

## SOME MATHEMATICAL MODELS IN HEALTH PLANNING

William P. Pierskalla<sup>1</sup>  
University of Pennsylvania

ABSTRACT. There are many areas in which mathematical models are useful in health care delivery. Two of these will be discussed here: 1) Diagnostic screening for early detection of disease, and 2) Planning regional blood banking systems.

Non contagious diseases arise in a population in a seemingly stochastic manner. If testing procedures exist which are capable of detecting the disease before it would otherwise become known, and if such early detection provides benefit, the periodic administration of such a test procedure to the members of the population, that is, a mass screening program, may be advisable. Moreover, if the population is composed of sub-populations which exhibit different disease incidence rates and different unit costs of test applications, and if different tests which have different reliabilities for detecting the disease are available, then the question of allocating limited screening resources among the sub-populations arises. The optimal allocation depends upon the form of the disutility functions of the sub-populations. Comprehensive analytic models are needed to perform this allocation.

Health planning can be viewed from many perspectives. Perhaps the most critical one facing the United States today is to contain the costs of health care and yet deliver quality care to the entire population of the U.S. Certain aspects of planning to achieve these objectives must be undertaken on a regional level, others at a sub-regional level, and still others at the institutional level. An integrated hierarchy of analytical models is needed to link the decisions at each of these levels. Decisions at the macro level involve the appropriate numbers of people by skills, numbers of facilities, and technological sophistication for a region. At the middle level, the decisions involve facility locations, their levels of technology and services and personnel needs to achieve minimum cost yet provide accessibility and quality of care in the sub-regions. At the institutional or micro level, analytic models are used to determine admissions and appointments, inventory levels and capital equipment, daily and weekly staffing, and facility scheduling. These different levels of modeling are illustrated in the context of planning for regional blood-bank systems.

---

<sup>1</sup>1980 Mathematics Subject Classification 90-02, 90B05, 90B25.

This work was partially supported by the National Science Foundation under grant ENG77-07463 and by Office of Naval Research Contract N00014-75-C-0797 (Task NR042-322).

## I. DIAGNOSTIC SCREENING FOR EARLY DETECTION OF DISEASE

### 1. INTRODUCTION

There are many situations where a defect can occur randomly among the members of a population -- either a population of human beings, inanimate objects, or perhaps livestock -- and once present, exist and develop, at least for a time, without any manifest symptoms. If the early detection of such a defect provides benefit, it may be worthwhile to employ a test capable of revealing the defect's existence in its earlier stages. (A defect, disorder, or disease will generally be referred to simply as a defect; and the word unit or individual will refer to a member of the population.)

Of course, continuous monitoring would provide the most immediate such revelation. But considerations of expense and practicality will frequently rule out continuous monitoring so that a schedule of periodic testing -- a screening program -- may be the most practical means of achieving early detection of the defect. In general terms, the question then becomes one of how best to trade off the expense of testing against the benefits to be achieved from detecting the defect in an earlier stage of development. The expense of testing increases both with the frequency of test applications and with the cost of the type of test used. The benefits increase with the frequency and the quality of test used and the quality of a test is often directly related to its cost.

The benefits of early detection also often depend upon the application considered. For example, in a human population being screened for some chronic disease (cancer, glaucoma, heart disease, etc.) the benefits of early detection might include an improved probability of ultimate cure, diminished time period of disability, discomfort, and loss of earnings, and reduced treatment cost. If the population being screened consists of machines engaged in some kind of production, the benefits of early detection might include a less costly ultimate repair and a reduction in the time period during which a faulty product is being unknowingly produced. If the population being screened consists of machines held in readiness to meet some emergency situation, an early detection of a defect would reduce the the time the machine was not serving its protective function. This process

of inspecting a sizable population for defects is called mass screening.

The expense of testing includes easily quantifiable economic costs such as those of the labor and materials needed to administer the testing. However, there can also be other important cost components which are more difficult to quantify. For example, in the case of a human population subject to medical screening, the cost of testing includes 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., from the cumulative effect of X-ray exposure.

The rationale for constructing a mathematical model of mass screening is to provide a conceptual framework within which a mass screening program might best be designed and its worth evaluated. Such a design must determine which kind of testing technology will be used. Several candidate technologies may be available, each with different reliability characteristics and costs. In addition, the frequency of testing must be decided. If the target population can be partitioned into sub-populations according to susceptibility to the defect (e.g., by age, family background, time since last overhaul (for a machine), etc.), then the best allocation of the testing budget among the sub-populations must be determined. Lastly, a decision must be made, for the population and type of defect under consideration, whether a mass screening program is justified at all. It is felt that the above determinations are best carried out within the conceptual framework of a cogent model of mass screening.

In the following sections a reasonably comprehensive model for decision making with regard to mass screening is presented. The model utilizes a time-based approach. But rather than seeking to analyze or minimize detection delay, as in some of the literature on this topic, the objective function is an arbitrary increasing function of detection delay. (Detection delay is the time between the incidence of the defect and its detection, regardless of whether that detection is the result of a screening test or of the defect becoming self-evident.) The reason for choosing a general function is that the disutility experienced upon the delayed detection of a defect may well vary in a highly nonlinear way with the length of the delay.

Another objective sometimes used for inspection models is to maximize the lead time where the lead time is defined as the difference between the time of detection via a screening test and the time detection would otherwise have occurred had not a screening program been in existence. However, in the case in which the test being used is perfectly reliable then it be-

comes possible to consider the lead-time criterion in terms of the following simple function of detection delay. Suppose that at the (possibly random) age  $T$  (measured from the time of incidence of the defect) the defect, in the normal course of its development, would become manifest even without a screening test. Then, if the defect is detected at age  $t$ , the lead-time gained is  $(T - t)^+$ .<sup>1</sup> Therefore, if  $\sup T$  is finite, the function  $D(t) = \sup T - E(T-t)^+$  may be used as a disutility function which, when minimized, will maximize the expected lead-time.

Consequently, the results discussed here based on an arbitrary disutility function, generalize as well as extend results in earlier work. In addition, new results are presented.

In the next section a brief review of the literature is given. Since the work on inspection models is very large, only the most relevant papers are discussed. The interested reader may refer to the surveys by McCall [1965] and Pierskalla and Voelker [1976] for a more comprehensive review.

In the third section the model is formally stated and an expression derived for the expected disutility per unit time incurred under a regime of uniformly spaced test applications.

This model can be specialized to the case of a perfect test; i.e., when the test is administered to an individual with the defect, the defect will be detected with certainty. Under this assumption, a regime of uniform test intervals is optimal within a wider class of "cyclic" testing schedules for a single sub-population. The problem of allocating a screening budget among various sub-populations can then be analyzed. As part of this analysis, the explicit screening schedule  $(r_1, r_2, \dots, r_Q)$ , where  $r_j$  designates the optimal testing frequency for members of the  $j$ th sub-population, can be obtained in terms of the sub-population-specific incidence rates  $N_j \lambda_j$  and the budget constraint for the disutility function  $D(t) = at^m$ . For  $D(\cdot)$  convex, the above solution (when  $m = 1$ ) provides a bound on the ratios  $r_i/r_j$ .

Other uses of the model involve the analysis of the case when the probability of detection is a constant over all values of elapsed time since incidence. The long-run expected disutility per unit time can be derived; its differential qualities (with respect to variations in the test reliability parameter and testing frequency) exhibited; and explicit solutions provided for various special forms of  $D(\cdot)$ . Also for special forms of  $D(\cdot)$ , decision rules can be presented to select between two different kinds

---

<sup>1</sup>The function  $u^+ = \max(u, 0)$ .

of tests which differ with respect to their reliabilities and cost per application.

If it is assumed that the test will detect the defect if and only if the elapsed time since incidence exceeds (or equals) a critical threshold  $\hat{T}$  which characterizes the test, an expression for the long-run expected disutility per unit time can be derived. Then decision rules can be developed to select between two alternative test types which differ with respect to their critical threshold  $\hat{T}$  and their cost per application. Some interesting examples for linear and quadratic disutility are given in the fourth section.

The last section in Part I provides a technique to estimate the shape of the disutility function from empirical data in the case of a perfect test.

Finally, it should be mentioned that the population (or each sub-population, in the instances where a heterogeneous population is considered) is assumed to be of fixed size,  $N$ , and the defect to arise according to a stationary Poisson process with rate  $N\lambda$  ( $N$  could be a very large but finite number). It may be somewhat more realistic to set the defect occurrence rate (sometimes called the defect arrival rate) proportional to the number of defect-free units, rather than to the total number of units in the population. However, it is also assumed that no defect can remain undetected longer than  $T^*$ , even without any screening tests being given. Hence,  $T^*N\lambda$  is an upper bound on the expected number of undetected defects in the population. If it is the case that once a defect is detected, the afflicted unit is replaced in the population with a healthy unit, then  $T^*N\lambda$  represents a bound on the expected difference between the number of healthy units and the total number of units in the population. For  $\lambda$  small, as would be the case for a relatively infrequently occurring disorder, this should represent no difficulty. Consequently, it is assumed that  $\lambda$  is small relative to  $N$ , which is consistent with the examples of potential applications which have been or will be mentioned.

## 2. LITERATURE REVIEW

Some of the early papers which have a bearing on time dependent models of mass screening are: Derman [1961], Roeloffs [1963, 1967], Barlow, Hunter, and Proschan [1963] and Keller [1974]. Kirch and Klein [1974], whose paper is addressed explicitly to a mass screening application, seek an inspection schedule which will minimize expected direction delay subject to a constraint on the expected number of examinations an individual would incur over a lifetime; the test is assumed perfectly reliable. The point of view adopted here in Part I is similar to that of Kirch and Klein. However, one respect in which the approach here differs from that of Kirch and Klein is that several sub-populations are considered each with its own characteristic

incidence rate for the disorder. The idea is then to allocate optimally a fixed screening budget among the sub-populations. Kirch and Klein instead take a longitudinal view. An individual through his lifetime is subject to different probabilities of incurring the defect and a screening schedule is optimized subject to a constraint on the expected number of examinations over a lifetime.

McCall [1969] considered the problem of scheduling dental examinations under the assumption that the time between the incidence of a cavity and the scheduled dental examination controls whether the cavity results in a filling or an extraction. Cavities are assumed to occur according to a Poisson process. As a generalization, he permits the time required for a cavity to become beyond repair (by a filling) to be a random variable.

Lincoln and Weiss [1964] studied 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,  $p(t)$ , is a function of the defect's age,  $t$ . They derive equations, similar to renewal type equations, which relate the density functions for the following entities: the probability of detection at a test application ( $p(t)$ ), the time until the defect becomes potentially detectable and from this time the forward recurrence time to the first test, the probability of the event that at a particular time a test occurs and all prior tests had failed to detect the defect, and the detection delay. For the two special cases where  $p(t)$  is a constant and where  $p(t)$  is exponential, the moments for the detection delay are derived in closed form. For uniform testing intervals (and general  $p(\cdot)$ ), the distribution and moments of the detection delay are computed. For  $p(t) = 1$ , they solve for that testing schedule which maximizes the time between tests subject to a constraint on the performance of the screening program relative to detection delay. Two such constraints are considered. The first bounds the probability of detection delay exceeding some threshold  $T$ . The second bounds the mean detection delay.

THEOREM I.3.1 in the next section, which gives the expected disutility in terms of the test interval and test type, is similar to the objective function given by Lincoln and Weiss although their approach, as outlined above, is different.

Several recent papers have studied the statistical characteristics of the lead-time provided by a screening program -- either a one-shot screen of the population or a periodic screening program. Some of these papers are: Hutchison and Shapiro [1968], Zelen [1971], and Prorok [1973].

In some recent work, Eddy [1978a, 1978b] describes a semi-Markov Chain analysis for the screening and detection of cancer. This modeling effort was

undertaken to devise a practical planning tool based in the reality of a breast cancer screening program.

Finally a few authors, Schwartz and Galliher [1975], Thompson and Disney [1976] and Voelker [1976] let both the reliability of the test and the disutility (or utility) of detection be a function of the defect's state rather than of time since the defect's incidence. Although such models are more general and do utilize a general concept of disutility, they have not been amenable to closed form evaluation of expected disutility.

To incorporate random defect occurrences into their models, previous researchers focus upon an individual who will incur the defect. They use the density function for the age when that individual incurs the defect as a fundamental element of their model. Since the density function reflects age-specific incidence rates, a "life time" testing schedule can, thereby, be developed to tailor testing frequency at each age to the probability that the defect will occur at the age.

Our way of modeling the randomness of these occurrences reflects a somewhat different perspective on the mass screening problem. We look through the eyes of a decision-maker charged with intelligently allocating a fixed budget. The time frame over which the allocation must be made is often short compared to a typical life time of a member of the client population. Therefore, the decision-maker does not plan lifetime screening schedules for particular individuals. Instead, he tries to maximize the benefit that can be derived from his available budget over a much shorter planning horizon.

With the problem viewed in this perspective, the random nature of defect occurrences is most naturally modeled as a Poisson process with its parameter determined by the incidence rate of the defect and the size of the population. This approach has proved particularly useful in the following context: if different segments of the client population exhibit different incidence rates, sub-populations can be defined with defect incidence within each of them modeled as a Poisson process with its respective parameter. Then the budget can be so allocated among the sub-populations as to permit appropriate relative testing frequencies (cf. Voelker and Pierskalla [1978]). In this way, age-specific incidence rates can be incorporated into the notion of Poisson defect occurrences. Moreover, factors other than age which affect defect incidence rates (family history, smoking habits, work environment, etc.) can also be incorporated into the model.

### 3. A MODEL FOR MASS SCREENING

Although most of the definitions are stated prior to their use in each section, listed below are some notation and conventions used as various times

throughout the paper.

$$(1) \quad \frac{l_m}{r}(t) = 1 \quad \left[ \frac{m}{r}, \frac{m+1}{r} \right] (t) = \begin{cases} 1 & \text{if } \frac{m}{r} \leq t \leq \frac{m+1}{r} \\ 0 & \text{otherwise} \end{cases}$$

$$(2) \quad \prod_{i=1}^n x_i = 1 \quad \text{for } n < 1$$

$$(3) \quad \sum_{i=1}^n x_i = 0 \quad \text{for } n < 1$$

(4)  $[x]$  is the largest integer not exceeding  $x$ .

(5) The phrase "increasing function" shall mean a strictly increasing function.

As mentioned previously, the objective function will not be to maximize expected utility, but rather to minimize expected disutility. The disutility associated with a particular detection will depend only on the elapsed time since the defect's incidence. The notation  $D(t)$  will express the disutility incurred if detection occurs  $t$  units of time after incidence.  $D(\cdot)$  is assumed throughout this paper to be a nonnegative increasing function.

In those cases where  $\sup_{s \geq 0} D(s) < \infty$ , there is a natural relation between the notion of a utility function and  $D(\cdot)$ ; namely,

$$U(t) = \sup_{s \geq 0} D(s) - D(t) . .$$

$U(t)$  is a decreasing function expressing the disutility avoided by detecting a defect  $t$  units of time after its incidence.

Initially, only a single susceptibility class of size  $N$  will be considered. The results will then be generalized to an arbitrary number of subclasses. The times of incidence for the defect in the population are assumed to form a Poisson process with parameter  $N\lambda$  and are designated by the sequence  $\{S^k\}$ ,  $k = 1, 2, \dots$ . This assumption is motivated by the fact that any occurrence or arrival process with the following characteristics is a Poisson process: at the time of an arrival there is almost surely (i.e., with probability one) only one arrival, that the number of arrivals in a time interval does not depend on past arrivals, and that the number of arrivals in intervals of equal length are identically distributed (Cinlar [1975]). For many applications such as the occurrence of diseases like cancer, heart trouble, etc., this



assumption is quite reasonable.

Were a screening test of type  $\ell$  to be administered to the individual with the  $k$ th defect (i.e.,  $k$ th in the order of defect incidence) at time  $S^k + t$ , then the test outcome random variable is:

DEFINITION:

$$Y_{\ell}^k(t) = \begin{cases} 0 & \text{if } k\text{th defect is not detected at time } t \\ & \text{after incidence} \\ 1 & \text{if } k\text{th defect is detected at time } t \text{ after} \\ & \text{incidence.} \end{cases} \quad (3.1)$$

Since a defect cannot be detected before its incidence,  $Y_{\ell}^k(t) = 0$  for  $t < 0$ . Notice that the argument of  $Y_{\ell}^k(\cdot)$  refers to time relative to the incidence of the defect. If the population is screened at time  $s$ , where  $s$  is absolute time, then the  $k$ th defect will be detected if and only if  $S^k \leq s$  and  $Y_{\ell}^k(s - S^k) = 1$ .

It is assumed that  $Y_{\ell}^k(t)$ ,  $Y_{\ell}^j(s)$  are independent except when  $k = j$  and  $t = s$ , and that  $P(Y_{\ell}^k(t) = 1)$  depends only on  $\ell$  and  $t$ . This latter assumption makes possible the

DEFINITION:

$$p_{\ell}(t) = P(Y_{\ell}^1(t) = 1). \quad (3.2)$$

The function  $p_{\ell}(\cdot)$  describes the reliability characteristics of the type  $\ell$  screening test, i.e., how likely such a test is to detect a defect as a function of the defect's age. Note that  $p_{\ell}(t) = 0$  for  $t < 0$ .

In this section the screening test is assumed to be administered to the entire population at the times  $1/r, 2/r, 3/r, \dots$ . The testing frequency  $r$  is a control variable. In the next section, upon assuming perfect test reliability, it is shown that the above schedule of uniform testing intervals is optimal within a wider class of "cyclic" schedules.

The random variable  $\bar{S}_{r,\ell}^k$  denotes the time at which the  $k$ th defect is detected.  $\bar{S}_{r,\ell}^k$  depends on the arrival time of the defect ( $S^k$ ), the type of test used ( $\ell$ ), and the testing frequency ( $r$ ).

DEFINITION:

$$\bar{S}_{r,\ell}^k = \min_{n=1,2,\dots} \{n/r | Y_{\ell}^k(\frac{n}{r} - S^k) = 1\} \quad (3.3)$$

Given the application of test type  $\ell$  at the times  $\{1/r, 2/r, \dots\}$ , the disutility incurred by the  $k$ th defect is  $D(\bar{S}_{r,\ell}^k - S^k)$ . The total disutility incurred due to those defects which occurred in the interval  $[j/r, (j+1)/r]$ , when test  $\ell$  is used is:

DEFINITION:

$$B_{r,\ell,j} = \sum_{k=1}^{\infty} D(\bar{S}_{r,\ell}^k - S^k) 1_{j/r}(S^k).$$

The following proposition provides an expression for  $E[B_{r,\ell,j}]$  in terms only of the disutility due to detection delay (as expressed by  $D(\cdot)$ ) and of the reliability of test type  $\ell$  (as expressed by  $p_{\ell}(\cdot)$ ). In addition, the theorem shows that  $E[B_{r,\ell,j}] = E[B_{r,\ell,0}]$  for  $j = 0, 1, 2, \dots$

THEOREM I.3.1:

$$E[B_{r,\ell,j}] = N\lambda \sum_{n=1}^{\infty} \int_{\frac{n-1}{r}}^{\frac{n}{r}} D(u) p_{\ell}(u) \prod_{m=1}^{n-1} [1 - p_{\ell}(u - \frac{m}{r})] du$$

for  $j = 0, 1, 2, \dots$

This proposition yields a relatively simple expression for the expected total disutility in any interval of length  $1/r$  and using test type  $\ell$  with probability of detection  $p_{\ell}(\cdot)$ . This expectation is used in the objective function of a mathematical program to determine the optimal testing frequency for a mass screening program for a heterogeneous population. In order to develop this mathematical program, the following definitions are useful.

DEFINITION:

$$\bar{B}_{r,\ell} = \lim_{n \rightarrow \infty} \frac{r}{n} \sum_{j=0}^{n-1} E[B_{r,\ell,j}].$$

$\bar{B}_{r,\ell}$  is the long-run expected disutility per unit time given the testing frequency  $r$  and the test type  $\ell$ . (The factor  $r$  enters the definition to convert disutility per unit testing-interval into per unit time.) By Theorem I.3.1,  $E[B_{r,\ell,j}] = E[B_{r,\ell,0}]$  for  $j = 0, 1, 2, \dots$ . Therefore,

$$\begin{aligned}\bar{B}_{r,\ell} &= rE[B_{r,\ell,0}] \\ &= rN\lambda \sum_{n=1}^{\infty} \int_{\frac{n-1}{r}}^{\frac{n}{r}} D(u) p_{\ell}(u) \prod_{m=1}^{n-1} [1 - p_{\ell}(u - \frac{m}{r})] du.\end{aligned}\quad (3.4)$$

Notice that if test type  $\ell$  provides perfect reliability, i.e.,  $p_{\ell}(t) = 1$  for  $t \geq 0$ , then (3.4) gives

$$\bar{B}_{r,\ell} = rN\lambda \int_0^{\frac{1}{r}} D(u) du. \quad (3.5)$$

Now consider the problem of selecting testing frequencies and test-types for each of  $Q$  different susceptibility classes which together comprise the whole population. These classes may differ from one another in the number of units they contain, in their defect incidence intensity, and in the cost per test application to an individual for a particular type of test. The sub-populations, however, are assumed to share a common  $D(\cdot)$  function.

Using (3.4) the expected long-run disutility per unit time for sub-population  $j$  with frequency  $r(j)$  and test  $\ell(j)$  where  $j = 1, 2, \dots, Q$  is given by:

$$\bar{B}_{j,r(j),\ell(j)} = r(j)N_j\lambda(j) \sum_{i=1}^{\infty} \int_{\frac{i-1}{r(j)}}^{\frac{i}{r(j)}} D(u) p_{\ell(j)}(u) \prod_{m=1}^{i-1} [1 - p_{\ell(j)}(u - \frac{m}{r(j)})] du.$$

This expression can be used to formulate a multi-sub-population screening problem subject to a budget constraint as follows:

$$\text{Minimize } (r(1), \dots, r(Q), \ell(1), \dots, \ell(Q)) \sum_{j=1}^Q \bar{B}_{j,r(j),\ell(j)} \quad (3.6)$$

such that

$$\sum_{j=1}^Q N_j c_{j,\ell(j)} r_j \leq b \quad (3.7)$$

$$r_j > 0 \quad j=1, \dots, Q \quad (3.8)$$

$$\ell(j) \in \mathcal{L} \quad j=1, \dots, Q \quad (3.9)$$

where

- $c_{j,\ell(j)}$  = cost per application of a test of type  $\ell(j)$  to an individual of sub-population  $j$ ,  
 $N_j$  = number of units (or individuals) in sub-population  $j$ ,  
 $b$  = budget per unit time.  
 $\mathcal{L}$  = set of all feasible tests .

In order to make this mathematical program even more comprehensive, it is possible to add constraints on the amount of testing labor available and on the capacity of the testing facilities in terms of the number of arrivals, the frequency of testing and the type of tests used. In addition, the cost of false positives can be included as a part of the test costs in inequality (3.7).

For example, a constraint on the total labor available is:

$$\sum_{j=1}^Q N_j \delta_{j,\ell(j)} r_j \leq L_\ell \quad (3.10)$$

and on the total testing facilities available is:

$$\sum_{j=1}^Q N_j f_{j,\ell(j)} r_j \leq F_\ell \quad (3.11)$$

for each type of test  $\ell(\cdot)$  used over the sub-populations being tested, where

- $\delta_{j,\ell(j)}$  = amount of labor needed to administer test type  $\ell(j)$  to an individual of population  $j$ ,  
 $f_{j,\ell(j)}$  = amount of testing facility time needed to administer test type  $\ell(j)$  to an individual of sub-population  $j$ ,  
 $L_\ell$  = total amount of labor available to administer test type  $\ell(\cdot)$  per unit time,  
 $F_\ell$  = total amount of facility time available to administer test type  $\ell(\cdot)$  per unit time.

#### 4. SOME EXAMPLES OF DISUTILITY FUNCTIONS

Suppose a production process is subject to a randomly occurring defect. Although production appears to proceed normally after the incidence of the defect, the product produced is, thereafter, defective to an extent which remains constant until the production process is returned to its proper mode

of operation. The only way to learn if the production process is in this degraded state is to perform a costly test. Now, if a test detects the existence of the degraded mode of production  $t$  units of time after its incidence, the harm done will be proportional to the amount of defection product (unknowingly) produced which, in turn, is proportional to  $t$ . Hence,  $D(t) = at$  for some  $a > 0$ .

Another example where a linear  $D(\cdot)$  function may be appropriate would be for the periodic inspection of an inactive device (such as a missile) stored for possible use in an emergency. If  $t$  is the time between the incidence of the disorder and its detection, the disutility incurred is proportional to the probability that the device would be needed in that time interval. If such "emergencies" arise according to a Poisson process with rate  $\mu$ , then the probability of an emergency in a time interval of length  $t$  is  $1 - e^{-\mu t}$ , which, for  $\mu$  small, is approximately  $\mu t$ . Hence, if  $b$  is the cost incurred should there be an emergency while the device is defective, and if  $\mu$  is the (small) arrival rate of emergencies, then  $D(t) = b\mu t$ .

A quadratic disutility could arise in the following situation. Suppose the magnitude of a randomly occurring defect increases linearly with time since the occurrence of the defect. For example, the magnitude of the defect might be the size of a small leak in a storage container for a fluid, and as fluid escapes, the leak gets larger. Further, suppose that the harm done accumulates at a rate proportional to the magnitude of the defect. Hence, the quantity of fluid lost (at least initially) increases the longer the defect exists, and the rate of fluid loss is proportional to the size of the leak.

Let the size of the leak (as measured by rate of fluid loss), at a time  $s$  since the leak's incidence, be  $cs$ . Then, if the defect is detected at time  $t$  since incidence, the disutility incurred (fluid lost) is  $D(t) = \int_0^t cs ds$   
 $= 1/2 ct^2$

##### 5. EMPIRICAL ESTIMATION OF $D(\cdot)$

Although data may not be available currently to support the utilization of particular models, that should not deter the development of such models. It is reasonable to expect that in many application areas in the future much additional data will become available; for example, more data will become available concerning the stochastic pattern of a particular disease's development. Furthermore, a good model will serve as a guide to the kinds of data which should be gathered.

It is also reasonable to expect the future development of improved testing technologies capable of detecting a disorder in a much earlier stage of development than is now the case. Such innovations will make screening

programs more attractive and, consequently, make more important the analytical tools to design such programs intelligently.

However, it is possible to provide a simple procedure to estimate empirically the disutility function  $D(\cdot)$  under the assumption of perfect detection. Such a procedure is necessary, since upon the detection of a disorder there may be no way of directly determining how long the disorder has been present, although that length of time could not exceed  $x = \frac{1}{r}$ . It is assumed, however, that at each detection of a defect, the degree of disutility incurred due to that defect can be observed. For example, in the case of medical screening, at the detection of a tumor its degree of development can be noted even if it is not possible to determine exactly when, since the last test, the tumor originated.

In order to allocate optimally a screening budget among differing sub-populations (susceptibility classes), as was discussed in Section 3, it is necessary to know the shape of the disutility function. Under the reasonable assumption that  $D(\cdot)$  is an increasing continuous function, the function may be derived in the following manner. For a particular population subject to a Poisson defect arrival  $\{S^k\}$  with rate  $N\lambda$ , arbitrarily select a value for  $x$  and set up a prototype mass screening program for the population in which tests are made (using a perfect test) at times  $0, x, 2x, \dots$ . The manner of defects detected at the time  $jx$ , which are characterized by a disutility less than or equal to  $y$ , record and designate by  $Q_j(y)$ . Then,

$$Q_j(y) = \sum_{k=0}^{\infty} 1_{[0,y]}(D(jx - S^k)) 1_{[jx-x, jx]}(S^k), \quad j=1,2,3,\dots$$

$Q_j(y)$  is observable for all values of disutility  $y > 0$ . Further,  $Q_i(y)$  and  $Q_j(y)$  are independent and identically distributed random variables for  $i \neq j$  and  $y > 0$ . Therefore, by the law of large numbers,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n Q_j(y) = E[Q_1(y)] \quad \text{for all } y > 0.$$

Hence, the function  $y \rightarrow E[Q_1(y)]$  may be estimated by collecting a sufficiently large sample of the functions  $Q_1, Q_2, \dots$ .

Let  $H(y) = E[Q_1(y)]$ . Once  $H(\cdot)$  is estimated in this manner, the following theorem shows how the desired function  $D(\cdot)$  may be obtained from  $H(\cdot)$ .

THEOREM I.3.2: If  $D(0) = 0$  and  $D(\cdot)$  is continuous and increasing on  $[0, x]$ . then

$$D(t) = H^{-1}(N\lambda t) \quad \text{for } 0 \leq t \leq x .$$

By observing the  $Q_j(y)$  and taking the inverse of their expectation, an estimate for  $D(t)$  may be obtained. For some diseases such as heart disease, glaucoma, and some cancers, there may be enough data currently available to begin estimating the disutility functions and to start computing optimal testing frequencies and tests.

## II. PLANNING REGIONAL BLOOD BANKING SYSTEMS

### 1. A HIERARCHY OF PLANNING MODELS

Although health planning involves many activities and mathematical modeling aspects, we will only discuss hierarchies of models to link decisions at the regional, sub-regional and institutional levels. Decisions at the macro level involve the appropriate numbers of people by skills, numbers of facilities, and technological sophistication for a region. At the middle level, the decisions involve facility locations, their levels of technology and services and personnel needs to achieve minimum cost yet provide accessibility and quality of care in the sub-regions. At the institutional or micro level, analytic models are used to determine admissions and appointments, inventory levels and capital equipment, daily and weekly staffing, and facility scheduling. At all levels it is necessary to make these decisions by trading off competing objectives such as minimizing costs, increasing quality of care, and increasing accessibility and availability of services.

Rather than discuss these modeling activities in an abstract form, which could easily be done, we will focus on regionalization in blood banking.

Blood banks are an important and integral part of health service systems. Their main functions are blood procurement, processing, cross-matching,<sup>2</sup> storage, distribution, recycling, pricing, quality control and outdating. The large blood banks are often also responsible for blood research, disease and

---

<sup>2</sup>Cross-matching is the procedure of testing the donor's blood with a sample of blood from a potential recipient (patient) to determine whether the two types of blood are compatible and therefore will not lead to medical complications when the patient is transfused.

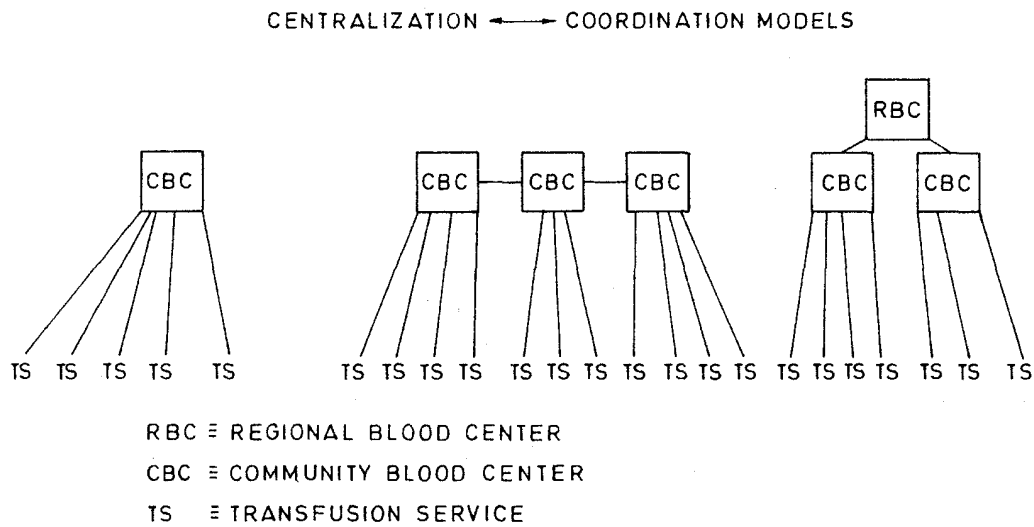
reaction prevention. In recent years, there has been much discussion on the issue of regionalization of blood banking systems, in the hope of decreasing shortages, outdates and operating costs, without sacrificing blood quality, research and education.

In a broad sense, regionalization is a process by which blood banks within a given geographical area move toward the coordination of their activities. Such coordination may range from cases in which the blood banks merge into a large, centralized unit, to cases where the existing structure remains unaltered and only certain functions, such as donor recruitment, processing and distribution, are coordinated among the blood banks. In most of these cases, questions of optimal region size, central and local bank locations, regional boundaries, optimal distribution and communication network configurations must be answered. Also, administrative policies, ordering and cross-matching policies, and donor recruitment and component therapy strategies must be analyzed and coordinated.

In an earlier paper on regionalization, (Cohen [1975]), hierarchical structures for different types of regional blood banking systems were discussed. The appropriate structure for any area depends on a complex interaction between the level of activity, the economies or diseconomies of scale, the cost effectiveness and efficiency, as well as the interactions of the interested parties. In this section, the impact of size and structure on the different costs of operation of blood banking activities in a region will be analyzed. The main determinants of cost are the personnel, the space, the equipment, the location and allocation of facilities and the transportation used to carry out the activities of the blood center.

Since regions vary in terms of their geography, number of donors, and number of recipients of blood services, the most effective and efficient organizational structure will also vary from region to region. It is not our purpose here to delve into all the many different ramifications of the different regional structures. Rather for this analysis it is sufficient to consider three generic structures of regional blood banking in the United States. Most systems which do not exactly fit into one of these structures can be closely approximated by one of them. The structures are given in Figure 1, and represent (i) a community blood center which services the entire needs of a particular region, (ii) a collection of (communicating) community blood centers each under its own independent control, which services the blood needs of a particular region, and (iii) a regional blood center which either coordinates or controls the activities of a collection of community blood centers which in turn fulfill the needs of the community in the region. As a general rule, as the size of a region changes due to an increase in demand or





Centralization may involve:

- Power Transfer
- Centralized Information System
- Mergers, Relocations

Coordination may involve:

- Cooperative Donor Recruitment
- Resource Sharing
- Distribution and Processing
- Share Information
- Coordinate Planning

Figure 1: Regional Blood Banking

geography, the single community blood center may not adequately fulfill the needs of the region and one of the other two structures tends to replace the single center over time. In some cases, the new structure consists of a regional community blood center with satellite facilities.

In the next section, we focus on the location-allocation-transportation aspects of regionalization. Our variables will be central bank locations, regional boundaries and blood distribution network configurations. In the first model, all the other aspects of regionalization are summarized by certain terms called system costs, which are primarily functions of two factors: the number of hospitals in a region and the amount of blood used by each hospital in the region. Both of these factors are functionally related to the variables considered in the location-allocation-transportation model (since they vary with the regional boundaries.) The other factors that affect the non-transportation aspects of regionalization tend to be independent of the variables in the model, consequently they are independent of the location-allocation-transportation decisions.

The third section describes some middle level and institutional level models in their abstract forms. Although these mathematical models are directly applicable to the blood bank decision process, they do provide qualitative insights into the considerations underlying the optimal decisions and the general structure of such decisions.

The final section briefly summarizes the findings.

## 2. THE REGIONAL MODEL AND DECISIONS

The regionalization model is verbally described as follows: "Within a given geographical area, there are  $N$  hospitals. Regionalization is to be achieved by dividing the area into  $M$  regions and establishing a central blood bank in each region. All blood banking activities in a region are to be coordinated. Supply generation is to be done mainly by each central bank and each hospital is to obtain its primary blood supply from the central bank in its region. The blood distribution operation consists of periodic and emergency deliveries. The hospitals in a region receive their periodic daily requirements from their central bank. The blood deliveries are made by vehicles which, starting from the central bank, visit one by one the hospitals they are scheduled to supply, and return to the central bank. These vehicles have given capacities and given limits on the number of deliveries they can make per day. Because of the wide fluctuations in demand, a hospital may deplete certain blood types before the next periodic delivery is due. In that case, a delivery vehicle is dispatched immediately, from its central bank.

The delivery vehicle makes an emergency blood delivery to that hospital and returns to the blood bank. The problem is to decide how many central blood banks to set up, where to locate them, how to allocate the hospitals to the banks, and how to route the periodic supply operation, so that the total of transportation costs (periodic and emergency supply costs) and the other system costs are a minimum." This problem will be called the Blood Transportation-Allocation Problem (BTAP).

In modeling the BTAP it is necessary to describe the costs as functions of the decision variables. The periodic delivery costs, which are a set of linear terms, depend on all three types of variables in this problem, that is, routing the periodic supply operations, locating the banks, and allocating the hospitals to the banks. The emergency costs are also a set of linear terms and depend on only two types of variables, locating the banks and allocating the hospitals. The system costs are nonlinear functions of the size of the blood banks and the number of hospitals allocated to the blood banks, and therefore, of all the variables in this problem they depend only on hospital allocations. So, if the system costs were constant and the emergency costs were negligible, the model would be equivalent to the General Transportation Problem (GTP) (see Or [1976], or Magnanti et al. [1975]). If the system costs were constant and the periodic delivery costs were negligible, the model would reduce to a Location-Allocation problem (LAP) (see Cooper [1963, 1964], Hurter and Wendell [1973a, 1973b], or Francis and White [1974]). So the model is a complex combination of these two large problems. The basic strategy we use in order to obtain a good solution depends heavily on these two sub-problems. We solve each subproblem independently and then combine them at the end, making tradeoffs between them and superimposing the system costs considerations, to obtain a good solution to the model. However, it should be noted that, unlike the GTP, solving the LAP does not produce a complete, feasible solution to the main model. It only gives the locations and the allocations, and in order to get the missing periodic delivery routes, one must solve a set of vehicle dispatch problems.

Other work on regionalization or centralization of blood bank activities does not consider the location-allocation-transportation aspects. Jennings [1970, 1972] used a simulation model to construct part of a regional blood banking system. He grouped a number of identical hospitals together; however, he did not have a central blood bank. Transshipment policies and inventory levels were studied to see their impact on shortages and outdates. Yen [1965] also studied multiechelon inventory systems. He concentrated his efforts on the optimal inventory levels and the optimal issuing policies. Prastacos [1977] and Prastacos et al. [1977] were interested in the allocation of exist-

ing stocks among the hospitals. Neither Jennings, Yen nor Prastacos et al. studied the location of central banks or the allocation of hospitals to them.

#### THE BLOOD TRANSPORTATION-ALLOCATION MODEL

Because the BTAP is a complex optimization problem, we will make a few reasonable assumptions to decompose the problem into smaller subproblems.

ASSUMPTION 1: The number of banks,  $M$ , is a given number.

Even in cases in which the above assumption doesn't hold,  $M$  is almost always restricted to a small, finite, feasible set ( $M$  is always an integer and  $1 \leq M \leq N$ ). So, in those cases one could solve the problem for each feasible value of  $M$  to get the optimal solution. In this respect, Assumption 1 is not restrictive.

ASSUMPTION 2: The blood delivery period is daily for each hospital.

Considering that some hospitals use more than 7000 units of blood per year, while some others use less than 10, this is an unrealistic assumption. Unfortunately, determining the optimal multiple delivery periods as well as the location-allocation and routings increases the complexity of the problem considerably and makes it almost impossible to find a direct solution procedure. A simple, multiple period problem should have three options for the periodic deliveries (daily, biweekly, weekly). However, the problem of choosing optimal periods for each hospital would be a very large (and time consuming) combinatoric process. If the periods are set in advance (daily, biweekly, weekly) these multiple periods can be easily incorporated into the present location-allocation model merely by adjusting the costs to reflect costs per day. The routes of the delivery vehicles however would have to be adjusted later.

ASSUMPTION 3: The potential locations of the  $M$  banks are a finite (and usually small) number.

The set of practical feasible locations is almost always a small, finite set (usually one does not want to build a blood bank from scratch; instead, they are most frequently located in the area's largest hospitals or at existing blood centers). So, if necessary, we could solve the problem for all

combinations of feasible locations, to get the optimal solution. In a design problem, this is not an impossible enumeration, since even a region the size of the Chicago metropolitan area (using the amount of blood transfused as a measure of hospital size) has only six central blood banks and has only seven hospitals with consumption rates of over 7000 units per year.

Finally, it is assumed that each vehicle makes one (non-emergency) trip per day. If multiple trips per day are allowed in a real setting this model would need to be modified appropriately. Most of the latter do not qualify to be central blood banks for various reasons. So, in this respect Assumption 3 is not very restrictive.

The following notation will be used in formulating the BTAP.

- i)  $N$  is the number of demand points (i.e., hospitals).
- ii)  $M$  is the number of supply points (i.e., banks)
- iii)  $n$  is the maximum number of supply vehicles available.
- iv)  $\mathcal{D} = \{H_1, \dots, H_N\}$  is a set of  $N$  demand points.
- v)  $\mathcal{S} = \{H_{N+1}, \dots, H_{N+M}\}$  is a set of  $M$  supply points.
- vi)  $\mathcal{H} = \mathcal{D} \cup \mathcal{S}$  is the set of all points involved in the problem.
- vii)  $d_{ij}$  is the "distance" from  $H_i$  to  $H_j$ . It should be noted that although Euclidean distances among locations of hospitals and central banks are used in the solution procedure, one could obtain a matrix of accurate travel times between all pairs of hospitals and banks, and one could use this matrix or any other "distance measure" instead of the Euclidean distance matrix.
- viii)  $C_k$ ,  $k = 1, \dots, n$  is the capacity of supply vehicle  $k$ .
- ix)  $Q_i$ ,  $i = 1, \dots, N$  is the requirement of demand point  $i$ .
- x)  $D_k$ ,  $k = 1, \dots, n$  is the maximum distance supply vehicle  $k$  may travel on a non-emergency delivery route.
- xi)  $\gamma_i$ ,  $i = 1, \dots, N$  is the expected number of emergency deliveries to hospital  $H_i$  per period.  $\gamma_i$  is the probability that the demand at  $H_i$  exceeds the supply at  $H_i$  given the optimal inventory level at  $H_i$  is used.
- xii)  $s(\ell, q)$  is the systems cost function of a region, where  $\ell$  is the number of hospitals in that region, and  $q$  is the amount of blood used per year in that region.

- xiii)  $y_{ij}$ ,  $i = 1, \dots, N$ ;  $j = N+1, \dots, N+M$  is a zero-one variable such that  $y_{ij}$  is 1 if hospital  $H_i$  is assigned to central bank  $H_j$  and is 0 otherwise.
- xiv)  $x_{ijk}$ ,  $i = 1, \dots, N+M$ ;  $j = 1, \dots, N+M$ ;  $k=1, \dots, n$  is a zero-one variable such that  $x_{ijk}$  is 1 if vehicle  $k$  goes from hospital  $H_i$  to  $H_j$  and is 0 otherwise.

The BTAP is:

PROBLEM 1:

$$\min z^1(x,y) = \sum_{i=1}^{N+M} \sum_{j=1}^{N+M} \sum_{k=1}^n d_{ij} x_{ijk} + \sum_{i=1}^N \sum_{j=N+1}^{N+M} \gamma_i d_{ij} y_{ij} + \sum_{j=N+1}^{N+M} s \left( \sum_{i=1}^N y_{ij}, \sum_{i=1}^N Q_i y_{ij} \right) \quad (1)$$

subject to

$$\sum_{k=1}^n \sum_{j=1}^{N+M} x_{ijk} = 1 \quad i = 1, \dots, N \quad (2)$$

$$\sum_{j=1}^{N+M} \sum_{i=1}^N Q_i x_{ijk} \leq C_k \quad k = 1, \dots, n \quad (3)$$

$$\sum_{j=1}^{N+M} \sum_{i=1}^{N+M} d_{ij} x_{ijk} \leq D_k \quad k = 1, \dots, n \quad (4)$$

$$\sum_{\{i:H_i \in S\}} \sum_{\{j:H_j \in \bar{S}\}} \sum_{k=1}^n x_{ijk} \geq 1 \quad \text{for all } (S, \bar{S}) \quad (5)$$

where  $S$  is any proper subset of  $\mathcal{H}$  containing  $\mathcal{L}$  and  $\bar{S}$  is the complement of  $S$ .

$$\sum_{j=1}^{N+M} x_{hjk} = \sum_{i=1}^{N+M} x_{ihk} \quad k = 1, \dots, n; h = 1, \dots, N+M \quad (6)$$

$$y_{ij} \geq \sum_{h=1}^{N+M} x_{ihk} + \sum_{h=1}^{N+M} x_{jkh} - 1 \quad i = 1, \dots, N; j = N+1, \dots, N+M \quad (7)$$

$$k = 1, \dots, n$$

$$x_{ijk} = 0, 1 \quad i = 1, \dots, N+M; j = 1, \dots, N+M; \quad (8)$$

$$k = 1, \dots, n$$

(note that  $x_{iik} = 0$ )

$$y_{ij} = 0, 1 \quad i = 1, \dots, N; j = N+1, \dots, N+M. \quad (9)$$

The explanation of these constraint sets are as follows. Constraints (2) require that every hospital receive a shipment from some vehicle; (3) are the vehicle capacity constraints; (4) are the maximum travel distance constraints (note, it is implicitly assumed that  $Q_i \leq C_k$  for  $i=1, \dots, N$  and  $k=1, \dots, n$ ); (5) require that graph  $\mathcal{L}$  corresponding to  $x$  is connected; (6) imply that a vehicle departs from a point  $h$  if and only if it enters there (conservation of flow); (7) contains the coupling constraints between variables  $x = \{x_{ijk}\}$  and  $y = \{y_{ij}\}$ . It means that if there is vehicle  $k$  passing

from both hospital  $i$ , ( $\sum_{h=1}^{N+M} x_{ihk} = 1$ ), and from bank  $j$ , ( $\sum_{h=1}^{N+M} x_{jkh} = 1$ ), then hospital  $i$  is assigned to bank  $j$ , ( $y_{ij} \geq 1+1-1 = 1$ ). As shown in Or [1976], these constraints imply that there is an optimal solution in which each vehicle is based at a particular supply point.

In Problem 1 the variables  $x = \{x_{ijk}\}$  correspond to the routing of the periodic delivery vehicles and the variables  $y = \{y_{ij}\}$  correspond to the allocations of the hospitals to the blood banks. For a given  $x = \{x_{ijk}\}$ ,  $y = \{y_{ij}\}$  is uniquely determined, but the converse is not true; if we are given the allocations, a series of  $M$  vehicle dispatch problems have to be solved, in order to obtain the routings. Problem 1 has a finite feasible solution set and a nonempty optimal solution set. However, the underlying Multiple Vehicle Dispatch Problem (MVDP) makes it a complex integer programming problem. For  $N$  of any significant size ( $N \geq 20$ ), the BTAP is too large to be solved by conventional mathematical programming techniques in a reasonable amount of time.

In Problem 1, if  $\gamma_i$ ,  $i = 1, \dots, N$  are small or emergency costs negligible (actual  $\gamma_i$ 's range from .0002 to .06 when optimal ordering policies are followed, see Pierskalla and Yen [35]) and the function  $s(\ell, k)$  is essentially constant, then

$$z^2(x) = \sum_{i=1}^{N+M} \sum_{j=1}^{N+M} \sum_{k=1}^n d_{ij} x_{ijk}$$

would be the dominating term in the objective function (1). Then we could just solve the MVDP,

PROBLEM 2

$$\min \sum_{i=1}^{N+M} \sum_{j=1}^{N+M} \sum_{k=1}^n d_{ij} x_{ijk} \quad (11)$$

subject to

$$\sum_{k=1}^n \sum_{j=1}^{N+M} x_{ijk} = 1 \quad i=1, \dots, N \quad (12)$$

$$\sum_{j=1}^{N+M} \sum_{i=1}^N Q_i x_{ijk} \leq C_k \quad k=1, \dots, n \quad (13)$$

$$\sum_{j=1}^{N+M} \sum_{i=1}^{N+M} d_{ij} x_{ijk} \leq D_k \quad k=1, \dots, n \quad (14)$$

$$\sum_{i: H_i \in S} \sum_{j: H_j \in \bar{S}} \sum_{k=1}^n x_{ijk} \geq 1 \quad \text{for all } (S, \bar{S}) \quad (15)$$

$$\sum_{j=1}^{N+M} x_{hjk} = \sum_{i=1}^{N+M} x_{ihk} \quad k=1, \dots, n; \quad h=1, \dots, N+M \quad (16)$$

$$x_{ijk} = 0, 1 \quad i=1, \dots, N+M; \quad j=1, \dots, N+M \quad (17)$$

$$k=1, \dots, n$$

in order to obtain the optimal  $x^*$  for Problem 1. The optimal allocations,  $y^*$ , would then be uniquely determined by  $x^*$ .

On the other hand, if  $\gamma_i$ ,  $i=1, \dots, N$  are relatively large (which might happen under nonoptimal ordering policies) or system costs and periodic



delivery costs are negligible, then

$$z^3(y) = \sum_{i=1}^N \sum_{j=N+1}^{N+M} \gamma_i^d y_{ij}$$

would be the dominating term in the objective function. Then we could just solve the allocation problem,

PROBLEM 3

$$\min \sum_{i=1}^N \sum_{j=N+1}^{N+M} \gamma_i^d y_{ij} \quad (18)$$

subject to

$$\sum_{j=N+1}^{N+M} y_{ij} = 1 \quad i = 1, \dots, N \quad (19)$$

$$y_{ij} = 0, 1 \quad \begin{array}{l} i = 1, \dots, N; \\ j = N+1, \dots, N+M \end{array} \quad (20)$$

in order to get the optimal  $y^\circ$  for Problem 1. Then, optimal routings,  $x^\circ$ , would be obtained by solving a vehicle dispatch problem for each one of the  $M$  regions determined by  $y^\circ$ .

Let  $x^*$  be an optimal solution of Problem 2. Let  $y^*$  be the allocations determined by  $x^*$ . Let  $y^\circ$  be an optimal solution of Problem 3.

It directly follows from the above definitions that

$$z^2(x^*) \leq z^2(x^\circ)$$

$$z^3(y^\circ) \leq z^3(y^*)$$

and if the systems costs are essentially constant, then

$$z^2(x^*) + z^3(y^\circ)$$

would be a good lower bound on the optimal value of Problem 1.

In order to minimize the cost of operation under different regional system configurations, the economies of scale curves representing feasible

combinations of the functional areas of blood banking were incorporated into the model. These curves form the basis for determining the operating costs for each of the different regional structures.

For example, if one considers the regional structure given by the first figure in Figure 1, then the only questions which arise relative to costs involve where should the community blood center be located in order to minimize total transportation costs for donors, recruiting, phlebotomy<sup>3</sup> on mobiles, and routine and emergency deliveries to the transfusion services. Since the other costs, such as processing, administration, inventory control and phlebotomy at the Center are relatively independent of location, these costs would not be included in the decision process, because they would be incurred no matter where the community blood center was located.

On the other hand, if the regional structure is that of the second figure in Figure 1, then all of the costs from the economies of scale curves are relevant, since not only the location of the community blood centers but also their sizes in the region are important decision variables. Consequently, all of the curves are needed, and tradeoffs among these costs and locations in sizes must be made.

Finally, if one were to analyze the third figure of Figure 1, the appropriate use of the economies of scale curves would depend upon the authority structure and governance relationships between the regional blood center and the community blood centers. For example, if the regional blood center were only a coordinating body of information and did not really have any authority over the community blood centers, then the costs of operation would be very similar to those in the preceding paragraph. That is, all of the functional areas of blood banking would be located at the community blood centers; consequently, virtually all of the costs would be incurred there, and hence if one were to do a regional design, one would be interested in the location of the blood centers, as well as their operating size. On the other hand, if for example, donor recruiting were done at the regional blood center, then the costs of donor recruiting would not be charged to the community blood centers but would be a regional blood center cost. Since donor recruiting costs are related to the distances the donor recruiters have to travel, then the location of the regional blood center would be a factor in the cost structure; however, none of the other costs, such as phlebotomy, processing, inventory control and distribution administration at the community blood

---

<sup>3</sup>Phlebotomy is the procedure of drawing blood from a donor. These drawings often occur at the hospital, blood center, or at a distant location such as a school, church, company, etc. When drawings are made at a distant location, mobile blood vehicles are used and the drawings are thus called "phlebotomy on mobiles."

centers would be included in the cost of the regional blood center itself. Consequently, wherever the functional areas and authority for administration of those areas are located in the system, then those costs should be charged at those appropriate places.

### 3. DECISIONS AT LOWER ECHELONS

The preceding analysis focused on macro and some middle level decision making models in a regional system. In this section, we discuss decisions on inventory control, issuing, and crossmatch release policies at lower levels in the modeling hierarchy. These decisions impact the decisions at the higher levels since they affect the system cost structures mentioned in the previous section.

The basic community blood center scheme is that of a "wheel" structure as shown in Figure 2, which is another view of Figure 1a.

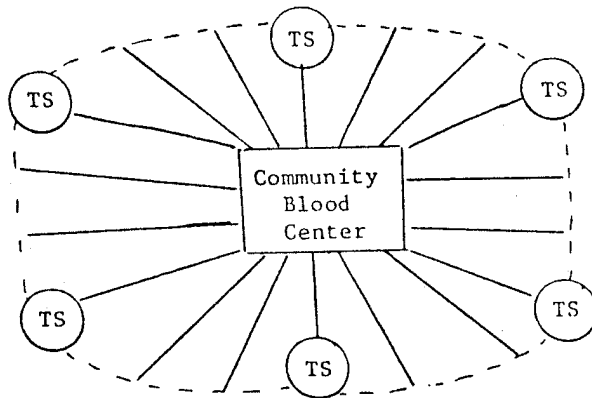


Figure 2

The Community Blood Center and Its  
Satellite Transfusion Service (TS's)

The community blood center acts as the hub of activities for its satellite blood banks and transfusion services. In the centralized system, the central blood bank supplies a number of TS's and maintains the authority to redistribute all blood in the system. Most TS's maintain a supply of whole blood and/or packed red cells. Other components may also be maintained at some facilities.

Another way to view a CBB and the decisions needed to answer the questions posed earlier is shown in Figure 3. In this figure, the CBB is shown at the top and the lines represent the flow of units through the different activities and TS's in the system. That is, Figure 3 is a schematic drawing of the inputs, outputs, and flows in a centralized system. Some of the decisions needed by the CBB are shown in the diamond-shaped boxes. The variables  $S_0, S_1, \dots, S_N$  represent the number of units needed on hand each day at the CBB and at the TS's in order to meet the system needs without excess outdates. Of course, the  $S_i$ 's are different for each ABO-Rh type and each component<sup>4</sup> and they change over time depending upon the changing needs at the TS's.

Basically the inventory flow system for the CBB operates in the following manner. Forecasts of future ABO-Rh blood needs and component needs are made. The CBB periodically constructs mobile phlebotomy schedules and forecasts the corresponding quantities to be drawn at each mobile site. Individual drawings are also often made at the CBB itself and/or its TS's. These drawings are scheduled to meet the forecasted demands at the TS's and maintain a stock of inventories on hand at the CBB. On a daily, semi-weekly or weekly basis depending on the level of activity and proximity to the CBB for each TS, orders to the TS's must be filled.

After the CBB receives all the requests from the TS's, the orders are filled by drawing from the inventories in the CBB. The decision policy as to which units to send is called the issuing policy. The most common issuing policies are first-in-first out (FIFO) or last-in-first-out (LIFO). For purposes of simplification as well as good medical practice, each ABO type and Rh factor is considered independent of the other types and Rh factors. When the sum of all TS demands for whole blood/packed red cells (WB/PRC's) or for particular components exceeds the total inventory in the CBB, the CBB may backlog the excess demand or may fill all demands by calling in donors, by contacting other CBB's, by using frozen packed red cells if appropriate, or

---

<sup>4</sup>There are many derivative products which comprise whole blood and which can be separated from the whole blood. These derivatives are called components. Some commonly derived components are: platelets, granulocytes, leukocytes, cryoprecipitate, plasma and red cells.

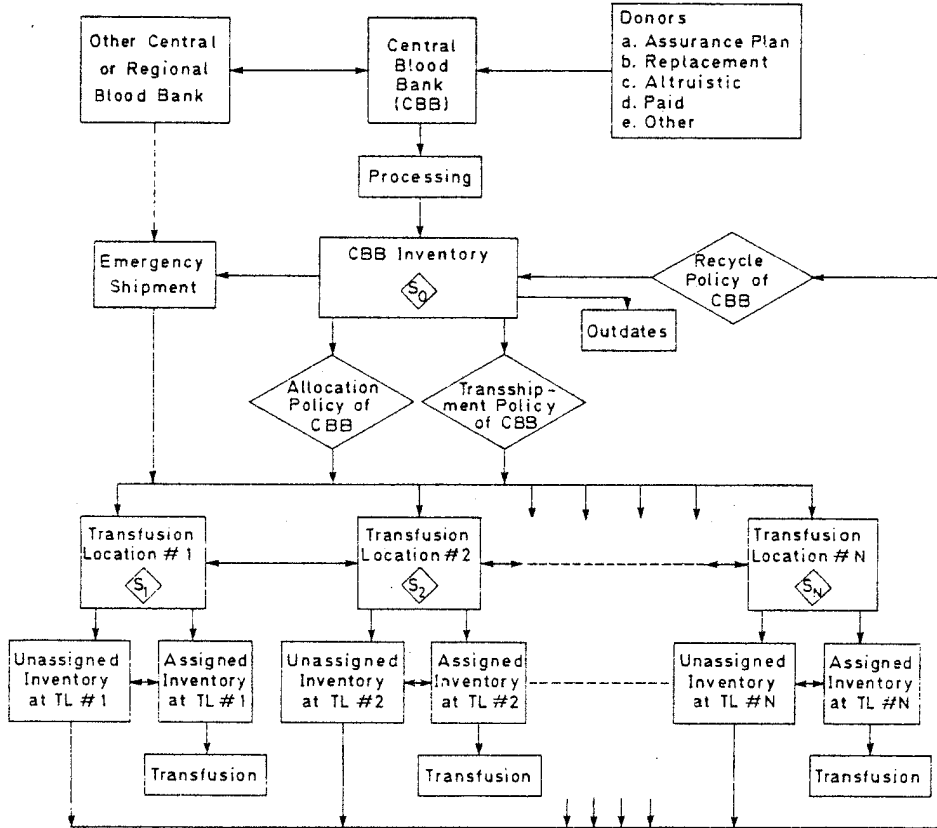


FIGURE 3.

FLOW CHART FOR A CENTRAL BLOOD BANKING SYSTEM  
SHOWING INVENTORY LOCATIONS AND CBB DECISION POLICIES

by requesting an emergency shipment from still higher echelon blood banks. The CBB uses different approaches to handle the excess demand depending upon whether the orders are routine or emergency.

Two examples of mathematical models to analyze the inventory levels for fresh red cells, frozen red cells and platelets are given here. Models for issuing and allocation of blood units are given in Yen [1975], Pierskalla and Roach [1972] and Prastacos, et al., [1977].

#### 4. A TWO PRODUCT PERISHABLE/NON PERISHABLE INVENTORY PROBLEM

As units of blood are drawn from donors they may be refrigerated, in which case the shelf life is twenty-one days, or frozen, in which case the shelf life is 365 days, which for all practical purposes is non-perishable. When demands exceed the available supply of fresh blood, frozen blood may then be thawed.

##### The One Period Model

We will make the following assumptions:

- 1) All orders are placed at the start of a period and received instantly.
- 2) All stock arrives new.
- 3) Demands in successive periods are independent identically distributed random variables with distribution  $F$  and density  $f$ . In addition we will assume that  $f(t) > 0$  when  $t > 0$ .
- 4) Inventory of product 1 (the perishable product) is depleted according to a FIFO policy.
- 5) All costs are linear. They include:
  - a) Ordering in both products (at unit costs  $c_1$  and  $c_2$ ) charged at the start of the period;
  - b) Holding in both products (at unit costs  $h_1$  and  $h_2$ ) charged on what is on hand at the end of the period;
  - c) Shortage (at unit cost  $r$ ) charged on the unsatisfied demand at the end of the period (the shortage cost is either the cost of emergency shipments from some source or the marginal cost of an additional day in the hospital for postponed surgery);
  - d) Outdating (at unit cost  $\theta$ ) charged on what deteriorates at the end of the period.
- 6) If product 1 has not been depleted by demand before reaching

age  $m$ -periods, it must be discarded at the unit cost given in 5(d).

- 7) There is a single demand source. Demands first deplete from product 1 (the perishable product) and then product 2. Excess demand is backlogged in the second product.

A number of the assumptions may be relaxed or altered. It should be noted that assumption (4) is quite mild. The optimality of FIFO for depletion of perishable inventory has been established under far more general circumstances (Nahmias [1974] and Pierskalla and Roach [1972]).

The approach will be to charge the outdating cost against the expected outdating of the present order which will not occur for  $m$  periods. The motivation behind this method is discussed in Nahmias [1975]. If  $x = (x_{m-1}, \dots, x_1)$  is the vector of perishable inventory on hand,  $x_i$  = number of units on hand which will outdate in exactly  $i$  periods, and  $y$  is the amount of new perishable inventory ordered, then it has been shown that the expression  $\int_0^y G_m(u; \underline{x}) du$  represents the expected outdating of  $y$ ,  $m$  periods into the future, where

$$G_n(t; \underline{x}(n-1)) = \int_0^t G_{n-1}(x_{n-1} + v; \underline{x}(n-2)) f(t-v) dv$$

for  $1 \leq n \leq m$ ;  $\underline{x}(n) = (x_n, \dots, x_1)$  and  $G_0(t) = \begin{cases} 1 & \text{if } t \geq 0 \\ 0 & \text{if } t < 0 \end{cases}$ . For each

$n \geq 1$ ,  $G_n(t; \underline{x}(n-1))$  is a C.D.F. in its first argument, and may possess a discontinuity at  $t = 0$ .

Letting  $x^2$  = amount of product 2 on hand,

$z$  = amount of product 2 on hand after ordering,

the total expected cost of ordering  $y$  of product 1 and  $z - x^2$  of product 2 is

$$\begin{aligned} c_1 y + h_1 \int_0^{x+y} (x+y-t) f(t) dt + \theta \int_0^y G_m(u; \underline{x}) du + c_2 (z-x^2) + h_2 z F(x+y) \\ + h_2 \int_{x+y}^{x+y+z} (x+y+z-t) f(t) dt + r \int_{x+y+z}^{\infty} (t-x-y-z) f(t) dt \end{aligned}$$

where  $x = \sum_{i=1}^{m-1} x_i$  for convenience. Collecting all terms independent of  $x^2$  we

will write this as  $L(\underline{x}, y, z) - c_2 x^2$ . A point which should be noted here is that the decision variables for each product have different interpretations:  $y$  represents the actual quantity of product 1 ordered, while  $z$  is the inventory level of product 2 after ordering. Our interest in the single period model is secondary to that of the multi-decision dynamic problem. The optimal ordering policy over the finite horizon will satisfy the functional equations

$$C_n(\underline{x}, x^2) = \inf_{\substack{y \geq 0 \\ z \geq x^2}} \{L(\underline{x}, y, z) - c_2 x^2 + \alpha \int_0^{\infty} C_{n-1}(s_1(\underline{x}, y, t), s_2(x+y, z, t))f(t)dt\}$$

for  $1 \leq n \leq N$  ( $N$  is a fixed positive integer). The transfer functions are given by:

$$s_1(\underline{x}, y, t) = (s_{1,m-1}(\underline{x}, y, t), \dots, s_{1,1}(\underline{x}, y, t))$$

where

$$s_{1,i}(\underline{x}, y, t) = (x_{i+1} - (t - \sum_{j=1}^i x_j)^+)^+, \quad 1 \leq i \leq m-1$$

(we interpret  $x_m = y$ )

and  $s_2(x+y, z, t) = z - (t - (x+y))^+$  where  $g^+ = \max(g, 0)$ . The discount factor is  $\alpha \in (0, 1)$ .

Again collecting all terms independent of  $x^2$ , we will write

$$C_n(\underline{x}, x^2) = \inf_{\substack{y \geq 0 \\ z \geq x^2}} \{B_n(\underline{x}, y, z) - c_2 x^2\}$$

in which

$$B_n(\underline{x}, y, z) = L(\underline{x}, y, z) + \alpha \int_0^{\infty} C_{n-1}(s_1(\underline{x}, y, t), s_2(x+y, z, t))f(t)dt.$$

The functions  $C_n(\underline{x}, x^2)$  have the usual interpretation as the minimum expected discounted cost for an  $n$  period problem when  $(\underline{x}, x^2)$  is on hand. As is customary with dynamic programming models, the periods are numbered backwards. The goal of our analysis will be to answer the two questions: when should an order be placed and how much of each product should be ordered?



The Ordering Regions

It becomes notationally and mathematically convenient to introduce the assumption that inventory remaining at the end of the horizon can be salvaged; inventory of product 1 remaining may be salvaged at a return  $c_1x$  and of product 2 at a return  $c_2x^2$ . Backlogged demand in product 2 may be made up by an emergency order at a cost  $-c_2x^2$ . That is,

$$C_0(x, x^2) = -c_1x - c_2x^2.$$

The assumption is identical to that made by Veinott [1965] in the analysis of non-perishable multi-product problems. The convenience of this assumption is that the somewhat artificial fact that the problem terminates in a given time horizon with inventory remaining is removed and the problem behaves more like a problem under steady state conditions. In some inventory problems this assumption allows the decomposition of the n-period problem into n one-period problems which are easier to solve.

In addition we will make the following four assumptions regarding the cost parameters:

- i)  $0 \leq h_2 \leq h_1$
- ii)  $0 < c_1 < c_2$
- iii)  $r > (1-\alpha)c_2$
- iv)  $0 \leq (1-\alpha)(c_2 - c_1) + (h_2 - h_1) < \theta$

Since perishable inventory must often be stored under special conditions, assumption (i) is not unreasonable. Assumption (ii) is necessary to insure that it is economical to stock product 1. Assumption (iii) is a usual one made for non-perishable inventory (Arrow, et al., [1958]). The expression of assumption (iv) is precisely the difference in the costs of procuring and holding one unit of each type of inventory and salvaging it the following period. If this term were negative then it would never be optimal to order in product 1; or if it exceeded the unit cost of outdating, it would never be optimal to order to a positive level in product 2.

We will have need to refer to the following constants in the analysis of the ordering regions in the first period:

$$u^* = F^{-1} \left[ \frac{r - c_2(1-\alpha)}{r+h_2} \right]$$

$$w^* = F^{-1} \left[ \frac{r - c_1(1-\alpha) + (h_2 - h_1)}{r+h_2} \right].$$

When  $F$  is strictly increasing, assumptions (iii) and (iv) guarantee that  $u^*$  and  $w^*$  both exist and are strictly positive. If the additional condition  $\alpha c_2 - h_2 < c_1$  is satisfied (although we will not require it to be), then we may also define the constant

$$v^* = F^{-1} \left[ \frac{r - c_1 + \alpha c_2}{r+h_2} \right]$$

With the inclusion of the salvage value assumption it follows from the definition of the transfer functions that

$$\begin{aligned} B_1(\underline{x}, y, z) &= L(\underline{x}, y, z) + \alpha F(x_1) \left[ -c_1 \left( y + \sum_{i=2}^{m-1} x_i \right) - c_2 z \right] \\ &\quad - \alpha c_1 \int_{x_1}^{x+y} (x+y-t) f(t) dt - \alpha c_2 [F_2(x+y) - F(x_1)] z \\ &\quad - \alpha c_2 \int_{x+y}^{\infty} (x+y+z-t) f(t) dt \end{aligned}$$

We will adopt the following notational convention: if  $h: \mathbb{R}^n \rightarrow \mathbb{R}^1$  and  $h \in C^{(2)}$ , then  $h^{(i)}$  is the first partial derivative of  $h$  with respect to its  $i$ th argument and  $h^{(i,j)}$  is the second cross partial derivative with respect to the  $i^{\text{th}}$  and  $j^{\text{th}}$  arguments respectively.

We have the following

**THEOREM II.4.1:**  $B_1(\underline{x}, y, z)$  is convex in  $(y, z)$  for all non-negative  $\underline{x}$ . The functions  $y_1(\underline{x}), z_1(\underline{x})$  solving  $B_1(\underline{x}, y_1(\underline{x}), z_1(\underline{x})) = \min_{y,z} \{B_1(\underline{x}, y, z)\}$  satisfy  $B_1^{(m)}(\underline{x}, y_1(\underline{x}), z_1(\underline{x})) = B_1^{(m+1)}(\underline{x}, y_1(\underline{x}), z_1(\underline{x})) = 0$  and are unique for all  $\underline{x}$ .

Since  $y_1(\underline{x}) > 0$  for all  $\underline{x}$ , a necessary and sufficient condition that it be possible to order to the global minimum of  $B_1(\underline{x}, y, z)$  is  $x^2 < z_1(\underline{x})$ . If  $x^2 \geq z_1(\underline{x})$ , it may still be optimal to order product 1. In this case we define the function  $p_1(\underline{x}, x^2)$  to satisfy

$$B_1(\underline{x}, p_1(\underline{x}, x^2), x^2) \equiv \inf_y (B_1(\underline{x}, y, x^2)).$$

If  $p_1(\underline{x}, x^2) > 0$  then it is optimal to order this amount of product 1.

Hence the following characterization follows:

THEOREM II.4.2: A necessary and sufficient condition that it is optimal to order a positive amount of product 1 is  $B_1^{(m)}(\underline{x}, 0, x^2) < 0$ . If one reasons analogously to THEOREM II.4.2, then it is tempting to assume that  $B_1^{(m+1)}(\underline{x}, 0, x^2) < 0$  implies it is optimal to order in product 2. However this is not the case. To see why, let  $t(\underline{x}, y)$  satisfy  $B_1^{(m+1)}(\underline{x}, y, t(\underline{x}, y)) = 0$ . Differentiating implicitly with respect to  $y$  we obtain

$$t^{(m)}(\underline{x}, y) = \frac{-B_1^{(m+1, m)}(\underline{x}, y, t(\underline{x}, y))}{B_1^{(m+1, m+1)}(\underline{x}, y, t(\underline{x}, y))} < 0.$$

Since  $t(\underline{x}, y_1(\underline{x})) = z_1(\underline{x})$  and  $y_1(\underline{x}) > 0$  it follows that  $t(\underline{x}, 0) > z_1(\underline{x})$ . If  $x^2$  satisfies  $z_1(\underline{x}) \leq x^2 < t(\underline{x}, 0)$  then  $B_1^{(m+1)}(\underline{x}, 0, x^2) < 0$  and it is optimal not to order in product 2.

Hence there are exactly three distinct ordering regions:

Region I - Optimal to order in both products:  $x^2 < z_1(\underline{x})$ ,

Region II - Optimal to order in product 1 only:  $x^2 \geq z_1(\underline{x})$   
 $B_1^{(m)}(\underline{x}, 0, x^2) < 0, (p_1(\underline{x}, x^2) \geq 0)$ .

Region III - Optimal not to order:  $B_1^{(m)}(\underline{x}, 0, x^2) \geq 0, (p_1(\underline{x}, x^2) \leq 0)$ .

Note from the proof of THEOREM II.4.2 that if it is optimal to order in product 2 then it is optimal to order in product 1. The boundary between Regions I and II is quite complex, as it depends on the entire vector  $(\underline{x}, x^2)$ . However the boundary between Regions II and III depends on the vector  $\underline{x}$  only through the sum of its components,  $x$ .

Define  $g(x) = (1/r+h_2) \cdot \{r+\alpha c_2 - c_1 - F(x)[(h_1 - h_2) + \alpha(c_2 - c_1)]\}$ .

Then

THEOREM II.4.3: A necessary and sufficient condition that  $(\underline{x}, x^2)$  is in Regions I or II is that

$$F(x + x^2) < g(x).$$

The function  $g(x)$  is strictly decreasing in  $x$  with  $g(+\infty) = F(w^*)$ . If the condition  $\alpha c_2 - h_2 < c_1$  is satisfied then  $F(w^*) \leq g(x) < 1$  for all  $x \geq 0$  and  $v^* = F^{-1}(g(0))$  will exist. However, if  $\alpha c_2 - h_2 \geq c_1$  then  $g(0) \geq 1$ . In this case there exists a unique number  $p^* \geq 0$  which solves  $g(p^*) = 1$ . The boundary between Regions II and III may be pictured in the  $(x, x^2)$  plane

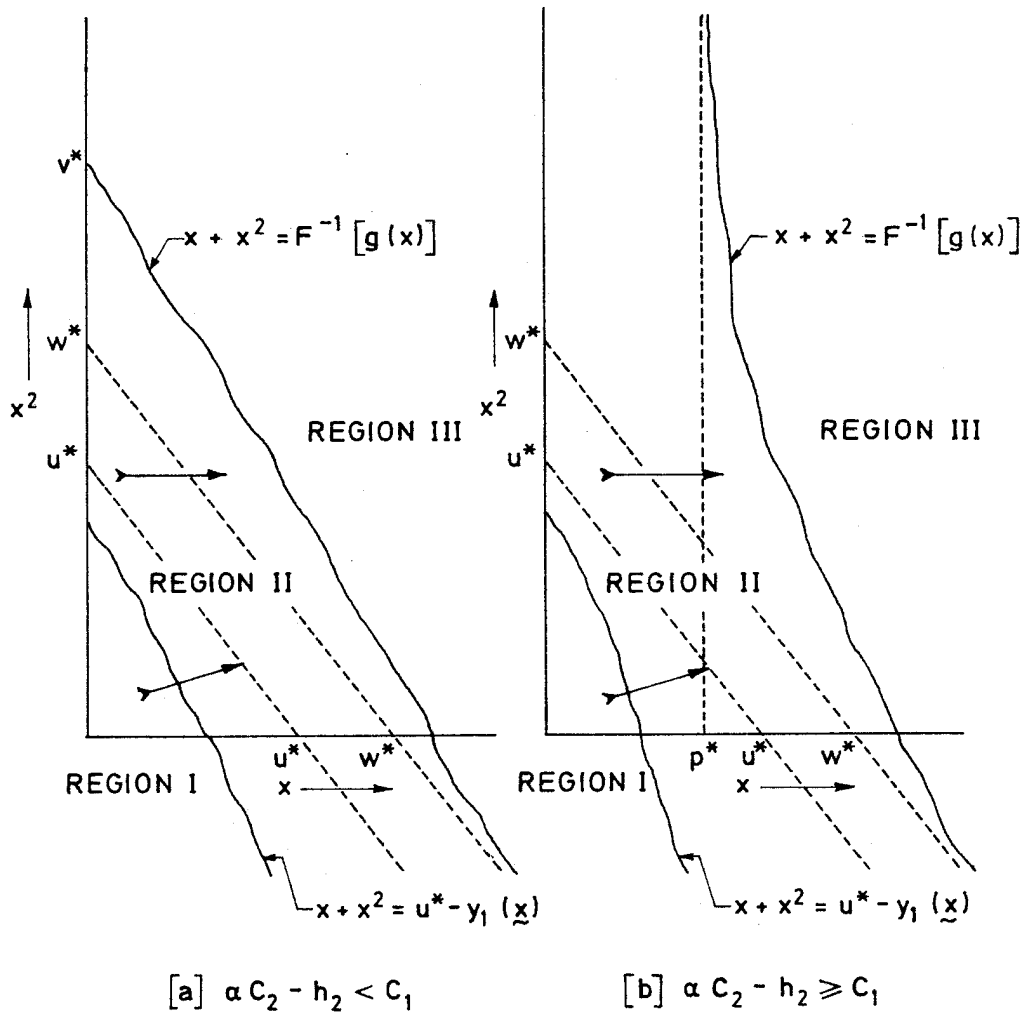


FIGURE 4. THE ORDERING REGIONS

independent of  $m$ . Figure 4 pictures this boundary for each of the two cases above. The boundary between Regions I and II may be pictured in the  $(x, x^2)$  plane only if  $m = 2$ . The arrows indicate the inventory position after ordering. Notice that  $(\tilde{x}, x^2) \in$  Region I guarantees that  $x + y_1(\tilde{x}) + z_1(\tilde{x}) = u^*$  (which follows from  $B_1^{(m+1)}(\tilde{x}, y_1(\tilde{x}), z_1(\tilde{x})) = 0$ ) while  $(x, x^2) \in$  Region II will yield  $F(x + x^2 + p_1(\tilde{x}, x^2)) < g(x + p_1(\tilde{x}, x^2))$ .

These results can be extended to the multiperiod dynamic problem. Even with the inclusion of the salvage value assumption, neither the ordering regions nor the ordering policies are stationary in time.

5. A BY-PRODUCT PRODUCTION SYSTEM WITH AN ALTERNATIVE

The idea behind this example is that components are made from whole blood, e.g., red cells and platelets. In such cases, two simultaneous decisions must be made. First, the inventory level of each product must be determined to best meet (random) demands. Second, the best (optimal) production level (e.g., number of runs) must be determined to achieve these inventories. Often these decisions cannot be made independently because production capacities or design characteristics of the production lines may prohibit arbitrary combinations. The system we consider is describe in Figure 5. There are two inventory items, called products 1 and 2, and two different production processes called types A and B. Type A is capable of making both products, simultaneously, according to the production coefficients  $\eta_1$  and  $\eta_2$ . When type A is operated at the unit level (e.g., a single run),  $\eta_1$  and  $\eta_2$  units of products 1 and 2 are obtained. We choose to call type A a by-product production process due to its multi-product capability. Type B is a single item production process and is capable of making only product 2. In this context, type B is the alternative method of obtaining product 2.

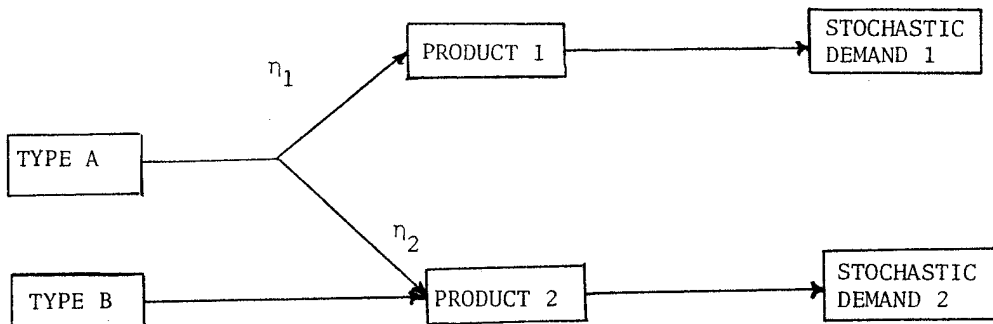


FIGURE 5  
Block Diagram of the Production System

The model is of the periodic review type, where the planning horizon is  $N$  periods long. At the beginning of period  $n$ ,  $n=1,2,\dots,N$ , the initial inventories (before production)  $x_n = (x_{n,1}, x_{n,2})$  are reviewed, where  $x_{n,j}$  is the initial inventory of product  $j$ ,  $j=1,2$ . Then, the starting inventories (stock levels after production but before demand)  $y_n = (y_{n,1}, y_{n,2})$  and the production levels  $p_n = (p_{n,A}, p_{n,B})$  are jointly determined.  $y_{n,j}$  is the starting inventory of product  $j$ ,  $j=1,2$  in period  $n$  and  $p_{n,A}$ ,  $p_{n,B}$  the production levels of types A and B, respectively. Then, a random demand  $D_n = (D_{n,1}, D_{n,2})$  is realized. We use a reverse recursion numbering scheme so that index  $N$  refers to the first period and 1 refers to the last period. When focusing attention to quantities within a period, we frequently drop the period index on the above quantities.

The following is a list of specific assumptions governing the development of our model.

#### 1) Production Processes

It is assumed that each production process is controlled by specifying the production level. For type A this level is  $p_A$ ; for type B it is  $p_B$ . Thus, the amount of each product produced is  $(\eta_1 p_A, p_B + \eta_2 p_A)$ . We assume that  $\eta_1$  and  $\eta_2$  are constant; they are not decision variables. For the present, we assume that there are no lead times and that there are no upper bounds on production.

#### 2) Demands

We assume that  $D_n$ ,  $n=1,2,\dots,N$  form a sequence of nonnegative independent and identically distributed random vectors with a continuous joint density function  $f(\cdot)$  and finite expected value. Although we allow dependencies among demands within a period, we do require that  $D_{n,j}$  be satisfied only from inventories of product  $j$ ,  $j=1,2$ . It is further assumed that unsatisfied demands are backlogged and are not lost.

#### 3) Production Costs

- a)  $c_A$  is the cost per unit of production on A;
- b)  $c_B$  is the cost per unit of production on B;
- c) there are no fixed costs of production.

#### 4) Inventory Costs

Let  $g_i(\cdot)$  be the holding and shortage cost for product  $i$ ,  $i=1,2$ , over any single period. We define the expected total holding and shortage over any period as given by

$$L(y) = \int_0^{\infty} \int_0^{\infty} [g_1(y_1 - t) + g_2(y_2 - u)] f(t, u) dt du.$$

We assume that

- a)  $L(\cdot)$  is strictly convex over  $R_2$ ;
- b)  $L(\cdot)$  has continuous second partial derivatives;
- c)  $\frac{1}{\eta_1} (c_A - \eta_2 c_B) y_1 + c_B y_2 + L(y) \rightarrow +\infty$

whenever  $\|y\| \rightarrow +\infty$  (where  $\|\cdot\|$  is the Euclidean norm). (3) assures the existence of an optimal policy.

### 5) Discount Parameter

Let  $\alpha \in [0, 1]$ .  $\alpha$  denotes the parameter which relates costs in future periods to the present.

We use the following notation for derivatives and partial derivatives of functions. For a twice continuously differentiable scalar function  $h(x)$ , let  $h'(\cdot)$  and  $h''(\cdot)$  represent its first and second derivatives, respectively. For a function  $g(\cdot)$  defined on  $R_n$  with continuous second partial derivatives, let

$$g^{(i)}(x) = \frac{\partial}{\partial x_i} g(x), \quad i=1,2,\dots,n$$

$$g^{(i,j)}(x) = \frac{\partial^2}{\partial x_j \partial x_i} g(x), \quad i,j=1,2,\dots,n.$$

The problem is to determine a policy that leads to the minimum expected discounted costs over the  $N$  period horizon. This leads to an optimization problem that is most easily conceptualized as a dynamic programming problem. Before formulating the dynamic program, we first show the relationship between a feasible set of production levels  $p = (p_A, p_B)$  and starting inventory levels  $y = (y_1, y_2)$  given any initial inventory  $x$ . Given  $p$  and  $x$ , the starting inventories  $y$  must satisfy

$$y_1 = \eta_1 p_A + x_1$$

$$y_2 = \eta_2 p_A + p_B + x_2.$$

Given  $y$  and  $x$ , the production levels  $p$  are determined by

$$p_A = \frac{1}{\eta_1} (y_1 - x_1)$$

$$p_B = (y_2 - \frac{\eta_2}{\eta_1} y_1) - (x_2 - \frac{\eta_2}{\eta_1} x_1) = y_2 - (x_2 + \frac{\eta_2}{\eta_1} (y_1 - x_1))$$

Although the dynamic program is most easily defined in terms of  $p$  and  $x$ , we favor using  $y$  and  $x$  since they characterize the optimal policy.

The dynamic programming recursion is given by

$$c_n(x) = \inf \{G_n(y) - \frac{1}{\eta_1} (c_A - \eta_2 c_B)x_1 - c_B x_2\}$$

subject to

$$y_1 \geq x_1$$

$$y_2 \geq x_2 + \frac{\eta_2}{\eta_1} (y_1 - x_1)$$

for

$$G_n(y) = \frac{1}{\eta_1} (c_A - \eta_2 c_B)y_1 + c_B y_2 + L(y)$$

$$+ \alpha \int_0^\infty \int_0^\infty C_{n-1}(y_1 - t, y_2 - u) f(t, u) dt du$$

where  $n=1,2,\dots,N$ ,  $y \in R_2$  and  $C_0(x) = 0$  for all  $x$ .

The first result is of a purely technical nature, but provides the properties of  $C_n(\cdot)$  and  $G_n(\cdot)$  crucial to the characterization theorem.

THEOREM II.5.1: Under the assumptions made above, the following hold.

1.  $G_n(\cdot)$  is continuous and strictly convex.
2. All sets of the form  $Q_n(\beta) = \{y \in R_2; G_n(y) \leq \beta\}$  are compact and convex for each bounded  $\beta \in R$ .
3. The point  $y_n(x)$  that solves the constrained dynamic program exists and is unique. Hence,  $\inf$  can be replaced by  $\min$  for each bounded  $x \in R_2$  and  $n=1,2,\dots,N$ .



4.  $C_n(\cdot)$  is convex and nonnegative on  $R_2$ .

The optimal policy,  $y_n(x)$  is completely characterized by a point  $S_n = (S_{n,1}, S_{n,2})$  and three scalar functions  $r_n(\cdot)$ ,  $w_n(\cdot)$ , and  $q_n(\cdot)$  as graphically described in Figure 6. Each arrow represents the path from an initial inventory (the tail) to its beginning inventory (the head) by some production combination. In Region I both A and B are used and notice that the starting inventory level will always be  $S_n$ . In Region III only type B is used so the starting stock always lies on the graph of  $q_n(\cdot)$ . In Region II only A is used. The path from  $x$  to  $y_n(x)$  in this region is a line parallel to  $w_n(\cdot)$ . It is also important to point out that  $r_n(\cdot)$ ,  $w_n(\cdot)$  and  $q_n(\cdot)$  are functions of  $x_1$ .

Many of the assumptions in this example can be relaxed or removed so the model is more realistic. The demands and costs may be nonstationary, the production capacities may be finite, unmet demands may be lost, fixed lead times may be incorporated and the general m-process-n-product case follows naturally.

In summary, then the analysis presented in these section allows one to look at the economics of regionalization, including the location of community blood centers, the allocation of hospital blood banks and transfusion services to them, and the costs of different structures for a region. As mentioned at the beginning, these are not the only variables which one would want to consider in implementing a regional system. However, they are extremely important variables since they affect the system operation and its short and long-run costs. These costs in turn affect the subsequent charging mechanisms for blood and components to the patient. This quantitative analysis is a guide to making important decisions in a more enlightened manner.

#### 6) Concluding Remarks

Finally, it is hoped that the preceding analysis has communicated to the reader some of the flavor of past and current research in several pertinent methodological areas of Operations Research: Inventory Theory, Facility Location, Vehicle Scheduling, and to a small extent Mathematical Programming and Optimization of Stochastic Models. In addition it is hoped the reader will be interested in pursuing the design and analysis of models which better measure health care delivery phenomena and which lead to better decisions for individuals and institutions in this important societal area.

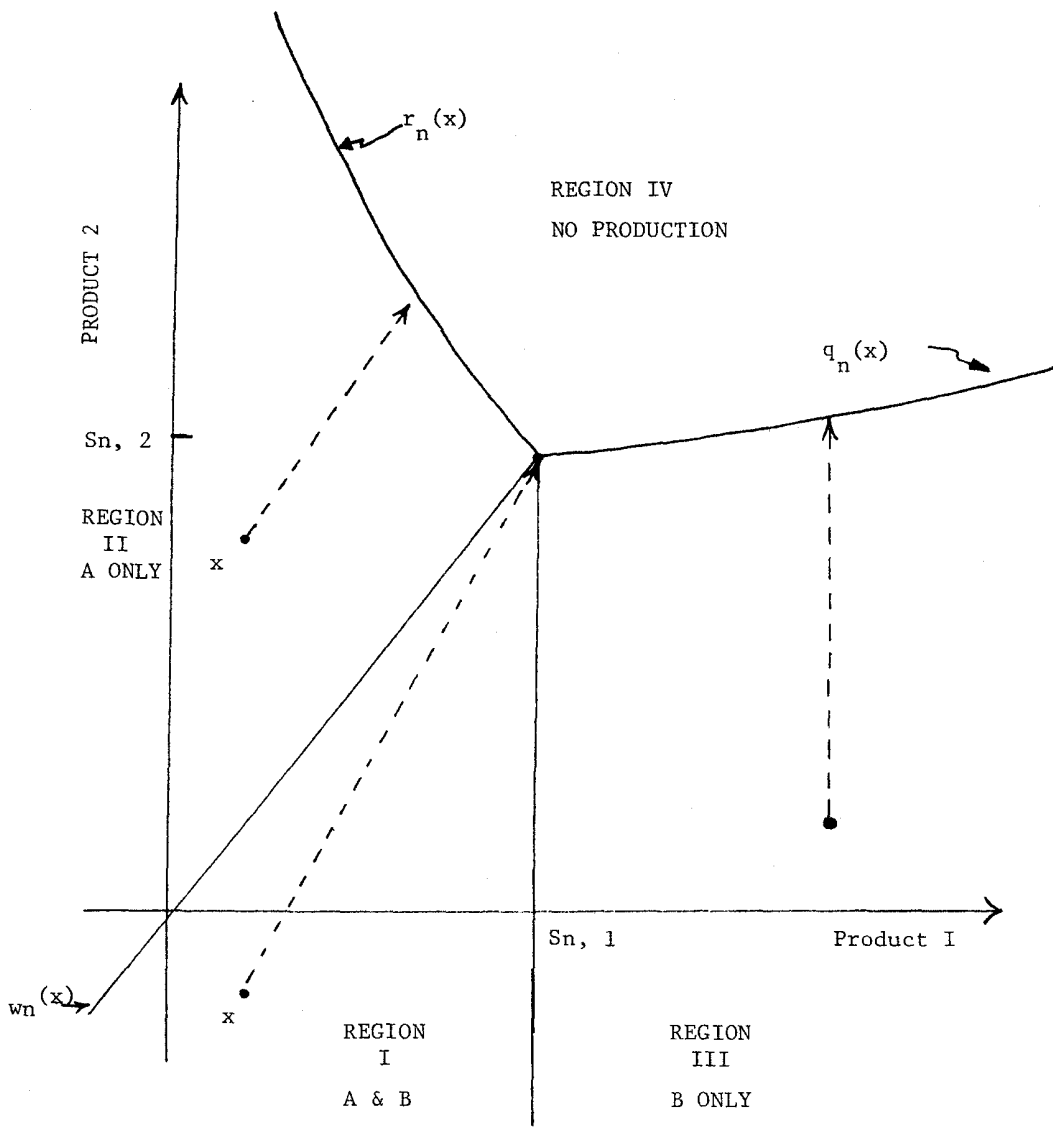


FIGURE 6

Characterization of the Optimal Policy in Period  $n$

## REFERENCES

- Arrow, K.J., S. Karlin and H. Scarf, eds. Studies in the Mathematical Theory of Inventory and Production, Stanford University Press, Stanford, Calif. (1958).
- Barlow, R., L. Hunter and F. Proschan, "Optimum Checking Procedures," J. Soc. Ind. Appl. Math., 11, 4 (1963), pp. 1078-1095.
- Calvo, A.B. and D.H. Marks, "Location of Health Care Facilities: An Analytical Approach," Socioeconomic Planning Science, 7 (1973), pp. 407-422.
- Cohen, M.A. and W.P. Pierskalla, "Management Policies for a Regional Blood Bank," Transfusion 15 (1975), pp. 58-67.
- Cohen, M.A. and W.P. Pierskalla, H.S. Yen, I. Or, "Regionalization and Regional Management Strategies in the Blood Banking System: An Overview." Proceedings of the Second National Symposium on the Logistics of Blood Transfusion Therapy, June 19-20, 1975. Sponsored by the Michigan Association of Blood Banks and the Michigan Community Blood Center.
- Cooper, L., "Location-Allocation Problem," Operations Research, 11 (1963).
- Cooper, L., "Heuristic Methods for Location-Allocation Problems," SIAM Review, 6 (1964).
- Cooper, L., "The Transportation-Location Problem," Operations Research, 20, (1972).
- Derman, C., "On Minimax Surveillance Schedules," Naval Research Logistics Quarterly, 8 (1961), pp. 415-419.
- Deuermeyer, B. and W.P. Pierskalla, "A By-Product Production System With An Alternative," Management Science, 24, (1978), pp. 1373-1383.
- Eddy, David M., "Screening for Cancer: Theory, Analysis, and Design," Department of Family, Community, and Preventive Medicine, Stanford University, (1978).
- Eddy, David M., "Implementation Plan: Part III. Rationale for the Cancer Screening Benefit Program Screening Policies," Department of Family, Community, and Preventive Medicine, Stanford University, for Blue Cross Association, (1978).
- Francis, R.L., and J.A. White, Facility Layout and Location, Prentice-Hall, Inc., Englewood Cliffs, N.J. (1974).
- Fries, B., "Optimal Ordering Policy for a Perishable Commodity With Fixed Lifetime," Operations Research, 23 (1975), pp. 46-61.
- Hutchinson, G. and S. Shapiro, "Lead Time Gained by Diagnostic Screening for Breast Cancer," Journal of the National Cancer Institute, 41 (1968), p. 665.
- Jennings, J., "Blood Bank Inventory Control," Management Science, 19 (1973), pp. 637-645. Also appears in Analysis of Public Systems, R. Keeney, P. Morse (editors), MIT Press, Cambridge, Mass., (1972).

- Jennings, J., "Inventory Control in Regional Blood Banking Systems," Technical Report Number 53, Operations Research Center, MIT (1970).
- Keller, J., "Optimum Checking Schedules for Systems Subject to Random Failure," Management Science, 21 (1974), pp. 256-260.
- Kirch, R. and M. Klein, "Surveillance Schedules for Medical Examinations," Management Science, 20 (1974), pp. 1403-1409.
- Kirch, R. and M. Klein, "Examination Schedules for Breast Cancer," Cancer, 33 (1974), pp. 1444-1450.
- Kolesar, P., "A Markovian Model for Hospital Admission Scheduling," Management Science, 16, 6 (February 1970), pp. B-384-396.
- Lincoln, T., and G.H. Weiss, "A Statistical Evaluation of Recurrent Medical Examinations," Operations Research, 12 (1964), pp. 187-205.
- McCall, J., "Maintenance Policies for Stochastically Failing Equipment: A Survey," Management Science, 11 (1965), pp. 493-524.
- McCall, J., "Preventive Medicine Policies," Rand Corporation, P-3368-1, (1969).
- Magnanti, T., B.L. Golden, and H.Q. Nguyen, "Implementing Vehicle Routing Algorithms," Technical Report No. 115, Operations Research Center, Massachusetts Institute of Technology (September 1975).
- Miller, H.E., W.P. Pierskalla, and G.J. Rath, "Nurse Scheduling Using Mathematical Programming," Operations Research, 24, 5 (September-October 1976), pp. 857-870.
- Nahmias, S. and W.P. Pierskalla, "Optimal Ordering Policies for a Product That Perishes in Two Periods Subject to Stochastic Demand," Naval Research Logistics Quarterly, 20 (1973), pp. 207-229.
- Nahmias, S., "Optimal Ordering Policies for Perishable Inventory," Operations Research, 23 (1975), pp. 735-749.
- Nahmais, S., "Inventory Depletion Management When the Field Life is Random," Management Science, 20 (1974), pp. 1276-1283.
- Nahmais, S. and W.P. Pierskalla, "A Two-Product Perishable/Nonperishable Inventory Problem," SIAM J. Applied Math., 30, No. 3 (1976), pp. 483-500.
- Or, I., "Traveling Salesman-Type Combinatorial Problems and Their Relation to the Logistics of Regional Blood Banking." Doctoral Dissertation, Northwestern University, Evanston, Illinois - 60201 (1976).
- Or, I., and W.P. Pierskalla, "BTAP: A Computer Program to Obtain Solutions to the Blood Transportation-Allocation Problem and Other Traveling Salesman Type Problems," Technical Report, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois (August 1976).
- Or, I., and W.P. Pierskalla, "A Transportation Location-Allocation Model for Regional Blood Banking," AIIE Transactions, (Summer 1979).
- Pierskalla, W.P. and J.A. Voelker, "A Survey of Maintenance Models: The Control and Surveillance of Deteriorating Systems," Naval Research Logistics Quarterly, 23, 3 (1976), pp. 353-388.

- Pierskalla, W.A. and J.A. Voelker, "A Model for Optimal Mass Screening and the Case of Perfect Test Reliability," Technical Report No. 3, Department of Industrial Engineering and Management Science, Northwestern University, Evanston, Illinois - 60201 (1977).
- Pierskalla, W.P. and C. Roach, "Optimal Issuing Policies for Perishable Inventory," Management Science, 18 (1972), pp. 603-615.
- Pierskalla, W.P. and J.A. Voelker, "Mass Screening With a Heterogeneous Population," Working Paper, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois (1978). Submitted to Operations Research.
- Prastacos, G.P. and E. Brodheim, "A Regional Blood Distribution Model," Working Paper, The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania (1977).
- Prastacos, G.P., "Allocation of Perishable Inventory," Technical Report No. 70, Operations Research Group, Columbia University, New York, New York (January 1977).
- Prorok, P., "On the Theory of Periodic Screening for the Early Detection of Disease," Unpublished Ph.D. Dissertation, University of New York at Buffalo (1973).
- Prorok, P., "The Theory of Periodic Screening, I: Lead Time and Proportion Detected," Advances in Applied Probability, 8 (March 1976), pp. 127-143.
- Prorok, P., "The Theory of Periodic Screening II: Doubly Bounded Recurrence Times and Mean Lead Time and Detection Probability Estimation," Advances in Applied Probability, 8 (September 1976), pp. 460-476.
- Roeloffs, R., "Minimax Surveillance Schedules With Partial Information," Naval Research Logistics Quarterly, 10 (1963), pp. 307-322.
- Roeloffs, R., "Minimax Surveillance Schedules for Replacement Units," Naval Research Logistics Quarterly, 14 (1967), pp. 461-471.
- Schwartz, M. and H. Galliher, "Analysis of Serial Screening in an Asymptomatic Individual to Detect Breast Cancer," Technical Report, Department of Industrial and Operations Engineering, College of Engineering, University of Michigan (1975).
- Thompson, D. and R. Disney, "A Mathematical Model of Progressive Diseases and Screening," presented at the November 1976 ORSA/TIMS Meeting.
- Vienott, A.F., Jr., "Optimal Policy for a Multi-Product, Dynamic, Non-stationary Inventory Problem," Management Science, 12 (1965), pp. 206-222.
- Vienott, A.F., Jr., "The Status of Mathematical Inventory Theory," Management Science, 12 (1966), pp. A745-A777.
- Voelker, J.A., "Contributions to the Theory of Mass Screening," Ph.D. Dissertation, Northwestern University, 1976.
- Wendell, R.E. and A.P. Hurter, "Location Theory, Dominance, and Convexity," Operations Research, 21 (1973), pp. 314-320.
- Wendell, R.E. and A.P. Hurter, "Optimal Locations on a Network," Transportation Science, 7 (1973), pp. 18-33.

Yen, H., "Inventory Management for a Perishable Product Multi-Echelon System,"  
Ph.D. Dissertation, Northwestern University, 1975.

Zelen, M. and M. Feinleib, "On the Theory of Screening for Chronic Diseases,"  
Biometrika, 56 (1969), p. 601.

Zelen, M., "Problems in the Early Detection of Disease and the Finding of  
Fault," presented at the meeting of the International Statistics Institute,  
Washington, D.C. (1971).