must take steps to increase the mortality. This we do by reducing the number of screening tests N. In this particular region, cost "C" is an increasing function of "N", so that not only will mortality rise as N

falls, but cost will fall. This is shown by the curve in Figure 13 (left) labeled "2 reduce N."

Finally, we are obliged to increase the cost from its new value to the original, higher value. This we do by increasing the cost of late treatment, α_K, as shown in Figure 13 (left) by the line labeled "3 increase α_K."

However, if cost "C" increased in step 2 as N was reduced, rather than decreased, we would have the situation pictured in Figure 13 (right). This is in fact the situation that would occur if one started at the point $\phi_K = 0.2$, $\alpha_K = 6{,}800$, and wished to compute the value of α_K corresponding to $\phi_K = 0.15$. As before, reducing ϕ_K would leave the cost constant while reducing mortality. But while reducing N would increase mortality again, it would also increase the cost. This would necessitate a reduction in α as the third step in computation rather than an increase.

It is quite clear, therefore, that no real program should lie on the anomalous part of the ϕ_K versus α_K trade-off curve. If such a program existed, it could be improved by increasing the number of screening tests. Not only would this lower the mortality, but it would reduce the cost as well.

Our final observation of this section concerns Figure 12 (right). We note that if the coefficient of variation is high, one is willing to pay more for a given reduction in late stage mortality.

Conclusions

Clearly, our model of a medical screening program is incomplete. We have assumed a constant interval between successive screening tests. We have assumed that no case of the disease spontaneously disappears. We have assumed that only a very few options are available to the medical system, that (for example) observation without treatment is not possible for cases whose tests return an equivocal result. Thus our conclusions cannot be accepted without reservation. We believe, however, that these conclusions are at least reasonable, and should not be discarded without good reason.

Our conclusions are the following:

1) Regardless of how effective one intends his screening program to be, and regardless of the precise nature of the disease, it is always worthwhile to reduce the disease incidence. Because it is often very difficult and costly--or even impossible due to lack of data--to estimate the transit time distribution, this suggests that public health efforts to

link diseases with environmental and hereditary factors be encouraged.

2) If one's screening program aims at only a moderate reduction in mortality, the best way to reduce costs appears to be making late treatment less expensive. This is even worthwhile at the expense of making the treatment less effective, since the cost of such a program appears to be largely independent of ϕ_K.

3) If one's screening program aims to reduce mortality to very low levels, it is important to have a very good screening test. Thus reducing the false negative rate p on the cost of the test α_N will reduce total program cost significantly. It is also important to have an effective late stage treatment (i.e. a low ϕ_K). It is less important, however, to have an inexpensive late stage treatment.

4) If one's program consists largely of late stage treatment, and very few tests or early detections, one should consider whether increasing the screening component will not decrease both mortality and cost. This will especially be true if the cost of late treatment is very high.

5) If one's screening program aims at a large reduction in mortality, the cost of the program will depend heavily on the nature of the disease--i.e., the transit time distribution. If the disease has a short mean transit time, or if its transit time is highly variable, the program cost will be much larger than if the disease is less rapid or less variable. This means that the marginal effectiveness of further investment in screening is highly uncertain, unless the disease is well understood. Better economic decisions could be made in a more certain situation; therefore, when it is possible, it will be desirable to discover more of the nature of each screenable disease.

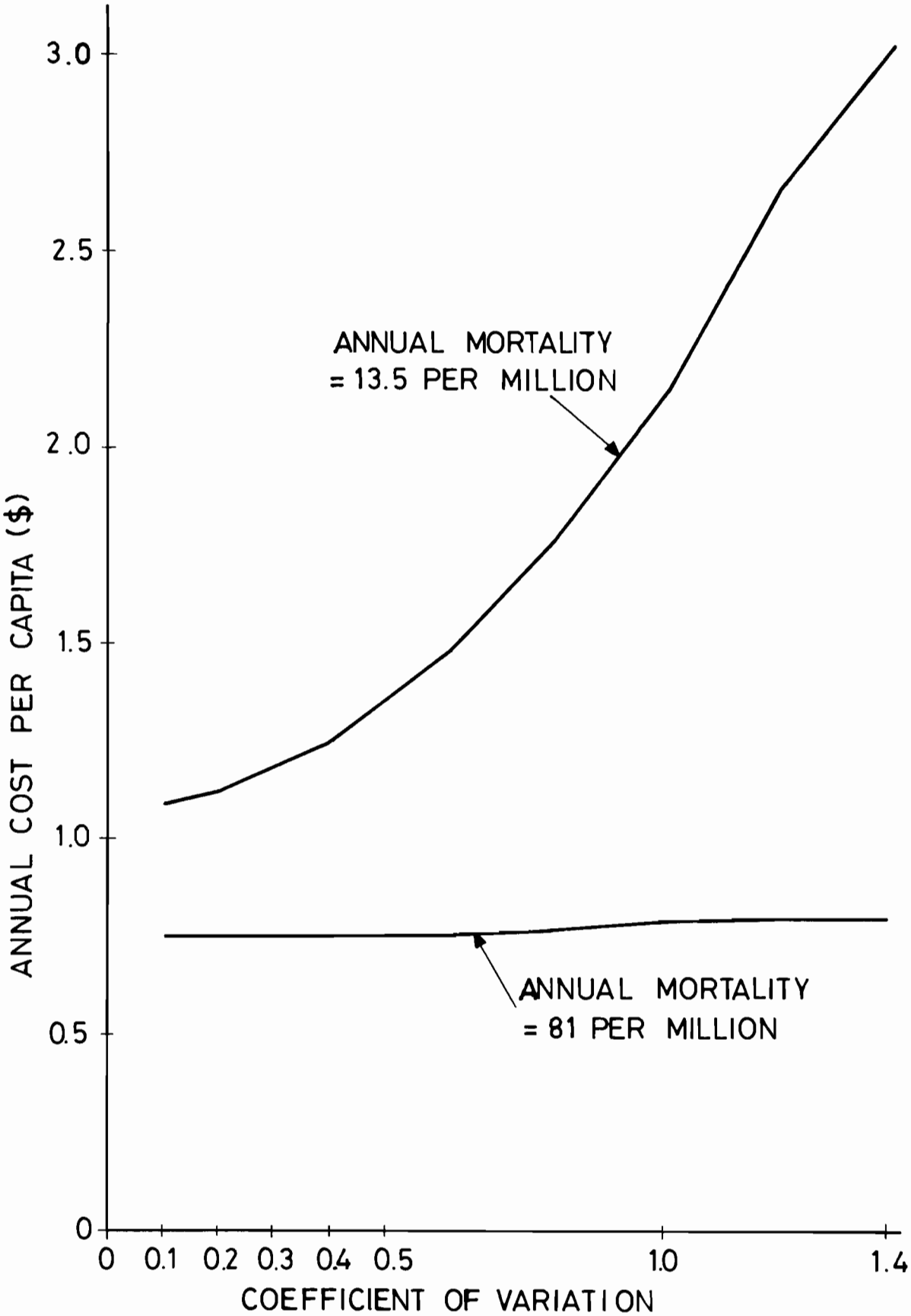


Figure 1. Effect on cost of variation in transit times.

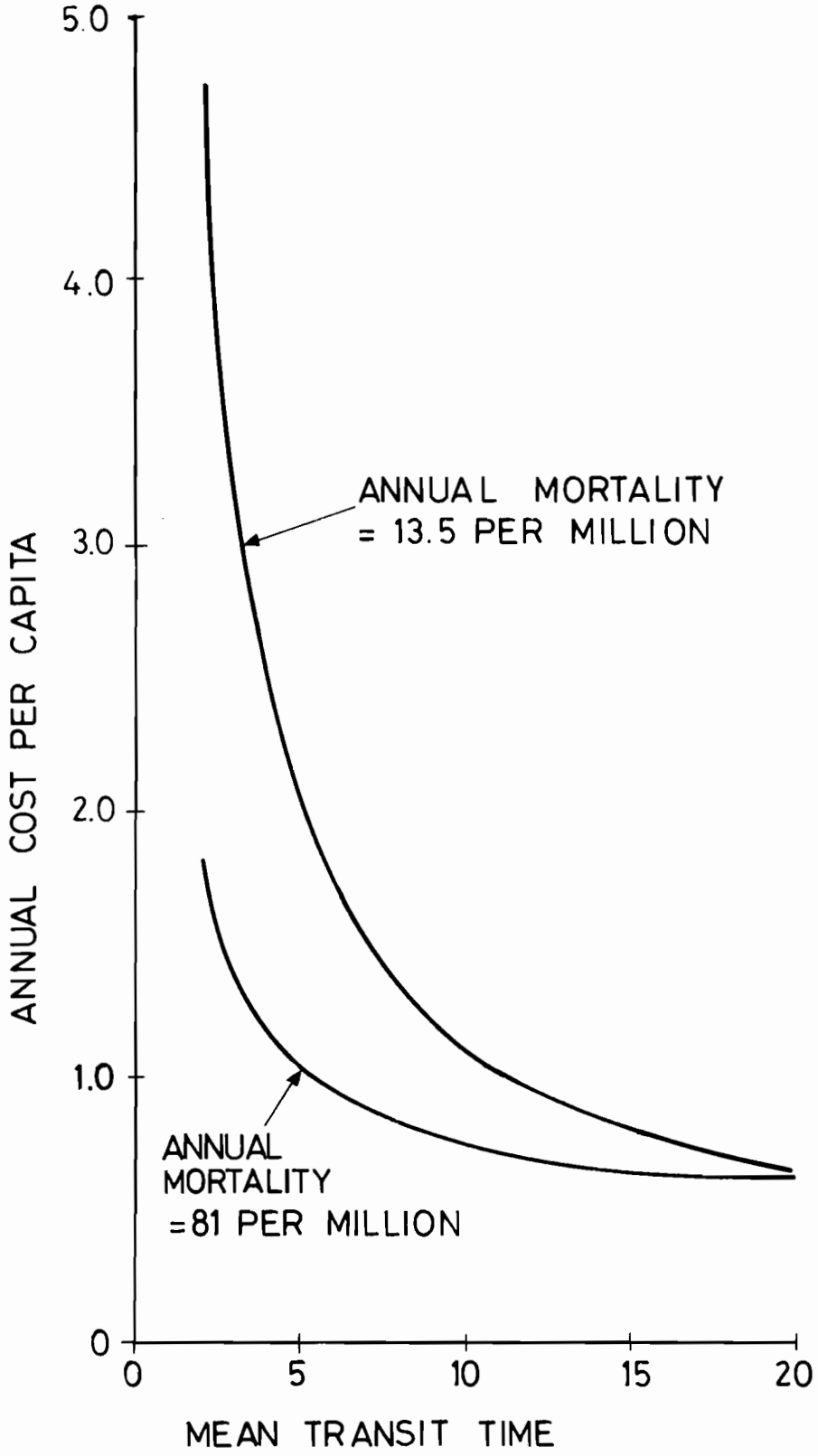


Figure 2. Effect on cost of mean transit time (coefficient of variation constant at 0.1).

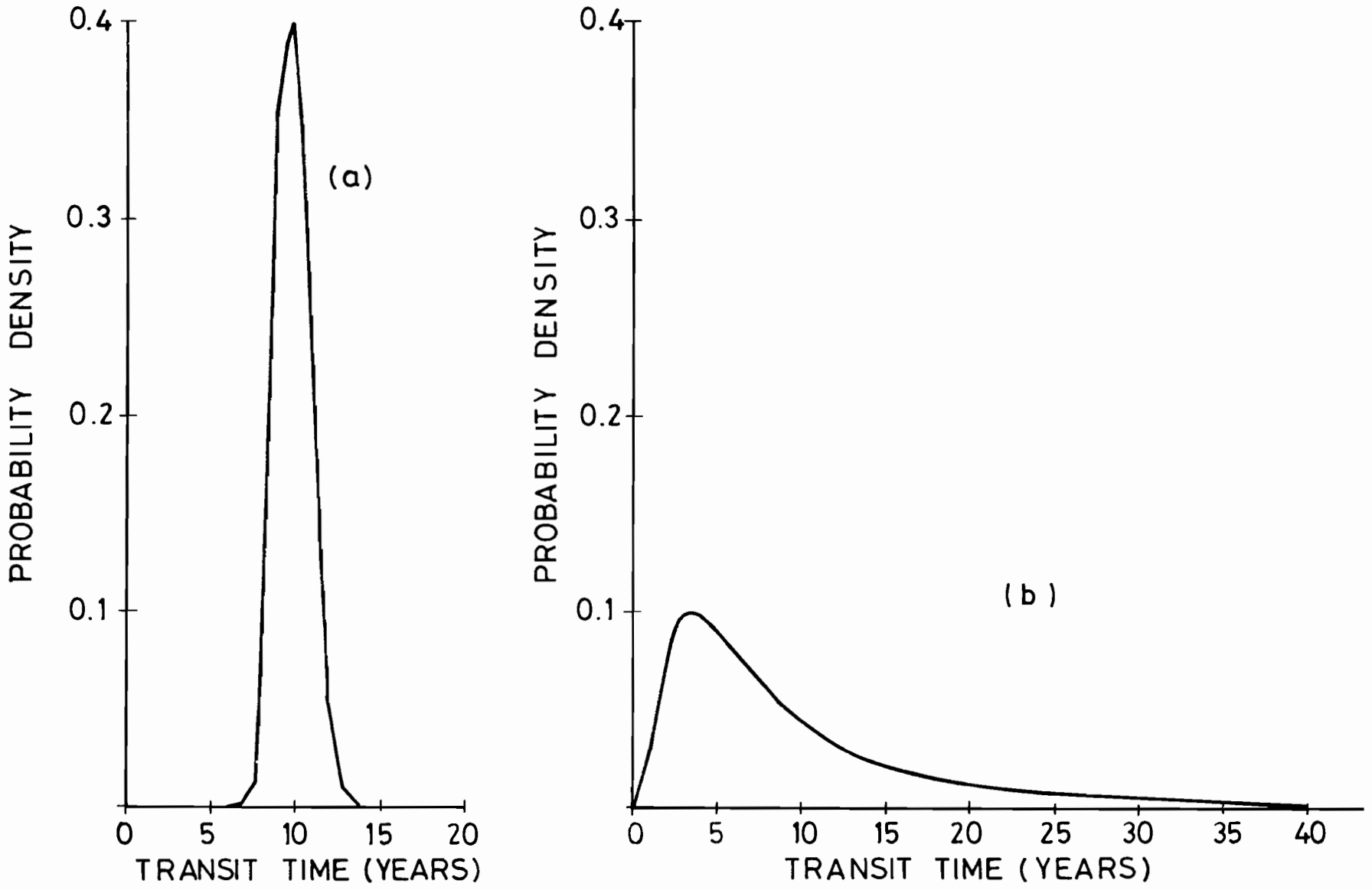
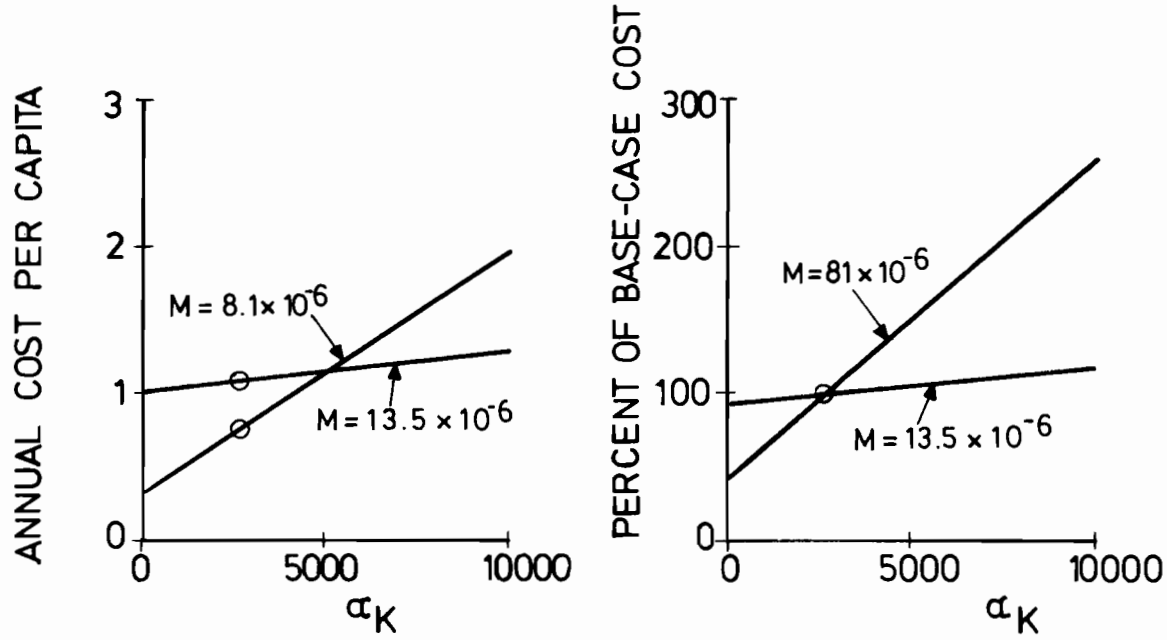
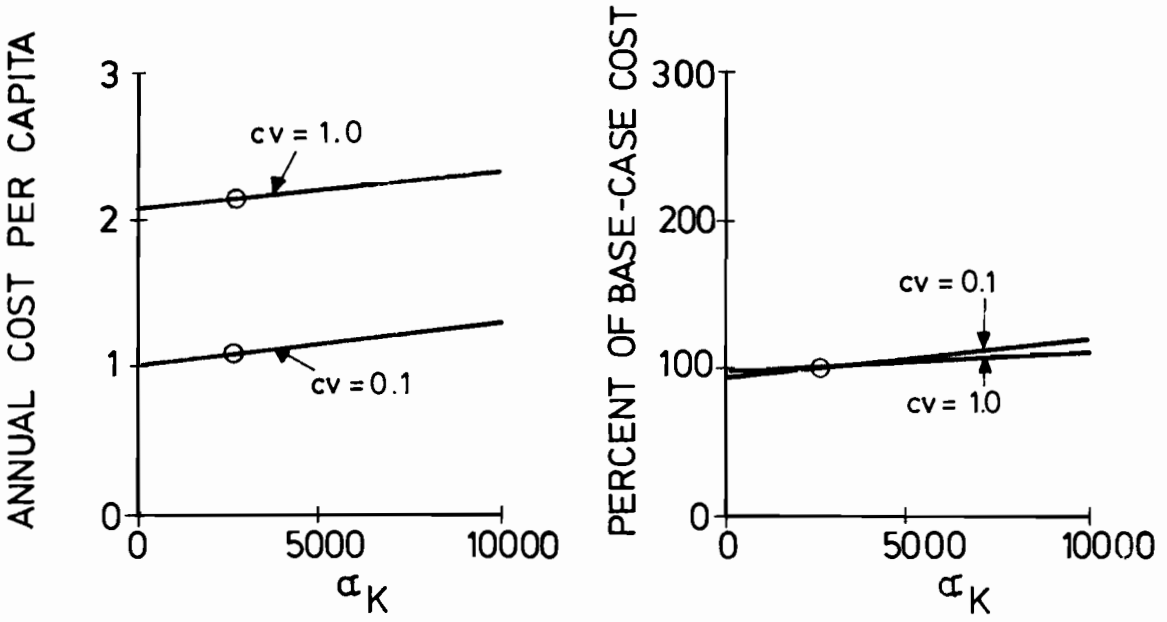


Figure 3. Transit time distributions. (a) c.v. = 0.1 (b) c.v. = 1.0

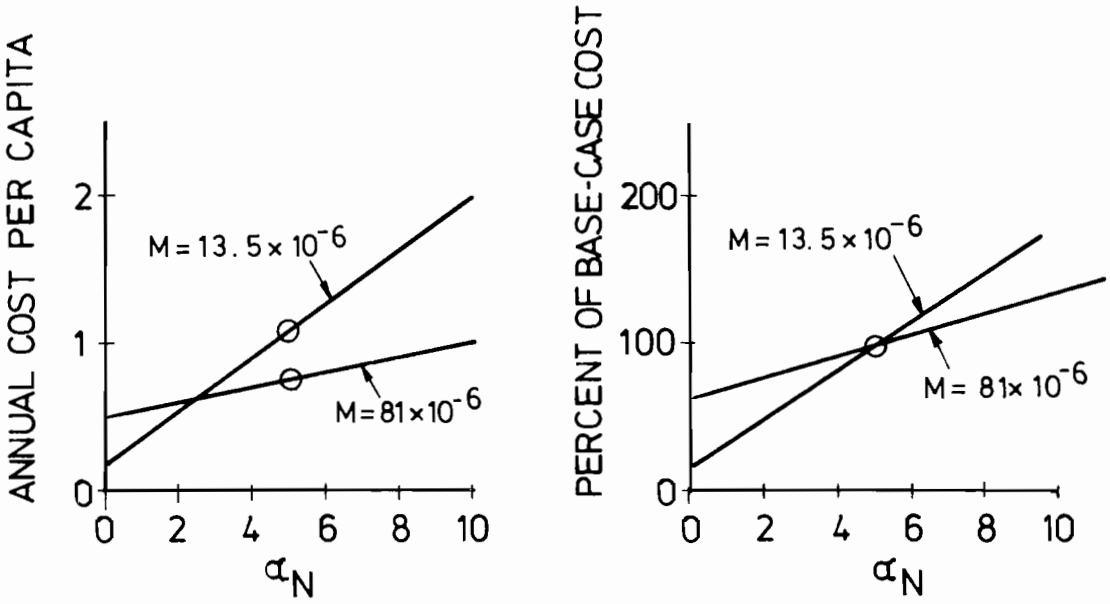


a) sensitivity to target mortality at constant coefficient of variation (cv = 0.1)

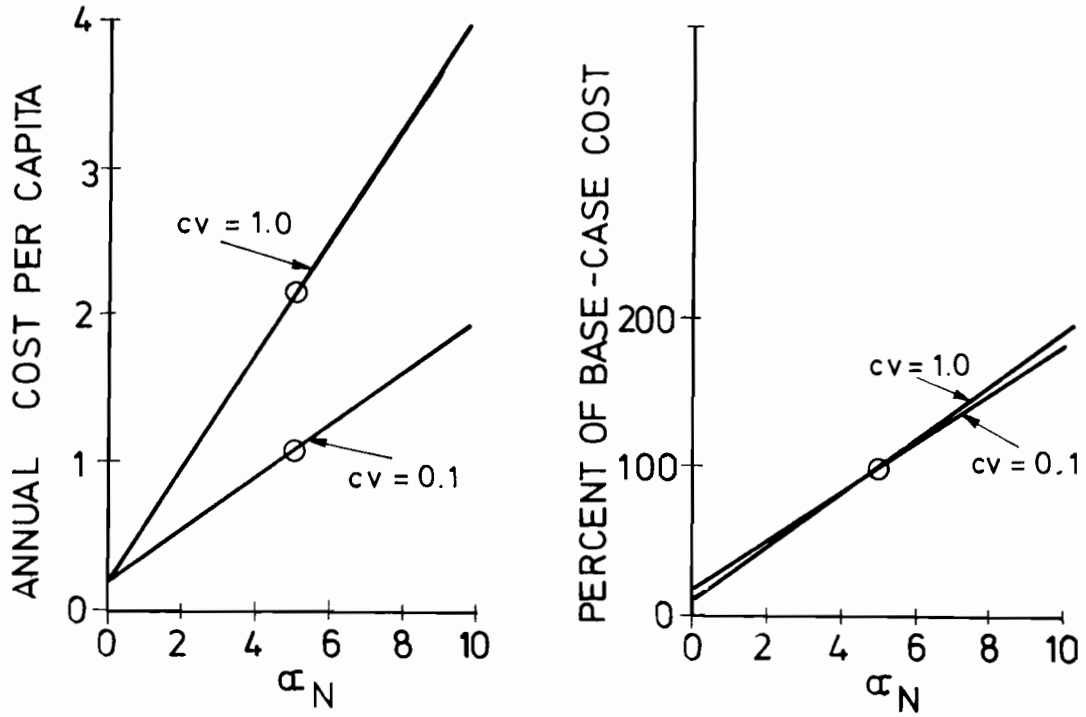


b) sensitivity to coefficient of variation at constant target mortality (M = 13.5 x 10^{-6})

Figure 4. Effect of late treatment cost on total cost.

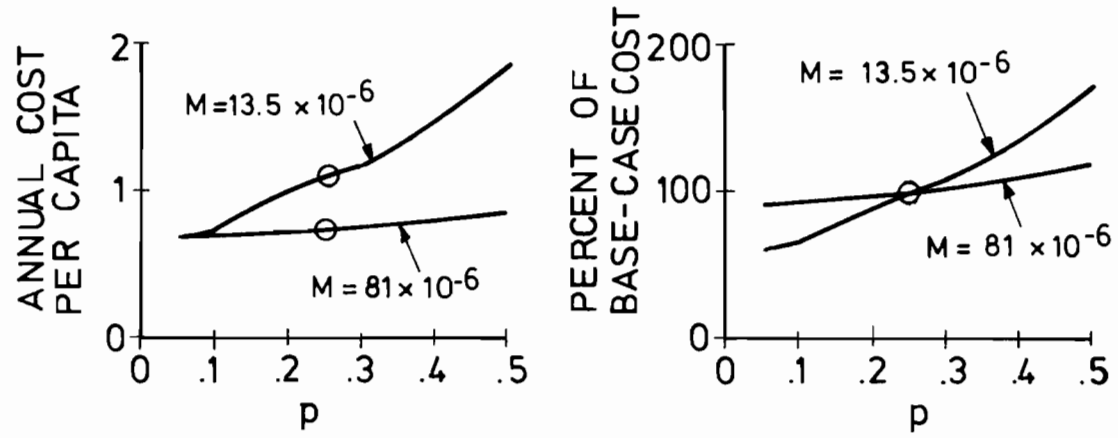


a) sensitivity to target mortality at constant coefficient of variation (cv = 0.1)

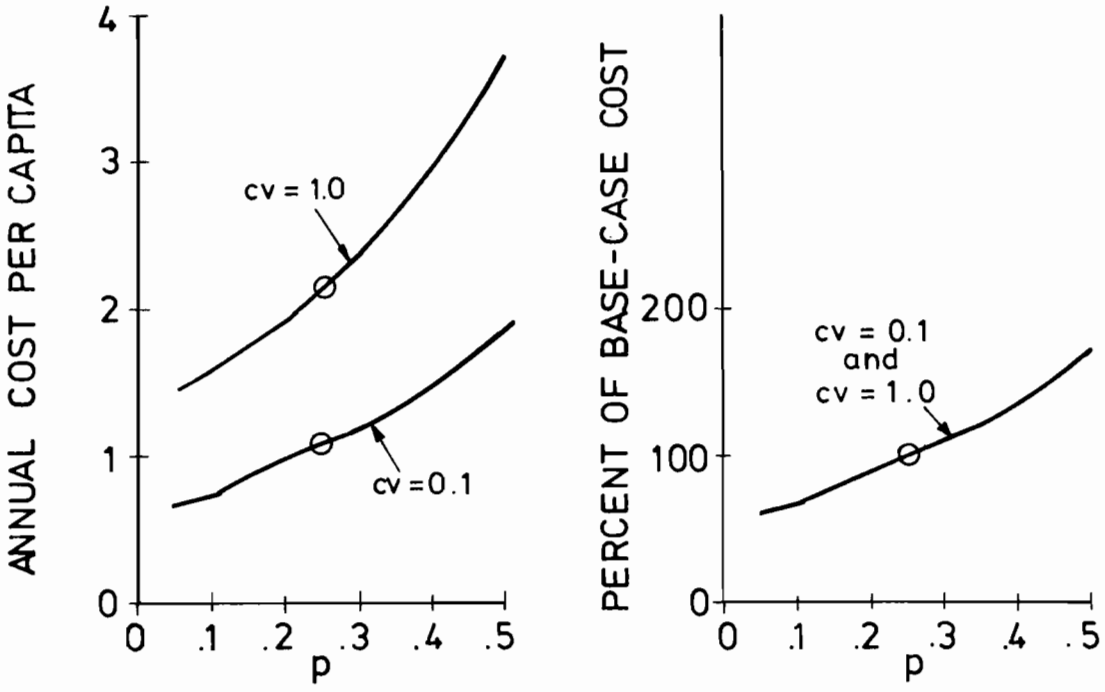


b) sensitivity to coefficient of variation at constant target mortality (M = 13.5 x 10^{-6})

Figure 5. Effect of cost per test on total cost.

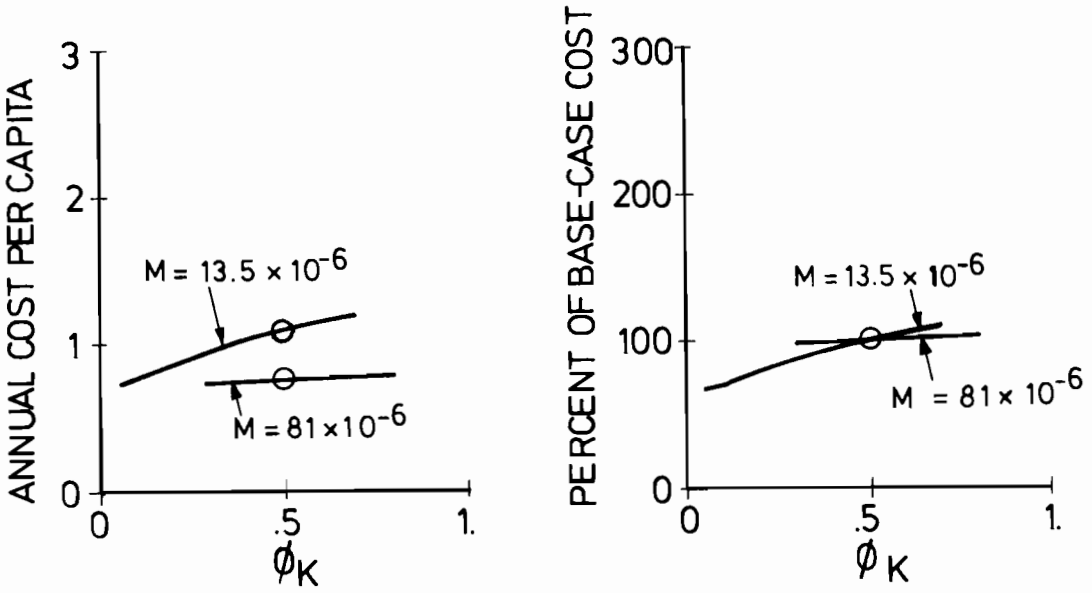


a) sensitivity to target mortality at constant coefficient of variation (cv = 0.1)

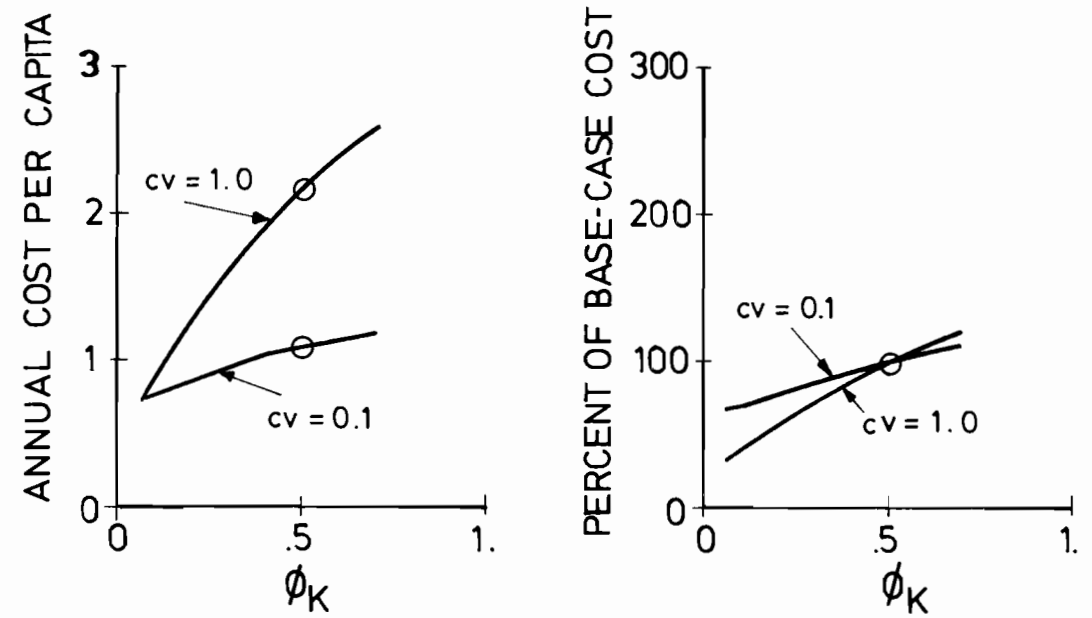


b) sensitivity to coefficient of variation at constant target mortality ($M = 13.5 \times 10^{-6}$)

Figure 6. Effect of changing the false-negative rate.

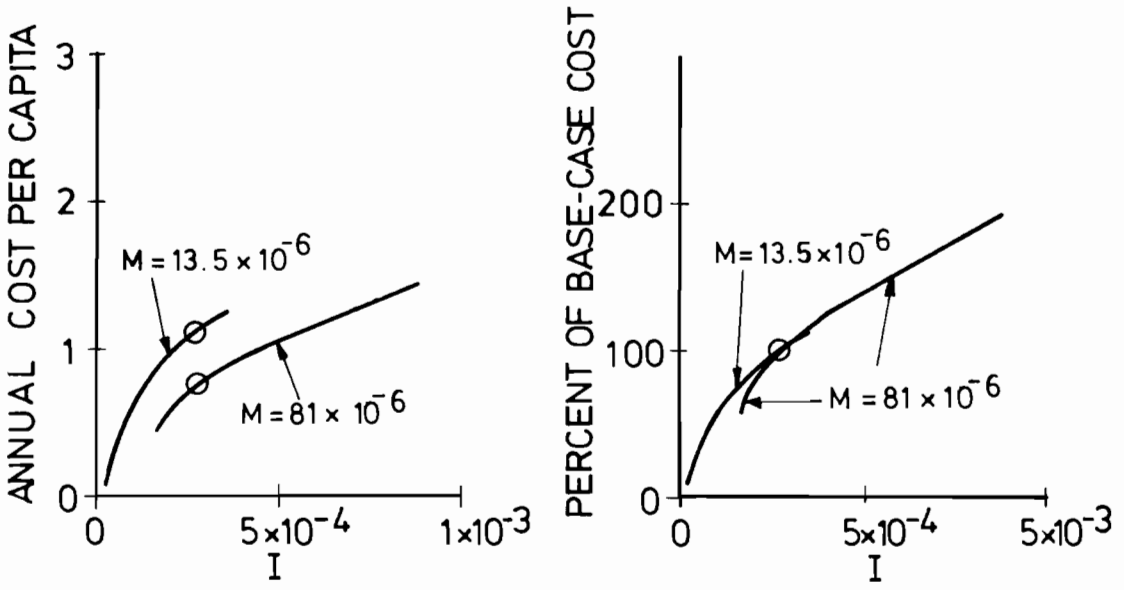


a) sensitivity to target mortality at constant coefficient of variation (cv = 0.1)

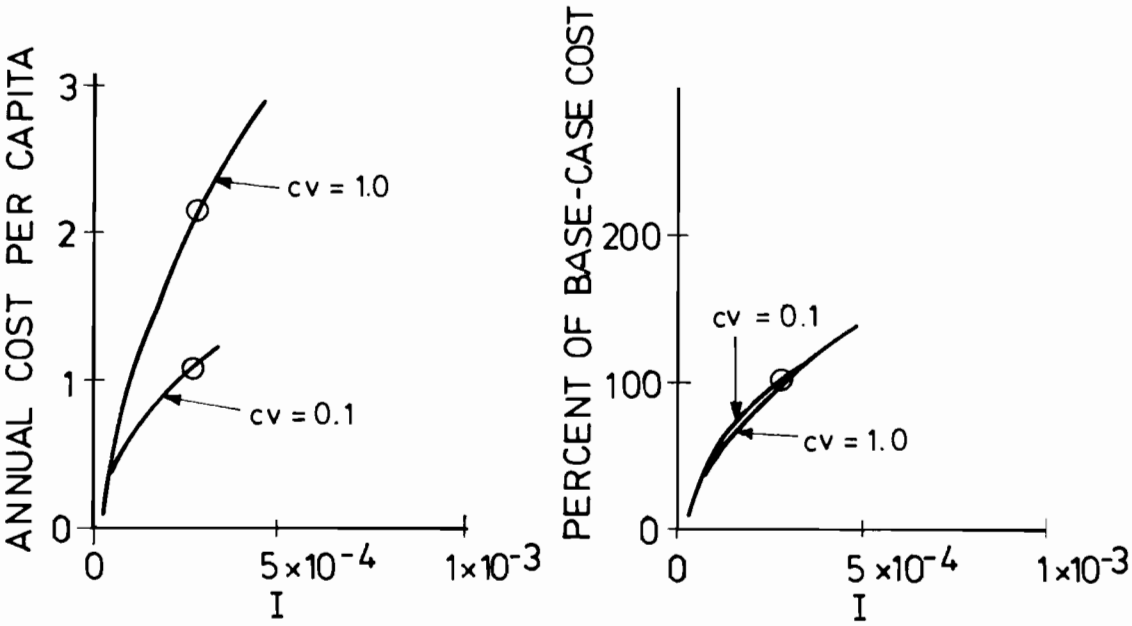


b) sensitivity to coefficient of variation at constant target mortality (M = 13.5 x 10^{-6})

Figure 7. Effect of changing mortality among late cases.

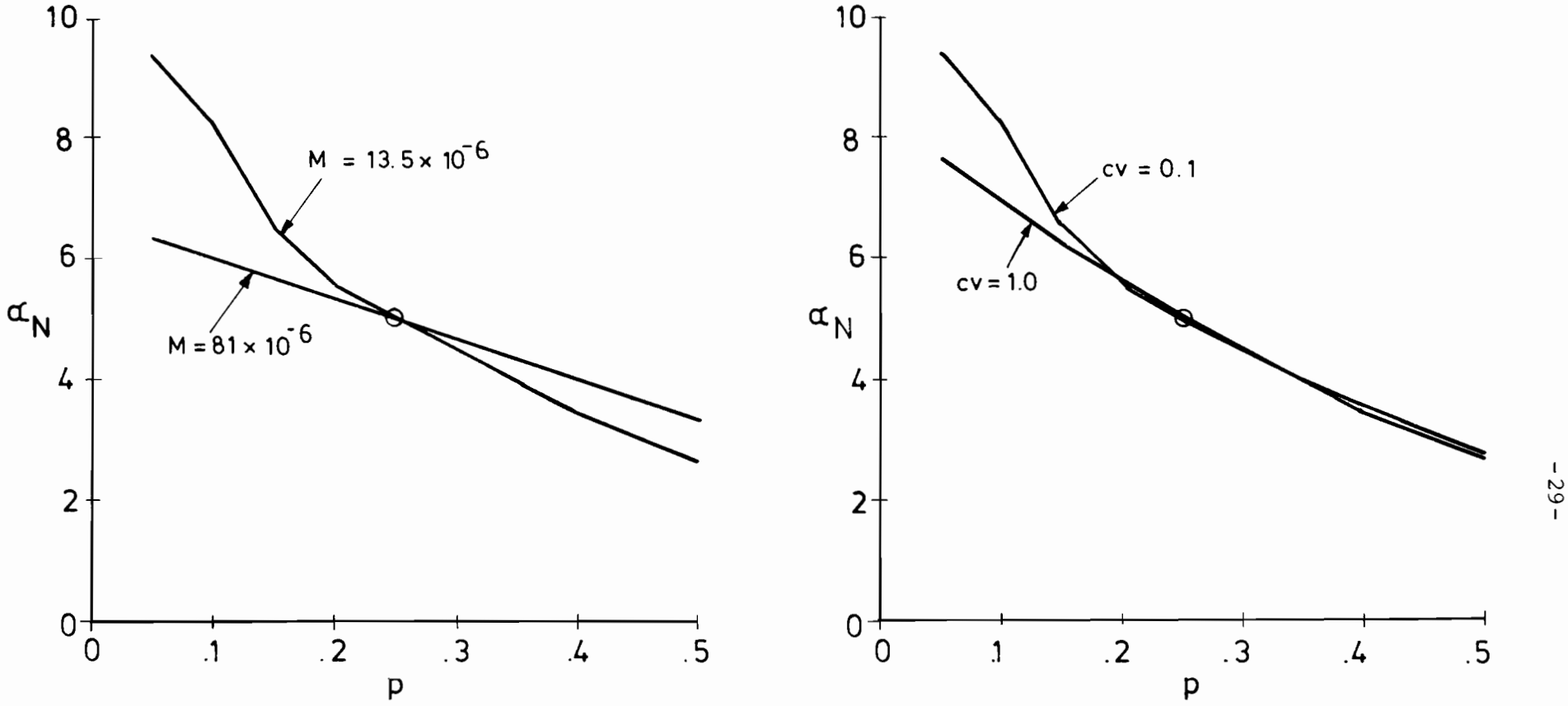


a) sensitivity to target mortality at constant coefficient of variation (cv = 0.1)



b) sensitivity to coefficient of variation at constant target mortality (M = 13.5 x 10^{-6})

Figure 8. Effect of changing the incidence.



(Left) sensitivity to target mortality at constant coefficient of variation (cv = 0.1)

(Right) sensitivity to coefficient of variation at constant target mortality (M = 13.5 x 10^{-6})

Figure 9. Trade-off between false negative rate (p) and cost per test (α_N).

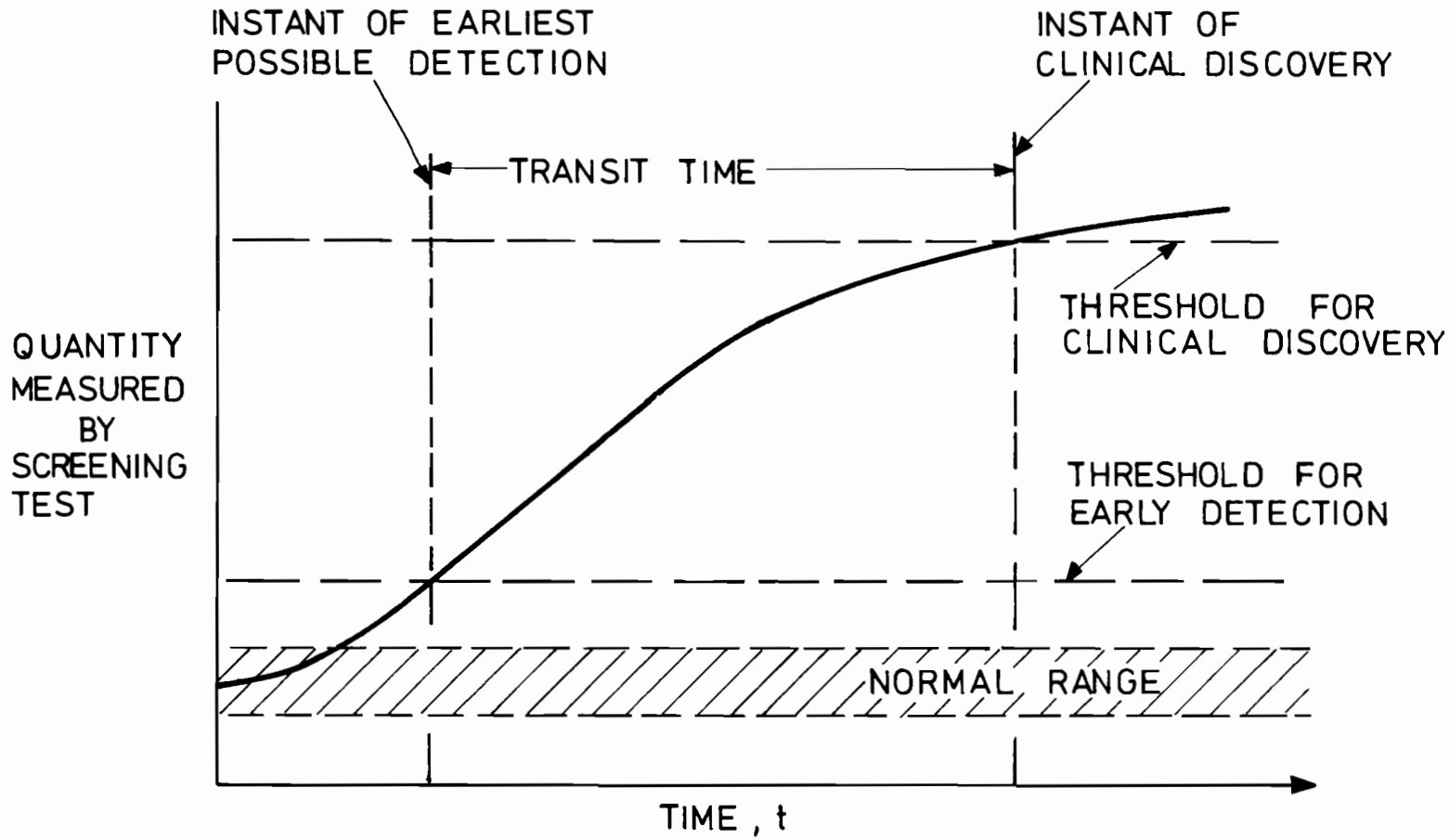
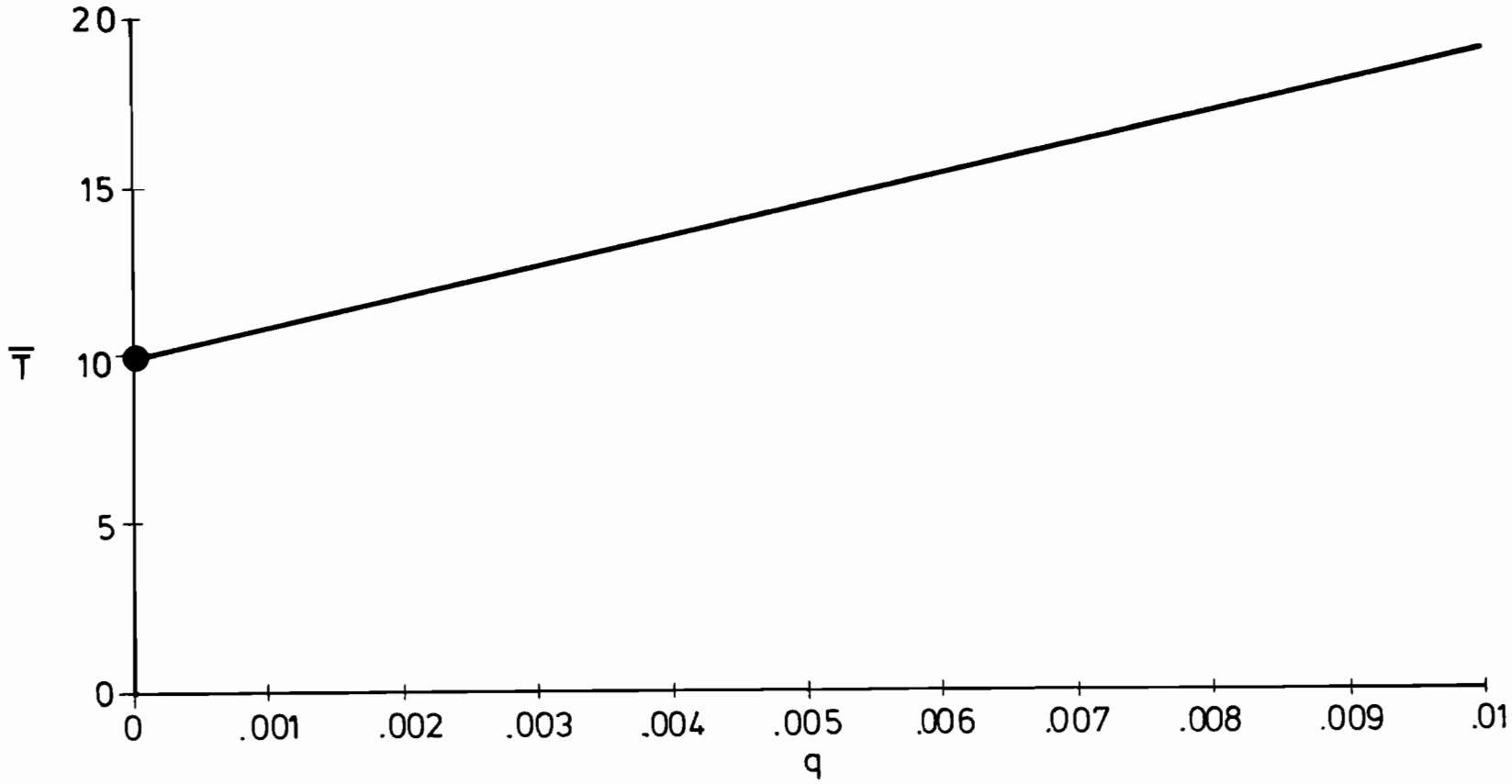
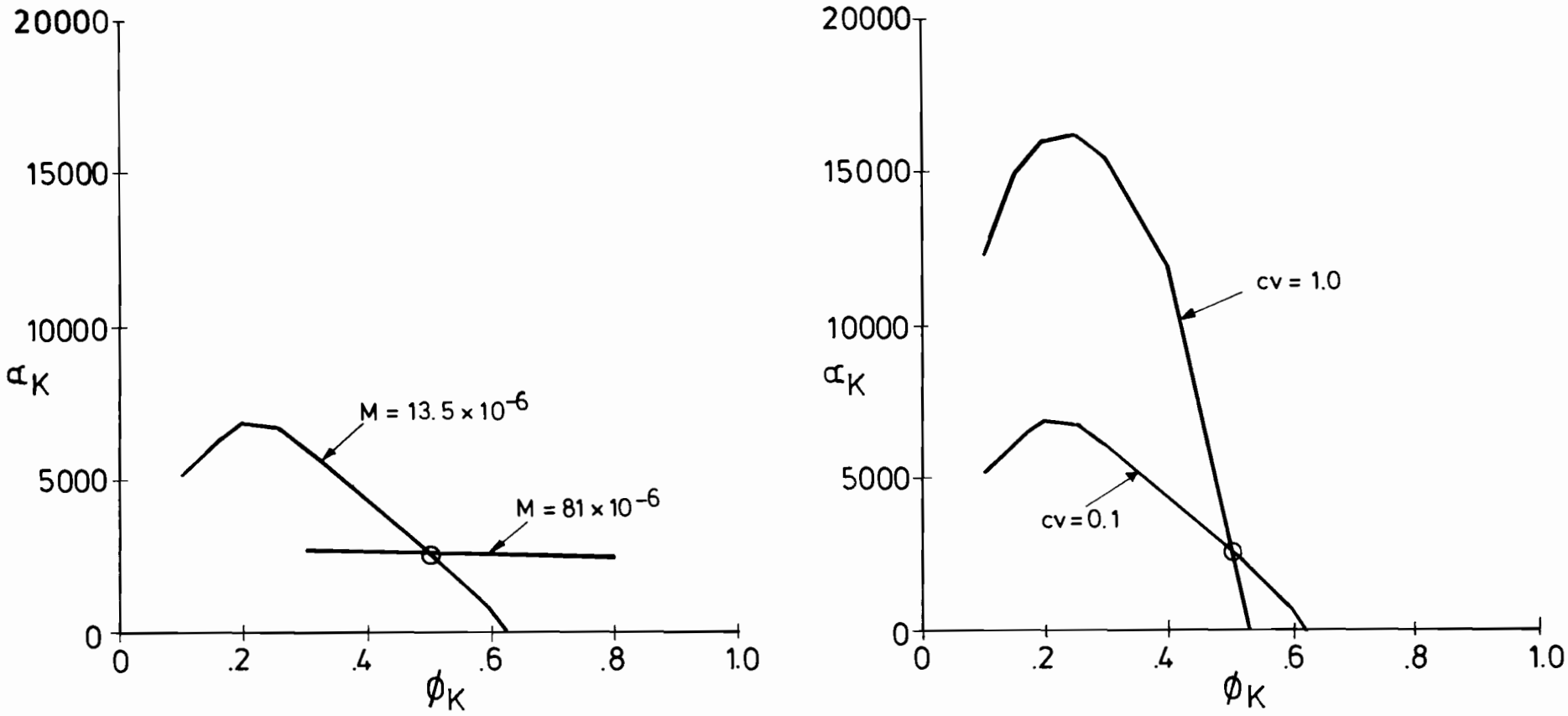


Figure 10. Interaction of disease and screening test.



(Note) This curve is not sensitive to changes in target mortality of coefficient of variation

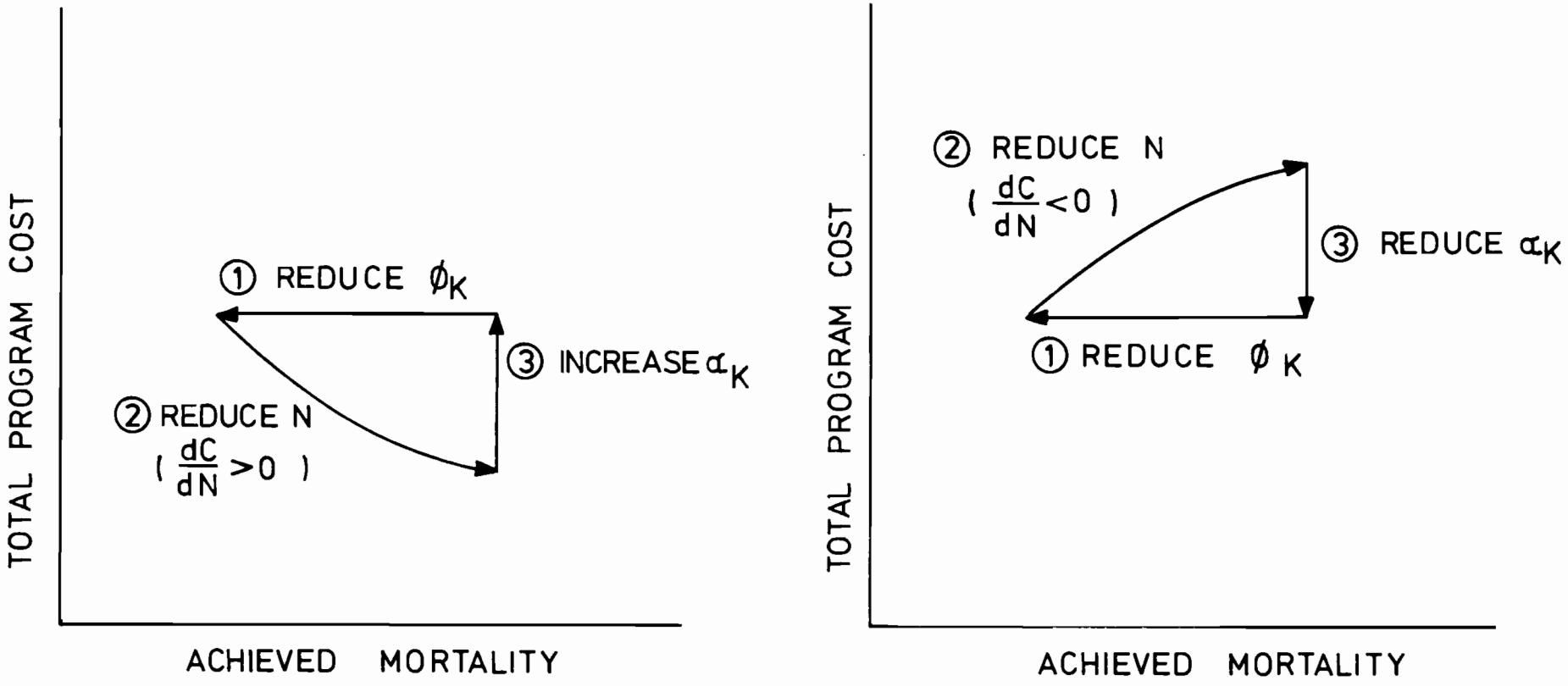
Figure 11. Trade-off between false-positive rate (q) and mean transit time ($\overline{T}$).



(Left) sensitivity to target mortality at constant coefficient of variation (cv = 0.1)

(Right) sensitivity to coefficient of variation at constant target mortality (M = 13.5 x 10^{-6})

Figure 12. Trade-off between mortality among late cases and cost of late treatment.



(Left) normal path, if cost (C) increases with number of tests (N).

(Right) anomolous path, if cost (C) decreases as number of tests (N) increases.

Figure 13. Calculation scheme for trade-off between ϕ_k and α_k.

References

[1] Beral, V. "Cancer of the Cervix: A Sexually Transmitted Infection?" The Lancet, 25 (1974), 1,037.

[2] Bigelow, J.H., and Ellis, D.B. "The False Negative Rate in Screening for Cervical Cancer." In Proceedings of the Joint IIASA/WHO Workshop on Screening for Cervical Cancer. Forthcoming.

[3] Boyes, D.A., Worth, A.J., and Fidler, H.K. "The Results of Treatment of 4,389 Cases of Preclinical Cervical Squamous Carcinoma." J. Obstet and Gynec. Brit. Cmlth., 77, 9 (1970), 769.

[4] Campbell, H. "Cancer of the Cervix - Survival: Comparative Study of 5-year Survival Rates from Cancer of the Cervix in 14 Countries from 1953-1957 Followed to 1962." J. Obstet and Gynec. Brit. Cmlth., 73, (1966), 27.

[5] Dunn, J.E., Slate, T.A., Mennit, J.W., and Montin, P.L. "Finding for Uterine Cancer from One or More Cytologic Examinations of 33,750 Women." J. Nat. Cancer Inst., 23, 3 (1959), 507.

[6] Fidla, H.K., Boyes, D.A., and Worth, A.J. "Cervical Cancer Detection in British Columbia." J. Obstet and Gynec, Brit. Cmlth., 75 (1968), 392.

[7] Green, G.H. "The Significance of Cervical Carcinoma In Situ." Am. J. Obstet and Gynec, 94, 7 (1966), 1,009.

[8] Knox, E.G. "Computer Simulation of Alternative Population Screening Policies." In Systems Aspects of Health Planning. Eds. N.T.J. Bailey and M. Thompson. Amsterdam, North-Holland Pub.Co., 1975, pp. 191-199.

[9] Schneider, J., and Twiggs, L.B. "The Costs of Carcinoma of the Cervix." Obstet and Gynec, 40 (1972).