Health Care Applications of Optimal Control Theory

by

Stefanos A. Zenios

Submitted to the Department of Electrical Engineering and
Computer Science
on May 23, 1996, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Operations Research

Abstract

This investigation is motivated by two important health care problems: group testing for HIV, and organ allocation to patients on the transplant waiting list. In the first part of the thesis, we investigate group (or pool) testing as a cost-effective method for screening donated blood products for HIV, and for estimating the prevalence of HIV. Rather than test each sample individually, this method combines various samples into a pool, and then tests the pool. We develop a hierarchical statistical model that relates the HIV test output to the antibody concentration in the pool, thereby capturing the effect of pooling together different samples. The model is validated using data from a variety of field studies. Using this model we develop novel group testing strategies that can be used to identify infected individuals and to estimate the prevalence of HIV. A simulation study shows that significant cost savings can be achieved without compromising the accuracy of the test.

In the second part of the thesis we investigate the allocation of kidneys to patients on the transplant waiting list. Using a queueing reneging model for the transplant waiting list, and a fluid population model that integrates the waiting list with the process of organ rejection, we develop novel kidney allocation policies. These policies distribute kidneys based on a dynamic index that captures criteria of medical efficiency and equity. To analyze the performance of these policies we develop a state of the art simulation policy model. This model dynamically tracks patients in the waiting list and post-transplant patients. The model is calibrated and validated on data from the United Network of Organ Sharing (UNOS), and the United States Renal Data System (USRDS). The output from the model provides estimates for cost, health outcomes and equity.

To demonstrate the computer simulation model we use it to test the performance of the dynamic index policies under a hypothetical waiting list scenario. Our results show that, at least for the hypothetical scenario considered here, there exists a weighted combination of the dynamic index policies that dominates the allocation policy currently used in the United States.

Part II

Allocation of Kidneys to Patients on the Transplant Waiting List

Chapter 3

Motivation

End Stage Renal Disease (ESRD), also known as chronic kidney failure, is a fatal disease unless treated with dialysis or kidney transplantation. Kidney transplantation is not only the most cost effective treatment, but also the treatment that improves the quality of life of the patients the most [1]. The first successful kidney transplantation was performed in 1954. In the following two decades, the procedure became widely used but the success rates were disappointingly low. The breakthrough came in 1978 with the introduction of the immunosuppresive drug cyclosporine. Since then, the success of kidney transplantation has improved dramatically and the post operative survival is now almost 100% [2]. The one year organ survival is between 79% and 90% [3]. Last year, approximately 40,000 transplantations were performed worldwide, with 25% of them in the United States [1].

However, the success of organ transplantation has created a dramatic shortage of organs. It is not uncommon for ESRD patients to wait for 21 months or more before a suitable donor is found; in the United States the median waiting time for candidates receiving transplants from cadaveric donors was 21 months in 1994. This raises the issue of organ allocation. Although in an environment with unlimited supply of organs, the organs should be allocated based only on expected clinical effectiveness, in an environment of serious organ shortage the organs should be allocated based on a weighted combina-

tion of medical benefit and equity. Allocation policies implicitly divide the patients to those that will benefit from transplantation and those that will not. A notion of fairness demands that this division is almost random, essentially a lottery. Therefore, to ensure the ethical integrity of any allocation policy, policy makers must understand how the policy takes into account clinical effectivess and equity [4].

Several empirical studies have analyzed the allocation policy that is currently implemented in the United States. The incomplete picture emerging from these studies is one of a persistent pattern of socioeconomic inequity [6, 7, 8, 9]. Furthermore, the Office of Inspector General have found that the current policy falls short of meeting the expectations of the public [10]. Therefore, to maintain public confidence in kidney transplantation, we must reassess the current policy and consider new policies.

In order to do that, it is necessary to develop a comprehensive policy model that dynamically tracks the population in the waiting list for kidney transplantation, as well as the post-transplantation population. Policy models have been used extensively in health policy analysis, the coronary heart disease policy model is perhaps the most well known example [11]. However, the success of these models depends on their realism, and the models that are specifically designed for studying kidney transplantation are not sufficiently realistic because they fail to explicitly capture the dynamics of the disease and post-transplantation survival; see for example Yuan et al. (1994) [13].

In contrast, in this thesis we develop a kidney transplantation policy model that explicitly captures these dynamics. The model is calibrated on data from the United Network of Organ Sharing (UNOS) and the United States Renal Data System (USRDS). The distinguishing characteristic of our model is that because of its versatility, it can be used to compare different allocation policies. The output from the model will include cost as well as measures of clinical effectiveness and equity.

However, comparing different policies may not be enough. What is also needed is an analytical framework for developing new policies. This is a second contribution of our research. We propose an analytical framework that complements several candidate

policies that are presented in the literature. Our framework is unique in that it employs a mathematical objective function that combines clinical effectiveness and equity. Motivated from this objective, we develope simple heuristic policies that are easy to implement. Although our objective is neither unique nor the best, it provides a starting point for future investigations. The derived policies exhibit a rich structure and provide useful insights.

Although a careful analysis of allocation polices can lead into better policies, it cannot relieve the organ shortage problem. This problem can be resolved only by increasing the supply of organs. Providing financial incentives to the family of the deceased is a method that is increasingly discussed as a potential way for inducing organ donation [14]. However, before it is recommended that financial incentives are provided, it is necessary to establish their cost utility. The analytical framework developed in this thesis provides the methodology for estimating cost utility. In particular, our model can provide estimates for the marginal cost benefit from an increase in the organ supply. These estimates can be used to establish the cost effectiveness of several financial incentive programs.

The findings from this thesis are expected to be of particular relevance to the Health Care Financing Administrations (HCFA) and the Medicare program. Specifically, in the US the cost of caring for ESRD patients is covered almost exclusively by the HCFA ESRD program since 1973, and the total cost of the ESRD program was US$5 billion in 1990. However, the ESRD program expanded at a rate substantially higher than the one predicted in 1973 and thus attracted considerable criticism. Our results can provide reliable estimates of the future cost and scope of the program. The estimates can be used both for planning and for resource allocation.

Although the results from this thesis focus on kidney transplantation, the underlying methodology is very general and can be adapted to study other transplantation programs, such as heart, liver, pancreas and heart-lung transplantation. Judicial use of our methodology is expected to provide reliable estimates of the future costs and effectiveness of these programs. These estimates will contribute to the ongoing debate of rationing

limited health care resources among new and expensive medical technologies [5]. Our methodology can also be used to study the efficacy of imperfect vaccines for AIDS.

The remainder of this Chapter is organized as follows: Section 3.1 gives a general overview of kidney transplantation and organ allocation. Section 3.2 gives a brief review of the literature about kidney transplantation. Finally, section 3.3 gives an overview of the thesis.

3.1 Kidney Allocation: An Overview

The allocation of kidneys in the United States is decentralized and is performed by 73 Organ Procurement Organizations (OPO). These OPOs are responsible for recovering and allocating organs within specific geographic regions. For example, the New England Organ Bank is responsible for recovering organs in the states of Massachusetts, Maine, New Hampshire and Vermont, and for allocating these organs to transplant candidates in these states. The activities of these OPOs are coordinated by the United Network of Organ Sharing (UNOS). UNOS maintains a centralized data base about all transplants performed and all transplant candidates, and oversees the allocation policies used by the 73 OPOs. Therefore, although each OPO can adopt a different allocation policy, its policy must be approved by UNOS. As a result, all OPOs adopt allocation policies that are similar to a "standard" policy, which we call the UNOS allocation policy; see Barton and Kallich (1994) [15] for a lucid description of the US allocation system.

In this section we will present a detailed description of the UNOS allocation policy. However, to better understand the rationale behind this policy, it is first necessary to describe the process of organ rejection and its relation to the immune system. The immune system is responsible for protecting the human body from any foreign agents, particularly parasites, bacteria and viruses. However, in the same way that the system responds to these intruders it also responds to a foreign transplanted organ. In particular, if the proteins on the surface of cells of the organ are different from the proteins of the

recipient, then the immune system will produce antibodies against the organ that will cause organ (or graft) rejection; graft is the medical term for the transplanted organ.

Organ rejection can be either *acute* or *chronic*. Acute rejection occurs when the recipient has antibodies against the organ proteins at the time of the transplantation. To prevent acute rejection, the blood of the recipient should be tested against the blood of the donor. If the test is negative, then this implies that the recipient does not have antibodies against the organ proteins and the transplant operations can be performed. If the outcome of the test is positive, then the organ is not transplanted and the recipient is referred to as being *presensitized* to the donor. The process of testing for antibodies to the donated organ is known as crossmatching. Because of time constraints, only patients that are presensitized to a large number of individuals are crossmatched. Highly sensitized individuals are transplant candidates that are crossmatched positive with more than 60% of a random sample of donors. Specifically, every candidate is crossmatched against a random sample of donors, and the proportion of tests with a positive outcome gives the candidate's panel reactive antibody (pra) level. Individuals with pra greater than 60% a will be referred to as beiing presensitized, whereas patients with pra less than 60% will be referred to as being non-presensitized; see [1] for further details.

Clearly, cross-matching reduces the chances of acute rejection. To reduce the chances of chronic rejection, it is necessary to match the HLA types of the recipient and the donor. The HLA type of an individual is a combination of six proteins that are present on all cells. These proteins appear in three pairs, or loci, known as the A, B and DR loci. Organs that match with the recipient at all six proteins have minor chances of chronic rejection. However, organs matched at less than six proteins have higher chances of chronic rejection. For further information about HLA matching see [1].

Age is another factor that is known to affect the chances of graft rejection. Specifically, older patients are rarely considered for transplantation because of documented higher chances of graft rejection [32].

UNOS uses a policy that explicitly considers the factors described above before al-

locating the organs. The policy is based on a hierarchical point scheme that allocates higher priority to the candidate with the higher number of points; see UNOS (1995) [34]. The hierarchical scheme is as follows :

1. **Blood Type Compatibility:** Organs must be allocated to patients with a blood type that is compatible with the blood type of the donor. In addition, organs of blood type O must be transplanted to patients of the same blood type. The only exception is when there exists a *zero antigen mismatched* candidate; a zero antigen mismatch occurs when the donor and the candidate have the same proteins identified at each pair of the A, B and DR loci (see Table 3.1 for an example). In this case, the organ must be offered to the zero antigen mismatched candidate.

2. **Mandatory Sharing of Zero Antigen Mismatched Kidneys:** If there exists a zero mismatched candidate in the data base maintained by UNOS, then the organ must be offered to this candidate.

3. **The Point Allocation System:** If there is no zero antigen mismatched candidate, the organ is allocated to the *local* candidate that has the higher number of points; a candidate is local if (s)he is registered with the OPO that procured the organ. The points are computed as follows:

   (a) Waiting Time: One point is assigned to the patient waiting the longest and fractions of points are assigned to all other patients according to their relative position in the waiting list. For example, if the kidney is of blood type O and there are 75 candidates of type O, then the candidate with the longest waiting time receives $\frac{75}{75}1 = 1$ points, the second candidates receives $\frac{74}{75}1 = .98667$ points, and so forth. In addition, each candidate receives one point for each full year (s)he has spent in the waiting list.

   (b) Quality of Antigen Matching: Points are assigned to candidates based on the number of mismatches between the candidate's tissue type and the donor's tissue type. The formula used is as follows:

Head	HLA-A	HLA-B	HLA-DR
Donor	2,3	8,44	2,4
Candidate	3,2	8,44	2,4

Table 3.1: An example of a zero mismatched kidney. Observe that the same antigens are identified at each pair of the the A, B and DR loci. It should be emphasized that although this kidney is zero mismatched, it is only matched at four out of the size loci. The antinges at loci A1 and A2 are different.

- If there are no B or DR mismatches, the candidate receives 7 points.
- If there is 1 B or DR mismatch, the candidate receives 5 points.
- If there are 2 B or DR mismatches, the candidate receives 2 points.

(c) Panel Reactive Antibodies: Pre-sensitized patients are assigned 4 points.

The allocation scheme described above has been in effect since July 31, 1995. Before that, a slightly different allocation scheme was in place. In this thesis, we will refer to the current allocation scheme as the *new UNOS policy* and to the previous scheme as the *old UNOS policy*.

Several characteristics of the policy are worth emphasizing. First, the policy forces zero mismatched organs to be allocated nationally. The rationale behind this is that zero mismatched organs are associated with superior graft survival; see Takemoto et al. (1994) [16]. Second, the policy assigns higher priority to better matched candidates; this policy is also justified by evidence presented in Takemoto et al. (1994) [16]. Third, the policy gives high priority to pre-sensitized patients. The rationale is that a highly sensitized patient must be offered a kidney an average of three to five time before a negative crossmatch is found; recall that pre-sensitized patients cross-match positive with more that 60% of an unbiased sample of donors. Finally, the policy attempts to promote equity by giving points based on the total waiting time. In particular, 7 years in the waiting list are "equivalent" with a kidney with no B or DR mismatches.

Although the UNOS policy attempts to balance medical benefit and equity, it is inherently unfair. To see this, consider the following hypothetical scenario: Suppose that ESRD patients can be of two HLA types X and Y, but donors are only of type

X. Furthermore, suppose that the supply of organs is equal to the demand of X type organs. Under the UNOS policy, type X candidates will be receiving organs almost right after they register in the waiting list, whereas type Y patients will rarely receive transplantation!

The last example provides a partial explanation for the current situation in the US. Specifically, in the US African American candidates wait on average longer than Caucasians [6]. This is partly because the HLA types are different between different races, and African Americans constitute 30% of the ESRD patients but only 10% of donors.

Racial inequity is one of several forms of inequity arising as a consequence of the point allocation system. Other forms of inequity are across age, gender and HLA type. The results in this thesis provide additional insights about the reasons for these inequities. Our analysis compares different allocation policies, and investigates how other policies balance the trade off between medical benefit and equity.

Although the UNOS allocation system takes into consideration several factors that can enhance graft survival, there are some additional factors that predict graft survival. A recent study by Chertow et al. (1996) [18] shows that the donor's and recipient's race, gender and age, the recipient's body surface area (bsa) and the history of previous transplantations are also strong predictors of graft survival; the body surface area is a clinical index of body size. Incorporating all these factors into an allocation policy is clearly unacceptable and it contradicts any notion of fairness and ethics. On the other hand, not investigating the behavior of such an allocation policy is doing injustice to the scientific community. So, in this thesis, we simulate the kidney transplant waiting list using such an inherently unfair allocation policy. The objective of this exercise is not to advocate the implementation of such a policy but to sharpen our understanding about the trade off between medical utility and equity.

3.2 Literature Review

The literature on kidney transplantation and allocation can be conveniently divided into three groups: (a) Empirical studies that identify factors that are associated with graft survival; (b) Empirical studies that identify factors that are associated with reduced access to transplantation; and (c) Studies that attempt to synthesize the lessons from the empirical studies to propose and test alternative allocation policies. In this section, we summarize the most influential studies in each of these three groups.

Graft Survival Studies:

The literature on this subject is vast. Several studies have shown that graft survival improves with HLA matching. In a recent study, Held et al. (1994) [17] analyzed data from 30,564 first time transplant recipients and showed that patients receiving zero mismatched kidneys have one year graft survival of 84.3%. This compares with 77.0% one year survival for grafts with four mismatches. These findings were confirmed by a more recent study by Chertow et al (1996) [18]. In this study, the authors use data from 31,515 transplantations performed between 1987 and 1991 and identify several factors that are strongly associated with graft survival. In addition to tissue matching, the authors have also shown that the donor's and recipient's race, gender and age, the recipient's body surface area, previous history of transplantation, and the cold ischemia time are also strong predictors of graft survival; cold ischemia time is the time that elapses between the procurement of the organ and the actual transplantation. However, both studies show that the survival difference between perfectly matched and poorly matched organs is small. This finding fuels a controversy in the nephrology community about the importance of HLA matching in organ allocation.

In a different set of studies, Brenner et al. (1992) [19] and Terasaki et al. (1994) [20] suggest that the size of the organ is also of essence and they propose the "hyperfiltration hypothesis". In more recent studies, Miles et al. (1996) [21] and Gaston et al. (1996) [21] argue that the impact of the organ size is insignificant.

Earlier studies have also attempted to study the impact of socioeconomic conditions

on graft survival. Kalil et al (1991) [23] analyzed data from 202 transplants to conclude that income level affects graft survival through a mechanism that appears to be independent from compliance and other known risk factors.

Age is another factor that attracted attention by the scientific community. In a very detailed study, Shah et al. (1988) [32] provided evidence that transplantation should be the preferred treatment for ESRD patients between 50 and 64 years old. This paper marked a change in transplant procedures and now transplantations are routinely performed to patients in that age group.

Access to Transplantation:

The widespread success of kidney transplantation in the late 70's and early 80's motivated a series of studies that investigated factors that affect access to transplantation. Gaylin et al. (1993) [7] analyzed a national sample of patients that started dialysis in 1986 and 1987. Using time until transplantation as the major outcome of interest, they have found that patients with pre-existing medical conditions, such as heart disease, experience longer waiting times. They have also confirmed earlier studies showing that women, nonwhite patients and lower income patients also experience longer waiting times. In a different study, Ellison et al. (1993) [24] analyzed transplant data from UNOS and concluded that the mean waiting time for nonwhite candidates is increasing at a faster rate than that for whites.

Synthesis:

The main conclusion from the studies described above is that under the current allocation policies, access to transplantation is restricted for minorities and women. Although this conclusion is not contested by anyone, its resolution became the subject of a heated debate. In fact, there are two schools of thought. Starzl et al. (1987) [25] argue that allocation decisions should be based on a point system that takes several factors into consideration; factors included in the Starzl paper are waiting time, antigen matching, pra, medical urgency and logistical factors. The authors also argue that it is unacceptable to base allocation decisions only on tissue matching because this would create a "genet-

ically determined bias in which some patients would never be able to find well matched kidneys". To counteract this argument, Opelz and Wujciak developed and analyzed a policy that allocates kidneys based on tissue type compatibility and the rarity of the recipient's tissue type. By taking the rariry of the tissue type into account, the policy ensures that all patients, independently of their tissue type, are guaranteed a minimum level of access to transplantation; see Wujciak and Opelz (1993) [26] and Opelz and Wujciak (1995) [27]. In an earlier study, Gjertson et al. (1991) [3] also argued in favor of HLA matching. Using data from 22,190 transplantations, they have concluded that allocating kidneys based on a hierarchical HLA matching scheme would enhance graft survival by 5 percentage points at ten years. However, selection bias in the data, and restrictive assumptions in their analysis cast doubts on the reliability of their estimate.

Operations Research and Decision Making Literature:

So far, we have reviewed the clinical literature about kidney transplantation. Despite the wealth of interesting operational and decision problems that are motivated by the problem of kidney allocation, the Operations Research literature on the subject is very limited. Righter (1989) [29], and references therein, study several variations of the stochastic assignment problem. The instance of problem studied by Righter assumes that there are n candidates that require transplantation and organs arrive sequentially according to a Poisson process. The objective is to assign organs to the candidates so as to maximize the total expected return. This is a sequential resource allocation problem and the author shows that under certain conditions the optimal policy is a threshold policy. David and Yechiali (1985,1995) [30, 31] study two problems that are also motivated by the problem of kidney allocation. Their first paper considers the problem of a single candidate that expects several transplant offers and wishes to decide which offer to accept. In their most recent paper, the authors generalize the original problem to consider M candidates and a random stream of N organs. The authors study this problem under several assumptions on the problem parameters and derive optimal policies.

In the Medical Decision Making community, Yuan et al. (1994) [13] develop a com-

puter simulation model which they use to assess the tradeoff between outcomes and access to transplantation (or equity). Their main conclusion is that the final choice about an allocation policy requires a value judgement on how to tradeoff clinical efficiency with equity.

3.3 Overview of this Investigation

The existing literature demonstrates that devising an "effective" kidney allocation policy is a non-trivial task. This is argued very elegantly in a brief article by Sanfilippo (1993) [28]. In that article, the author argues that there are six major factors that should be considered when making kidney allocation decisions:

1. Immunological constraints such as blood compatibility and presensitization.
2. Regulatory constraints such as congressional regulations.
3. Outcomes such as life expectancy and quality of life.
4. Access to transplantation.
5. Costs to society and to patients.
6. Effects on the total organ pool.

Clearly, any feasible allocation policy must be restricted by the first two constraints but can have a variable effect on the last four factors. This is further complicated because there exists a strong interrelationships among the last four factors so that it is impossible to simultaneously maximize all of them. The objective of this thesis is to develop the theoretical framework to study the interrelationship between these four factors, and in particular factor 3 (outcomes) and factor 4 (access). A companion objective is to develop a computer simulation model that supplements the theoretical framework and can provide quantitative answers to questions such as: "How will policy X affect access to transplantation and health outcome?"

The remainder of this thesis is organized as follows. In Chapters 4-5 we present two models that provide sharp insights about the impact of different allocation policies on access to transplantation and health outcomes. The first model is described in Chapter 4 and is a multiclass queueing system with reneging. This system captures the dynamics of the kidney transplant waiting list. Probabilistic analysis of this queuing system reveals the relationship between the allocation policy and several performance measure that reflect access to transplantation. This analysis also shows that the queueing system satisfies certain conservation laws that quantify the following intuition: improving access for one class of patients has adverse effect on other classes that compete for the same limited resources.

The second model is a *fluid demographics model* and is described in Chapter 5. This model captures the demographics of the ESRD population both before and after transplantation. Embedded in this model are several alternative mathematical objectives and constraints that reflect the two important variables of outcomes and access. These give rise to alternative formulations of optimal control problems where the objective is to obtain organ allocation policies that balance the trade off between clinical effectiveness and equity. Although we do not claim that the proposed objective and constraint fully capture the controversial concept of equity, we still believe that they provide some valuable insights and provide a useful starting point. We show how to derive efficient heuristics for the formulated optimal control problems and we interpret the heuristics in light of the underlying problem of kidney allocation.

Chapter 6 presents a computer simulation model for Kidney Transplantation. This model generalizes the fluid model of Chapter 5 and captures the complexities of kidney allocation that cannot be captured by the analytically tractable fluid demographic model. The computer model utilizes a sophisticated *graft survival submodel* that accurately predicts the process of graft failure. This model is calibrated and validated using data from more than 30,000 transplantations.

The final Chapter of this thesis, Chapter 7, presents results from the computer simulation program that compares the new UNOS allocation policy to several allocation policies motivated by the analysis in Chapters 5 and 6. Concluding remarks are given at the end of the last chapter.

Chapter 4

A Simple Queueing Model for the Transplant Waiting List

Kidney transplantation is the best treatment for chronic kidney failure; not only does it improve the quality of life of the patients, it also increases their life expectancy. However, the supply of organs for transplantation has failed to meet the ever increasing demand. As a result, it is now estimated that the demand exceeds the supply by at least 18% (see Chapter 6), the size of the waiting list increases at a rate exceeding 4000 candidates per year, and the death rate in the waiting list exceeds 1200 candidates per year; see UNOS, 1995. In this environment, scarce organs must be allocated so as to enhance medical benefit and promote equity.

It is now well accepted that the objectives of equity and medical benefit are conflicting. Specifically, to enhance medical benefit, allocation policies must match the donor's and recipient's tissue types; i.e. incoming organs should be allocated to the most well-matched candidate. However, because the frequencies of tissue types differ between donors and recipients, policies that allocate organs based only on tissue type compatibility generate disparities in access to transplantation and induce long delays for certain patients. These delays are an indication of inequity.

Several allocation policies have been recommended to achieve a balance in medical

benefit and equity. Even though such policies can have dramatic consequences on health outcomes and access to kidney transplantation, they are implemented with little knowledge about their expected impact on these outcomes of interest. As a result, a recent report from the Office of Inspector General has concluded that the allocation policy that is currently implemented in the United States has failed to promote equity.

The profound failure of the existing allocation policy is an indication that neither the dynamics of the waiting list, nor the nature of the trade-off between medical benefit and equity are well understood. It is our premise that to develop effective and equitable allocation policies, it is necessary to better understand the dynamics of the kidney transplant waiting list. To do that, we develop two mathematical models that provide answers to the following fundamental questions:

1. How does the allocation policy affect the average waiting time for patients in the waiting list?
2. What are the desirable properties of an equitable, or fair, allocation policy? Is there an *equity* measure that quantifies these properties?
3. What is the exact nature of the trade off between medical benefit and equity? Are there any laws that quantify this trade off?
4. What are the desirable properties of a policy that promotes medical benefit?
5. What are the desirable properties of a policy that optimally balances the trade off between medical benefit and equity?

To address the first two questions we analyze a multiple class, multiple server queueing system with reneging. In this system, customers correspond to patients, customer classes correspond to tissue types, and servers correspond to donated organs. Patients renege because they may die while waiting for a transplant. Assuming that the system is Markovian, we develop an analytical expression for the transform of the waiting time in

the system. Using a decomposition result and asymptotic expansions, we develop asymptotic approximations for the average waiting time and access to transplantation for each patient class. Our results show that the average waiting time depends both on the supply of organs and the reneging rates of the patients. Moreover, our analysis also shows that allocating organs based on tissue matching generates differences in the average waiting times between patient classes. Taking the point of view that these differences reflect inequity in access to transplantation, we define the total inequity in the system as the pairwise square difference in the average waiting times among different patient classes. The rationale behind this measure is that under an equitable policy, different patient classes should experience, on average, the same delays in the waiting list. To develop insights about the properties of an equitable policy, we obtain an allocation policy that minimizes the total inequity. This policy allocates a higher fraction of incoming organs to patients with lower reneging rates and coincides with the First Come First Transplanted policy (FCFT) when the reneging rate is independent of the patient class. However, this policy deviates from the FCFT policy when the reneging rates differ between patient classes. The relationship between allocation policies and waiting time is further clarified by a conservation law. This law states that giving preferential treatment to certain patient classes is afforded at the expense of less fortunate classes.

Although this queueing model contributes to a better understanding of the waiting list dynamics, it fails to further clarify the nature of the trade-off between equity and medical benefit. To clarify this trade-off, we develop a deterministic fluid model that integrates the waiting list dynamics with the population dynamics for patients with functioning transplanted organs. The model has several classes of patients and several classes of organ. Each patient class gives the patient's age, gender, race, health history, tissue type etc., and each organ class specifies the organ's tissue type and quality. Patients join the fluid model and flow between the different classes of the model. Similarly, organs are allocated to the patients and force an amount of flow out of the class receiving the organ and into the class of patients with functioning organs. Using simple properties of this

system we show that its state space satisfies simple conservation laws. These laws further illuminate the trade off between medical benefit and equity and quantify the intuition that to promote equity it is necessary to sacrifice medical benefit.

To develop allocation policies that optimally balance the two conflicting objectives, we analyze three optimal control problems. The first control problem is to allocate the flow of incoming organs so as to maximize the total number of quality adjusted life years (QALYs) in the system. The second problem is to allocate incoming organs so as to maximize the same objective but subject to constraints on the fraction of patients of each class that receive transplantation. The rationale behind these constraints is that they force the allocation policy to allocate no less than a prespecified number of transplants per patient class so that no class will receive a disproportionately small, or large, amount of transplants. The third problem is to allocated incoming organs so as to maximize a weighted combination of the total number of QALYs and equity in waiting times. In all three cases, the objectives and constraints reflect society's perspectives about medical benefit and equity.

Analyzing the three control problems is a mathematically difficult problem. Using Pontryagin's maximum principle, we show that the optimal policy for the first two problems is a dynamic index policy. We also develop simple approximations for these indices. The dynamic indices for the first problem give the marginal increase in QALYs, whereas the dynamic indices for the second problem are a weighted combination of the marginal increase in QALY's and the shadow price for the constraints. To analyze the third problem, we use a technique utilized by Wein, Zenios and Nowak (1996) to analyze a nonlinear optimal control problems. This technique gives a suboptimal heuristic that is derived from one step of the policy iteration algorithm. The derived policy is a dynamic index policy which adopts a weighted combination of the marginal increase in QALYs and the approximate marginal increase in equity. More importantly, this dynamic index policy generalizes the organ allocation policy that is currently being implemented in the United States.

The remainder of this chapter is organized as follows: Section 4.1 describes the multiple class reneging model and presents results from the steady state and asymptotic analysis. Following that, Section 4.2 derives some of the conservation laws that underlie the performance of the reneging system. Section 4.3 uses some of the lessons learned from the analysis of the reneging model to study the question of equity and to propose several quantities that capture this difficult-to-measure concept. The fluid demographics model is described in Chapter 5.

4.1 The Reneging Model:

The reneging model is more easily described in two steps. In the first step we analyze a single class M/M/1 queue with reneging, and in the second step we extend the analysis to the multiple class multiple server system.

In the single class single server system, customers arrive according to a Poisson process with arrival rate λ^+ and are serviced by an exponential server in a First Come First Served fashion; the service rate is λ^-. Customers can be lost before they are served if their sojourn time exceeds their expiration, or time-out, time. The expiration time is assumed to be exponentially distributed with rate μ; see Figure 4-1.

This system captures the general behavior of the waiting list for kidney transplantation when there is a single class of patients and a single class of organs. Customers correspond to patients, the server corresponds to transplants and reneging corresponds to the process of waiting list withdrawal; patients can withdraw from the waiting list either because they migrate or because they become ineligible for transplant. Although this model is a great simplification of the actual system that it attempts to describe, it provides a good starting point for our investigation.

The reader may wonder about the rationale behind the nontraditional notation used here. In particular, why use λ^- instead of μ for the service rate. The rationale comes from the literature on generalized queueing networks; see Gelenbe (1990). The simple

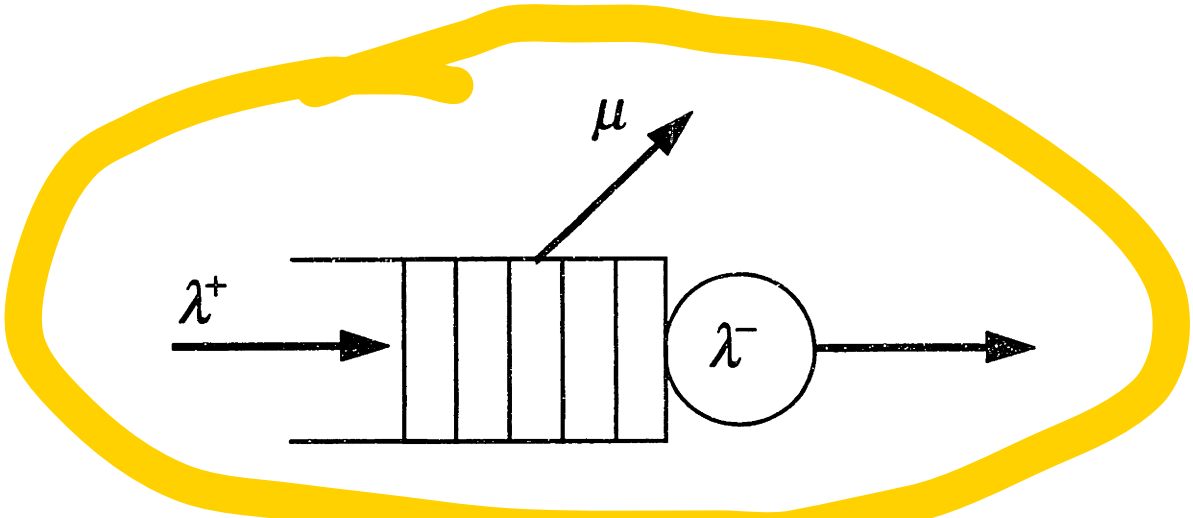

Figure 4-1: A simple reneging model for the single class kidney transplant waiting list: Patients arrive in a homogeneous Poisson process at a rate λ^+, organs arrive in a independent Poisson process at a rate λ^-. Patients leave the waiting list (renege) at a rate μ.

queueing system described here can also be presented as an infinite Markovian server with negative arrivals. In this context, the infinite server captures the time-out process and the negative arrivals model the incoming organs. Under this interpretation of our model, λ^- is the arrival rate of the *negative* arrivals and μ the service rate for our infinite server.

In our analysis, it is convenient to define two dimensionless quantities: the traffic intensity $\rho = \frac{\lambda^+}{\lambda^-}$ and the ratio $\nu = \frac{\lambda^-}{\mu}$ of the average time-out period to the average service time. Because our analysis is motivated by the problem of kidney allocation we assume that $\rho > 1$ and $\nu \gg 1$. This implies that the demand for organs exceeds the supply, and the waiting list withdrawal rate is much lower than the organ arrival rate. Moreover, because the steady state analysis of this queueing system does not give easy to use expressions, we will perform asymptotic approximations that assume that $\nu \to \infty$ and ρ remains fixed.

4.1.1 Single Class System and Fixed ρ and ν :

We start the analysis of the single class system with a Theorem that gives the stationary queue length distribution and the steady state distribution of the queueing time; the queueing time is defined to be the amount of time that a customer spends in queue and in service

Theorem 1 *Let π_k, $k \geq 0$ be the stationary queue length distribution for the single class M/M/1 queue with reneging. In addition, let L be the mean waiting list size, W be the steady state queueing time under the FCFS policy, and $G_W(s)$ the Laplace transform of the random variable W. The steady state distribution $\pi_k, k \geq 0$ and the steady state waiting list size are independent of the allocation policy and they satisfy*

$$\pi_k = \frac{1}{k!} \frac{\int_0^1 (z\rho\nu)^k (1-z)^{\nu-1} dz}{\int_0^1 \exp(\rho\nu z)(1-z)^{\nu-1} dz}, \tag{4.1}$$

and

$$L = \rho\nu \frac{\int_0^1 z(1-z)^{\nu-1} \exp(\rho\nu z) dz}{\int_0^1 (1-z)^{\nu-1} \exp(\rho\nu z) dz}. \tag{4.2}$$

Furthermore, the Laplace transform for the distribution of W is given by

$$G_W(s) = \frac{\mu}{\mu+s} \left(1 - G_Y(\mu+s)\right) + G_Y(\mu+s)\,, \tag{4.3}$$

where

$$G_Y(s) = \frac{\int_0^1 \exp(z\rho\nu)(1-z)^{\nu(s)-1} dz}{\int_0^1 \exp(z\rho\nu)(1-z)^{\nu-1} dz}, \tag{4.4}$$

and $\nu(s) = \nu + \frac{s}{\mu}$.

Proof: The proof of (4.1) and (4.2) is elementary and follows from the flow balance equations and from simple properties of the gamma and beta function. The proof of equation (4.3) is more interesting.

We study the system from the point of view of a newly arrived customer; using Kleinrock's notation we refer to this patient as the "tagged" customer (see Kleinrock,

1977). The waiting time for this customer can be expressed as the minimum of two random quantities: Y which gives the time the "tagged" customer leaves the system under the hypothetical scenario that his expiration time is infinite, and X which gives the expiration time of the "tagged" customer. Clearly,

$$W = \min(X, Y). \tag{4.5}$$

Equation (4.5) states that the "tagged" customer will exit the system either at the completion of service or at the expiration time, whichever occurs first.

Although the distribution of X is readily available, it is exponential with rate μ, the distribution of Y is more interesting and some work is needed: Let N be the random number of customers that are found in the transplant queue by our tagged patient. PASTA implies that the Laplace transform of Y satisfies

$$E[e^{-sY}] = \sum_{n=0}^{\infty} \pi_n E[e^{-sY}|N = n]. \tag{4.6}$$

Because the service discipline is FCFT, $Y|N = n$, denoted by Y_n, can be expressed as the sum of $n+1$ independent exponential variables $X_n, \ldots, X_0$,

$$Y_n = X_n + X_{n-1} + \ldots + X_0. \tag{4.7}$$

Equation (4.7) states that the "tagged" customer will exit the system only after all the customers ahead of him will depart from the system, and his service will be completed. The random variables $X_j : j = 0, \ldots, n$ have the following interpretation: X_n gives the time until the departure of the first customer and is exponentially distributed with rate $\lambda^- + n\mu$. In general, X_j gives the time from the departure of the $n-j+1$ customer until the departure of the $n-j$ customer and is exponentially distributed with rate $\lambda^- + j\mu$. Finally, X_0 is the time from the departure of the nth customer until the service completion of the "tagged" customer and is exponentially distributed with rate λ^-.

Equation (4.7) implies that the Laplace transform of Y_n is

$$G_Y^n(s) = \prod_{j=0}^{n} \frac{\lambda^- + \mu j}{\lambda^- + s + \mu j}. \tag{4.8}$$

Substituting equation (4.8) into equation (4.6), and using some simple properties of the beta and gamma function, we conclude that the Laplace transform of Y is

$$G_Y(s) = \frac{\int_0^1 \exp(z\rho\nu)(1-z)^{\nu(s)-1}dz}{\int_0^1 \exp(z\rho\nu)(1-z)^{\nu-1}dz}. \tag{4.9}$$

So far, we have characterized the distributions of X and Y. It now remains to obtain $G_W(s)$. Because $W = \min(X, Y)$, it follows that

$$\begin{aligned} G_W(s) &= \int_{y=0}^{\infty}\int_{x=0}^{y} e^{-sx}\mu e^{-\mu x} g_Y(y)dzdy + \int_{y=0}^{\infty}\int_{x=y}^{\infty} e^{-sy}\mu e^{-\mu x} g_Y(y)dzdy \quad (4.10) \\ &= \frac{\mu}{\mu+s}(1 - G_Y(\mu+s)) + G_Y(\mu+s). \quad (4.11) \end{aligned}$$

■

Several remarks are in order here. First, it is possible to use (4.3) to rederive Little's law; this provides a useful consistency check. To do that, use the standard property of Laplace transforms:

$$E[W] = -G'_W(0) \tag{4.12}$$

to compute $E[W]$. After some tedious algebra, we conclude that $E[W] = \frac{1}{\lambda^+}L$. It should be emphasized here that Little's law holds independently of the queueing discipline (Little, 1961). Second, it is possible to provide an alternative derivation of $E[W]$ by utilizing an auxiliary Markov process. The derivation is as follows: Let W_n denote the expected queueing time for the tagged customer *given* that there are n patients waiting ahead of him. Because the service discipline is FCFS, future customer arrivals do not affect the waiting time W_n. Therefore, to obtain W_n we consider a Markov chain where the external arrivals are turned off, and the initial number of customers is $n+1$; the $n+1$th

customer is the tagged customer. By conditioning on the next transition of this Markov chain, we see that W_n satisfies the recursion

$$W_n = \frac{\lambda^- + n\mu}{\lambda^- + (n+1)\mu}\left(\frac{1}{\lambda^- + (n+1)\mu} + W_{n-1}\right) + \tag{4.13}$$

$$+\frac{\mu}{\lambda^- + (n+1)\mu}\left(\frac{1}{\lambda^- + (n+1)\mu}\right) \tag{4.14}$$

$$= \frac{1}{\lambda^- + (n+1)\mu} + \frac{\lambda^- + n\mu}{\lambda^- + (n+1)\mu}W_{n-1}, \tag{4.15}$$

with initial conditions

$$W_0 = \frac{1}{\lambda^- + \mu}. \tag{4.16}$$

Solving the recursive equations (4.15)-(4.16), we conclude that

$$W_n = \frac{n+1}{\lambda^- + (n+1)\mu}. \tag{4.17}$$

It is then a simple exercise to show that

$$W = \sum_{k=0}^{\infty} \frac{k+1}{\lambda^- + (k+1)\mu}\pi_k \tag{4.18}$$

$$= \frac{L}{\lambda^+}. \tag{4.19}$$

It is worth emphasizing that $W_n \to \frac{1}{\mu}$, as $n \to \infty$. This states the intuitive fact that if the queue length is too large, then the expiration date will, almost certainly, precede the service time.

Theorem 1 provides the departure point for the analysis of several important quantities. We are particularly interested in the following quantities which measure the performance of the underlying kidney allocation process:

1. What is the fraction of customers that are serviced before the expiration date? Equivalently, what is the fraction of patients that receive transplantation?

2. What is the queueing time distribution *given* that the customer was serviced? Equivalently, what is the queuing time distribution *given* that the patient received transplantation?

3. What is the fraction of time that the queue is empty? Equivalently, what is the probability that an organ will become available while the queue is empty?

4. What is the *observed* steady state reneging rate (the observed reneging rate is not equal to μ because a fraction of the customers are serviced before their expiration time) ? Equivalently, what is the death rate in the waiting list?

To address these questions, let us first introduce some notation. Let τ denote the fraction of customers that are serviced before the expiration time, let $G_{W|T}(s)$ denote the conditional Laplace transform for the steady state queueing time *given* that service is received , let η denote the probability that the queue is empty and let σ be the observed steady state reneging rate.

Proposition 1 *In steady state, the fraction of patients receiving transplantation, the conditional Laplace transform of the queueing time given that service is received, the probability that a queue is empty, and the steady state observed reneging rate are given by the following expressions:*

$$\tau = G_Y(\mu), \tag{4.20}$$

$$G_{W|T}(s) = \frac{G_Y(\mu+s)}{G_Y(\mu)}, \tag{4.21}$$

$$\eta = \frac{\int_0^1 (1-z)^{\nu-1} dz}{\int_0^1 \exp(\rho\nu z)(1-z)^{\nu-1} dz}, \tag{4.22}$$

$$\sigma = \lambda + \frac{\int_0^1 z(1-z)^{\nu-1} \exp(\rho\nu z) dz}{\int_0^1 (1-z)^{\nu-1} \exp(\rho\nu z) dz}; \tag{4.23}$$

$G_Y(s)$ is defined in equation (4.4).

Proof. Let us start with equation (4.20). Let X denote the expiration time for a typical customer and Y the time the random customer enters service under the hypothetical

scenario that the expiration time is infinite. Then,

$$\begin{aligned} \tau &= P(X > Y) && (4.24)\\ &= \int_{y=0}^{\infty} P(X > y|Y = y) g_Y(y) dy dx && (4.25)\\ &= \int_0^{\infty} e^{-\mu y} g_Y(y) dy && (4.26)\\ &= G_Y(\mu). && (4.27) \end{aligned}$$

Next, consider equation (4.21). Let $f_{W|T}(w)$ denote the probability density of the waiting time given that the "tagged"customer was serviced; let T denotes the event that the "tagged" customer was serviced. Then,

$$G_{W|T}(s) = \int_0^{\infty} e^{-sw} f_{W|T}(w) dw. \tag{4.28}$$

However,

$$\begin{aligned} f_{W|T}(w) &= \frac{f_{W,T}(w,T)}{P(T)} && (4.29)\\ &= \frac{\int_{x=w}^{\infty} \mu e^{-\mu x} g_Y(w) dx}{P(T)} && (4.30)\\ &= \frac{e^{-\mu w} g_Y(w)}{G_Y(\mu)} && (4.31) \end{aligned}$$

Substituting (4.31) into (4.28) gives

$$G_{W|T}(s) = \frac{G_Y(\mu + s)}{G_Y(\mu)}. \tag{4.32}$$

Next, consider (4.21). Clearly, $\eta = \pi_0$ and the result follows from Theorem 1.

Finally, consider the death rate σ. This is equal to μL; where L is the steady state queue length. Therefore,

$$\sigma = \lambda + \frac{\int_0^1 z(1-z)^{\nu-1} \exp(\rho\nu z) dz}{\int_0^1 (1-z)^{\nu-1} \exp(\rho\nu z) dz}. \tag{4.33}$$

■

4.1.2 Asymptotic Expansions:

Unfortunately, because equations (4.1)-(4.4) are expressed in terms of integrals they provide little useful insights. To overcome this problem, we now use asymptotic expansions, and in particular Laplace's method, to derive asymptotically exact approximations for these equations; our approach imitates the approach in Cauffman, Puhlaski, Reimann and Wright (1994).

Theorem 2 *Consider the queueing system with customer arrival rate $n\lambda^+$, service rate $n\lambda^-$ and reneging rate μ. The following asymptotics hold as $n \to \infty$:*

$$\frac{L}{n} \sim \frac{\lambda^+ - \lambda^-}{\mu}, \tag{4.34}$$

$$W = (1 - \frac{\lambda^-}{\lambda^+})\frac{1}{\mu}, \tag{4.35}$$

$$\tau \sim \frac{\lambda^-}{\lambda^+}, \tag{4.36}$$

$$G_{W|T}(s) \sim \exp\left(-\frac{1}{\mu}\ln(\frac{\lambda^+}{\lambda^-})\right), \tag{4.37}$$

$$\eta \sim \frac{\sqrt{\pi}}{\rho^2 n^{\frac{3}{2}} \nu^{\frac{3}{2}}} \exp\left(-n\nu\left(\rho(1-1/\rho) + \ln(1/\rho)\right)\right), \tag{4.38}$$

$$\frac{\sigma}{n} \sim \lambda^+ - \lambda^-. \tag{4.39}$$

To prove this Theorem, we need the following lemma from Carrier, Krook and Pearson (1983).

Lemma 3 *Consider the following integral*

$$f(x) = \int_a^b e^{xh(t)} q(t) dt,$$

where all quantities are real. Suppose that $h(t)$ attains its maximum at the interior point t_0 and $h(t_0)$ exists and is twice differentiable. Then, the following asymptotic holds as

$x \to \infty$,

$$f(x) \sim q(t_0)\exp(xh(t_0))\left[-\frac{2\pi}{xh''(t_0)}\right]^{\frac{1}{2}}.$$

We can now outline the proof of Theorem 2.

Proof. Rather than present all the steps of the proof in excruciating detail, we will only derive the asymptotic expansion for L. Recall that

$$\begin{aligned} L &= \rho\nu\frac{\int_0^1 z(1-z)^{\nu-1}\exp(\rho\nu z)dz}{\int_0^1(1-z)^{\nu-1}\exp(\rho\nu z)dz} \\ &= n\frac{\lambda^+}{\mu}\frac{\int_0^1 z(1-z)^{\nu-1}\exp(\rho\nu z)dz}{\int_0^1(1-z)^{\nu-1}\exp(\rho\nu z)dz}. \end{aligned}$$

This implies that

$$\frac{L}{n} = \frac{\lambda^+}{\mu}\frac{\int_0^1 z(1-z)^{\nu-1}\exp(\rho\nu z)dz}{\int_0^1(1-z)^{\nu-1}\exp(\rho\nu z)dz}. \tag{4.40}$$

Let us now obtain asymptotic expansions for the two integrals. First, it is easy to show that

$$\int_0^1 z(1-z)^{\nu-1}\exp(\rho\nu z)dz = \tag{4.41}$$

$$= \int_0^1 z\exp\left(n\frac{\lambda^-}{\mu}(\rho z + \ln(1-z))\right)dz. \tag{4.42}$$

Because the maximum of $\rho z + \ln(1-z)$, given by $\left(1-\frac{1}{\rho}\right)$, is in the interior of [0,1], we can apply Lemma 3 to obtain

$$\int_0^1 z(1-z)^{\nu-1}\exp(\rho\nu z)dz \sim \left(1-\frac{1}{\rho}\right)\exp\left[n\frac{\lambda^-}{\mu}(\rho-1-\ln(\rho))\right]\left[\frac{2\pi}{n\frac{\lambda^-}{\mu}\rho^2}\right]^{\frac{1}{2}}. \tag{4.43}$$

Similarly, we can obtain the following asymptotic expansion :

$$\int_0^1 (1-z)^{\nu-1}\exp(\rho\nu z)dz \sim \exp\left[n\frac{\lambda^-}{\mu}(\rho-1-\ln(\rho))\right]\left[\frac{2\pi}{n\frac{\lambda^-}{\mu}\rho^2}\right]^{\frac{1}{2}}. \tag{4.44}$$

Combining equations (4.43) and (4.44) we conclude that

$$\frac{L}{n} \sim \frac{\lambda^+ - \lambda^-}{\mu}.$$

The remaining expansions are derived in a similar way. ■

Expressions (4.34)-(4.39) are particularly insightful. Let us start with the expressions for the mean queue time and the mean queue length. These expressions are equivalent to the expressions for the following hypothetical queueing system: Consider an M/G/∞ queue with arrival rate $n\lambda^+$, and service rate μ with probability $\left(1 - \frac{\lambda^-}{\lambda^+}\right)$ and ∞ otherwise. In this queuing system customers are divided into those that are serviced by the finite rate exponential servers and those that are serviced by the infinite rate servers. Those customers that are serviced by the infinite rate server correspond to the customers that are serviced before time-out in the original queueing system.

Next, consider the expression for τ. This expression states that as $n \to \infty$, i.e. as the system becomes more congested, the fraction of customers that are serviced before time-out is equal to $\frac{\lambda^-}{\lambda^+}$. This statement is more insightful when made with reference to the underlying kidney transplant problem. In that case, the asymptotic expansion implies that the fraction of patients that receive transplantation is the same as the ratio of the supply over the demand. Therefore, in a highly congested system no kidneys are wasted. This intuition is further verified by expression (4.38) which shows that the probability that the queueing system is empty is exponentially small. Thus, the probability that an organ will become available while the waiting list is empty is also exponentially small.

The expression for the Laplace transform of the time until service is also insightful: it implies that, asymptotically, this random variable is equal to $\frac{1}{\mu}\ln(\rho)$ time units with probability 1; a similar results can be derived from the analysis of the fluid analogue of this queueing system. Finally, consider equation (4.39). This states that the death rate equals the excess in demand. It is worth emphasizing that combining equations (4.39) and (4.36) we obtain the following conservation result: a fraction $\frac{\lambda^-}{\lambda^+}$ of patients receive transplantation and the remaining $\left(1 - \frac{\lambda^-}{\lambda^+}\right)$ withdraw from the waiting list before they

receive transplantation.

Equations (4.34)-(4.39) provide new insights about the performance of the waiting list queueing system. The results presented here are, to the best of our knowledge, new in the queueing literature. The only exceptions are equations (4.1), (4.2) and (4.34) which were first presented in Coffman, Puhalski, Reiman and Wright (1994).

4.1.3 The Multiple Class System:

Thus far we have focused on the analysis of the single class queueing system with reneging. Although this system provides some useful insights, it is a very primitive model of the kidney transplant waiting list. A more realistic model is obtained when we allow several classes of customers and several servers.

Assume that there are J classes of customers and J classes of patients. Each class of customer gives the customer's (patient's) tissue type, and each server corresponds to a different organ tissue type. Let $\lambda_j^+ : j = 1, \ldots, J$ denote the arrival rate for class j customers and $\lambda_j^- : j = 1, \ldots, J$ denote the service rates at server j; equivalently, λ_j^- is the arrival rate for class j organs. Let $\mu_j : j = 1, \ldots, J$ denote the reneging rate of class j customers, and define the dimensionless quantities $\rho_j = \frac{\lambda_j^+}{\lambda_j^-}$ and $\nu_j = \frac{\lambda_j^-}{\mu_j}$. For brevity of notation define the total arrival rate $\lambda^- = \lambda_1^- + \ldots + \lambda_J^-$ and the total service rate $\lambda^+ = \lambda_1^+ + \ldots + \lambda_J^+$. Without loss of generality assume that $\rho_J \geq \rho_{j-1} \geq \ldots \geq \rho_1$. Furthermore, assume that the parameters lie in the heavy traffic regime $\rho_1 > 1$ and $\nu_j \gg 1$.

Our objective here is to understand how the performance of the queueing system is affected by the choice of the service discipline; service disciplines correspond to organ allocation policies in the context of the kidney transplant waiting list. We study three service disciplines: First we analyze a very simple randomized service discipline. Under this discipline, the queuing system decomposes into J single class $M/M/1$ queues with reneging that are simple to analyze. Next, we consider the Best Matching Discipline. This is a head of the line (HOL) priority discipline where each server prioratizes the

customers based on a different order. Under this discipline, server j gives HOL priority to class j customers; i.e. organs are allocated to patients of the same tissue type first. We will argue that in the heavy traffic regime, this queueing system also decomposes into J independent single class $M/M/1$ queues which are simple to analyze. Finally, we will briefly discuss the performance of a policy that *imitates* the First Come First Served policy (FCFS), this policy will be referred to as "FCFT" - the quotations indicate that the policy is not a pure FCFT policy but an imitation of the FCFT..

Randomized Service Discipline:

Under the randomized service discipline the service effort of each server is shared between all customer classes as follows: A proportion γ_{kj} of server k's effort is directed into class k customers and is used to serve the customer that has been in the system for the longest time. It is assumed that $\sum_{j=1}^{J} \gamma_{kj} = 1$ and that if a class of customers is empty, the service effort directed into that class is lost. Let the service rate for class j customers be $\lambda_j^-(\gamma) = \sum_{k=1}^{J} \lambda_k^- \gamma_{kj}$ and assume that $\lambda_j^-(\gamma) < \lambda^+$ for all classes j.

It is easy to show that under the randomized service discipline the queueing system decomposes into J independent single class $M/M/1$ queues with reneging. The arrival rate into queue j is λ_j^+, the service rate is $\lambda_j^-(\gamma)$ and the reneging rate is μ_j. It follows from this decomposition result that the key performance characteristics of the queueing system can be extracted directly from Theorem 2.

Theorem 4 *Let $L_j : j = 1, \ldots, J$ denote the mean queue length for customer class j; $W_j : j = 1, \ldots, J$, the mean queuing time; $W_j|T : j = 1, \ldots, J$ the mean queueing time until reneging; $\tau_j : j = 1, \ldots, J$ the fraction of customers that are serviced before time-out; and $\sigma_j : j = 1, ..., J$ the observed steady state reneging rate for class j customers. Let the arrival and reneging rate for class $j = 1, ..., J$ customers be, respectively, $n\lambda_j^+$, and μ_j, and the service rate for server $j = 1, ...J$ be $n\lambda_j^-$. Then the following asymptotics hold as $n \to \infty$:*

$$\frac{L_j}{n} \sim \frac{\lambda_j^+ - \lambda_j^-(\gamma)}{\mu_j}, \tag{4.45}$$

$$W_j \sim \frac{1}{\lambda_j^+} \frac{\lambda_j^+ - \lambda_j^-(\gamma)}{\mu_j}, \tag{4.46}$$

$$W_j|T \sim \frac{1}{\mu_j} \ln\left(\frac{\lambda_j^+}{\lambda_j^-(\gamma)}\right), \tag{4.47}$$

$$\tau_j \sim \frac{\lambda_j^-(\gamma)}{\lambda_j^+}, \tag{4.48}$$

$$\frac{\sigma_j}{n} \sim \lambda_j^+ - \lambda_j^-(\gamma). \tag{4.49}$$

Proof. The proof of this theorem follows directly from the decomposition result and Theorem 2. ■

The results from Theorem 4 provide useful insights that clarify the relationship between organ allocation policies and the performance of the kidney transplant waiting list. In particular, our analysis shows that for each class of patients, the mean queueing time, the mean time until transplantation, the fraction of performed transplantations, and the observed reneging rates are affected by three factors:

- Candidate arrival rate.
- Candidate death rate.
- Allocation policy.

In the remainder of this paper, we will try to understand the extent to which the allocation policy affects the performance of the queueing system and its impact on equity, or fairness. To give a preliminary sense of our results let us investigate the following statement that summarizes the desirable properties of a fair policy: a policy is perfectly fair if the probability of receiving transplantation and the mean waiting time until transplantation are independent of the class of a patient. . Equations (4.47) and (4.48) show that if different classes of patients have different reneging rates μ_j, then it may not be possible to design an allocation policy that satisfies this statement.

Best Matching Policy: We will now utilize the results from Theorem 4 to analyze the performance of the Best Matching Policy. Our analysis assumes that the queueing system decomposes into J independent queues, one for each class of customers. Although it is not our purpose to rigorously justify this decomposition, we provide an intuitive argument about the plausibility of the decomposition result. Clearly, under the Best Matching Policy the J customer classes interact when one of the customer classes is empty. In that case, the server that is dedicate to that customer class will divert its effort to some other class of lower priority. Therefore, in reality the J customer classes cannot be viewed independently, and the extend of their interdependency depends on the amount of time each server spends serving low priority customers. However, our results from Theorem 2 suggest that in the heavy traffic regime, the queuing system is almost never empty. This implies that, in heavy traffic, the servers are always busy serving customers of the higher priority and the queueing system decomposes into J independent M/M/1 queue is with reneging. There is one queue for each server, and every server serves its highest priority class. However, this argument fails if $\lambda_j^+ < \lambda_j^-$ for some j.

The previous discussion motivates the following asymptotic approximations for the performance of the queueing system under the best matching discipline:

$$\frac{L_j}{n} \sim \frac{\lambda_j^+ - \lambda_j^-}{\mu_j} \tag{4.50}$$

$$W_j \sim \frac{1}{\lambda_j^+} \frac{\lambda_j^+ - \lambda_j^-}{\mu_j} \tag{4.51}$$

$$W_j | T \sim \frac{1}{\mu_j} \ln\left(\frac{\lambda_j^+}{\lambda_j^-}\right) \tag{4.52}$$

$$\tau_j \sim \frac{\lambda_j^-}{\lambda_j^+}. \tag{4.53}$$

$$\frac{\sigma_j}{n} \sim \lambda_j^+ - \lambda_j^- \tag{4.54}$$

"First Come First Served Discipline": Finally, let us analyze the "FCFS" discipline. Our approach is as follows: Rather than directly study the FCFS discipline, we analyze a "surrogate" discipline that imitates that FCFS policy. The surrogate policy is a randomized service discipline which adopts service effort fractions γ_{jk} that force the mean queueing time until transplantation to be the same for all customer classes. Intuitively, this discipline approximates the FCFS discipline because under the FCFS discipline the mean waiting time until transplantation should be the same for every customer class. This implies that the performance of the queueing system under this randomized discipline should approximate the performance of the same system under the FCFS discipline. The main result is given in the following Theorem:

Theorem 5 *Let $L_j : j = 1, \ldots, J$ denote the mean queue length for customer class j; $W_j : j = 1, \ldots, J$, the mean queuing time; $W_j|T : j = 1, \ldots, J$ the conditional mean queueing time given that service is completed; $\tau_j : j = 1, \ldots, J$ the fraction of customers that are serviced before time-out; and $\sigma_j : j = 1, ..., J$ be the mean reneging rate for class j customers. Let the arrival and reneging rates class $j = 1, ..., J$ customers be $n\lambda_j^+$, and μ_j, respectively, and the service rate for server $j = 1, ...J$ be $n\lambda_j^-$. Consider the "FCFS" service discipline with service rates*

$$\lambda_j^-(\gamma^*) = \lambda^+ e^{-k^*\mu_j}$$

where k^* is the unique nonnegative root of the nonlinear equation

$$\sum_{j=1}^{J} e^{-k\mu_j} = \frac{\lambda^-}{\lambda^+}.$$

The following asymptotics characterize the performance of this discipline as $n \to \infty$,

$$\frac{L_j}{n} \sim \frac{\lambda_j^+ \left(1 - e^{-k\mu_j}\right)}{\mu_j}, \tag{4.55}$$

$$W_j \sim \frac{\left(1 - e^{-k\mu_j}\right)}{\mu_j}, \tag{4.56}$$

$$W_j|T \sim k, \tag{4.57}$$

$$\tau_j \sim e^{-k\mu_j}, \tag{4.58}$$

$$\frac{\sigma_j}{n} \sim \lambda_j^+ \left(1 - e^{-k\mu_j}\right).$$

Proof. The proof of this Theorem is straightforward and is omitted. The main insight here is that the service rates are chosen such that the mean queueing time until service is constant.

■

Under the "FCFS" discipline, different classes of patients wait, on average, the same amount of time before they receive transplantation. Therefore, this discipline provides the "gold standard" for equity. However, equation (4.58) imply that the probability that a patient will receive transplantation is inversely proportional to the death rate. Patients with the lowest death rates, have the highest probability of receiving transplantation. This implies that although the "FCFS" policy does not generate inequity in the waiting time, it implicitly favors healthier patients by allocating them a higher fraction of the available organs.

4.2 Conservation Laws for the Reneging Model:

So far we have analyzed both a single class and a multiple class queueing system with reneging and we have derived expressions for the performance of these systems. We have also shown that although the "FCFS" reduces variations of the waiting time until transplantation between different classes of patinets, it does not allocates organs uniformly between the different classes. Rather, classes with low death rates experience preferential treatment. In this section, we will attempt to sharpen these insights even further by showing that the multiple class reneging system satisfies simple conservation laws.

In most physical systems we cannot get something for nothing. For example, in

priority queueing systems certain customers may receive preferential treatment at the expense of others. In the kidney transplant waiting list certain patients may experience shorter delays at the expense of other patients. In general, a system that gives preferential treatment to certain customer (patient) classes does so at the expense of others and induces inequities. This intuitive statements can be made more precise using conservation laws. For example, the multiclass M/G/1 queuing system satisfies a conservation law that states that the weighted sum of the weighting times can never change. Similarly, the multiclass M/G/c system satisfies a conservation law that states that the total unfinished work in the system is invariant; see Federgruen and Groenvelt (1988). These laws are particularly useful because they clarify the nature of trade off that is present in priority queueing systems, and help the controllers of the queueing systems identify desirable priority policies.

The conservation laws presented here are concerned with the performance vector $W = (W_1, \ldots, W_J)$ of steady-state mean queueing times for the J customer classes of the reneging model of section 4.1.3. These laws state that the average queueing times and average queue lengths vary in a very precise manner and are restricted to lie on a hyperplane. Moreover, the conservation laws characterize the achievable performance region for the queueing system and can be used to devise priority policies that minimize certain objectives.

We start with some notation. Let u denote an arbitrary service discipline, and let $W^u = (W_1^u, \ldots, W_J^u)$ denote the steady-state expected waiting time for classes $j = 1, \ldots, J$ under the policy u. Let $J = \{1, \ldots, J\}$ (in this section J will denote the collection of all classes, not the number of patient classes) and for every subset $E \subseteq J$ define

$$\lambda_E^+ = \sum_{j \in E} \lambda_j^+. \tag{4.59}$$

Furthermore, let $b^u(E)$ denote the long-run average service effort allocated to customers in set E under the service discipline u, and let $b(E) = \max b^u(E)$. The following theorem

states that the performance vector W^u satisfies a conservation law:

Theorem 2 *The performance vector W^u satisfies the conservation laws*

$$\sum_{i \in E} \lambda_i^+ \mu_i W_i^u \geq \lambda_E^+ - b(E). \tag{4.60}$$

for every $E \subseteq J$.

Proof. The proof is based on the result that, in steady state, the total flow into a subset of customer classes E is equal to the total flow out of E.

Let $L^u = (L_1^u, \ldots, L_J^u)$ be the steady-state expected queue length for classes $j = 1, \ldots, J$ under discipline u. The steady state flow out of E is equal to the steady state service effort plus the observed steady state reneging rate, and the steady state flow into E is λ_E^+. Therefore, in steady state,

$$\sum_{j \in E} \mu_i L_i^u + b^u(E) = \lambda_E^+. \tag{4.61}$$

By Little's law, $L_i^u = \lambda_i^+ W_i^u$. Therefore,

$$\sum_{i \in E} \lambda_i^+ \mu_i W_i^u = \lambda_E^+ - b^u(E). \tag{4.62}$$

The result follows from the definition of $b(E)$. ■

It is worth emphasizing that Theorem 2 does not hold if the reneging process is not exponential; because the reneging process is exponential, the steady state reneging rate for class customers i is $\mu_i L_i^u$. Nevertheless, it is not difficult to see that Theorem 2 holds even if the organ and (or) patient arrival processes are not exponential. In general, the weaker conditions for Theorem 2 are that the reneging process is exponential and the queue length process is ergodic.

Theorem 2 provides very general conservation laws for our queueing system. These laws can be strengthened by obtaining analytical expressions for the set functions $b(E)$.

Although this is not possible in general, asymptotic results from section 4.1 can be used to derive approximations that are exact in the heavy traffic regime.

Lemma 6 *Let the customer arrival rates be $n\lambda_j^+$, the service rates be $n\lambda_j^-$ and the reneging rate be μ_j for $j = 1, ..., J$. Then, the following asymptotics hold as $n \to \infty$*

$$\frac{b(E)}{n} \sim \min(\lambda^-, \lambda_E^+).$$

Proof. To prove this claim, observe that

$$\frac{b(E)}{n} \leq \min(\lambda^-, \lambda_E^+). \tag{4.63}$$

To see this, observe that the total service effort diverted into subset E can neither exceed the capacity of the system, nor the overall demand from subset E.

Now, suppose that $\lambda_E^+ > \lambda^-$ and let u denote the policy that gives head of the line priority to patients of class $j \in E$. The results from subsection 4.1.3 show that $\frac{b^u(E)}{n} \sim \lambda^-$. Therefore, if $\lambda_E^+ > \lambda^-$, then

$$\frac{b(E)}{n} \sim \min(\lambda^-, \lambda_E^+). \tag{4.64}$$

Now suppose that $\lambda_E^+ \leq \lambda^-$. Again, giving head of the line priority to classes $j \in E$ implies that $b(E) \sim \lambda_E^+$. This analysis show that in heavy traffic

$$\frac{b(E)}{n} \sim \min(\lambda^-, \lambda_E^+). \tag{4.65}$$

■

The conservation laws can be strengthened even more if we restrict our attention to *work conserving* policies.

Definition 7 *A service discipline is called work conserving if no server is idle when there is work in the system.*

The main implication from concentrating on work conserving policies is that, in heavy traffic,

$$b^u(J) \sim \lambda^-. \tag{4.66}$$

To prove equation (4.66) recall that in the heavy traffic regime, the probability that the queueing system is empty is exponentially small, see equation (4.38), therefore there is always work in the system and the server never idles. Now, combining equation (4.66) with the conservation laws in Theorem 2 we obtain the following refined conservation laws:

$$\sum_{i \in J} \lambda_i^+ \mu_i W_i^u = \lambda_J^+ - \lambda^-. \tag{4.67}$$

$$\sum_{i \in E} \lambda_i^+ \mu_i W_i^u \geq \lambda_J^+ - \min\{\lambda^-, \lambda_J^+\}. \tag{4.68}$$

Conservation laws (4.67) and (4.68) are particularly insightful. Equation (4.67) states that the weighted average of the waiting times W_i^u is constant and cannot be affected by the allocation policy, no matter how sophisticated or ingenious it is. Moreover, equations (4.67)-(4.68) characterize the achievable region for the performance vector $W = (W_1, \ldots, W_J)$. These conservation laws can also be used to obtain the service disciplines to minimize certain objectives. For example, using results from Shantikumar and Yao (1991) we can show that to minimize the mean queueing time of incoming customers, the servers should be working on the customers with the lower reneging rates.

4.3 Measuring the Immeasurable: Equity Revisited

Our discussion so far has focused on the performance analysis of the simple reneging model. The motivation for this analysis was to better understand the conditions that give rise to inequity and develop simple quantitative measures for equity. We have analyzed three performance measures of the queueing system: mean waiting time, mean waiting time until transplantatiuon, and steady state fraction of patients receiving trans-

plantation. It is expected that in a fair policy, all these quantities should be the same for all customer classes. However, our analysis shows this may not be always possible. Our discussion following Theorem 4 shows that it is not possible to develop an allocation policy that simultaneously forces the mean waiting time until transplantation and the fraction of performed transplantation to be independent of the patient class. This implies that an equitable, or fair, policy should not necessarily force all these three measures to be independent of patient class. Rather, it should force the differences of these measures between classes to be small. This motivates the following equity metrics:

1. Inequity in mean queueing time:

$$E_W = \sum_{j=1}^{J} \sum_{k=1}^{J} (W_j - W_k)^2 . \tag{4.69}$$

2. Inequity in mean queueing time *until* transplantation:

$$E_{WT} = \sum_{j=1}^{J} \sum_{k=1}^{K} (W_j|T - W_k|T)^2 . \tag{4.70}$$

3. Inequity in access to transplantation:

$$E_A = \sum_{j=1}^{J} \sum_{k=1}^{K} (\tau_j - \tau_k)^2 . \tag{4.71}$$

The intuition behind these three metrics is that they measure the between classes difference for the three performance measures: mean queueing time, mean queueing time until transplantation, and mean fraction of performed transplantations. Unfortunately, with the only exception of the queueing time inequity measure, working with these three measures is not easy. However, the three measures are interrelated, and when $\mu_j = \mu$ all three measures are minimized by the same randomized allocation policy: $\lambda_j^-(\gamma) = \frac{\lambda^-}{\lambda^+}\lambda_j^+$. When the reneging rates are different between classes, the three measures are minimized by different policies, but the policies are the same to a first order approximation.

To justify this last statement, we consider the following problem: Allocate incoming organs so as to minimize the mean queueing time inequity:

$$\sum_{j\in J}\sum_{k\in J}(W_j^u - W_k^u)^2. \tag{4.72}$$

Without loss of generality assume that $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_J$. Furthermore, define the average death rate $\overline{\mu} = \sum_{j\in J}\frac{\lambda_j^-}{\lambda_J}\mu_j$, and $\rho = \frac{\lambda_J^+}{\lambda^-}$ The following proposition provides insights about the optimal allocation policy:

Proposition 7 *In heavy traffic, the total queueing time inequity can be maintained to be zero if and only if*

$$\mu_1 \leq \frac{\rho}{\rho - 1}\overline{\mu}. \tag{4.73}$$

If condition (4.73) holds, then the following static allocation policy minimizes the total queueing time inequity

$$\lambda_j^-(\gamma) = \frac{\lambda_j^+}{\lambda^+}\lambda^- \frac{(1-\rho)\mu_k + \rho\overline{\mu}}{\overline{\mu}}. \tag{4.74}$$

Proof. First, suppose that there exists a policy that maintains the total inequity to be zero. Then, under such a policy $W_i = W$, and the conservation law (4.67) implies that

$$W = \frac{\rho - 1}{\rho}\frac{1}{\overline{\mu}}. \tag{4.75}$$

However, the average waiting time for class i cannot exceed $\frac{1}{\mu_i}$. Therefore,

$$W = W_i \leq \frac{1}{\mu_i}. \tag{4.76}$$

Combining (4.75) with (4.76) implies that

$$\mu_i \leq \frac{\rho}{\rho - 1}\overline{\mu}. \tag{4.77}$$

This gives condition (4.73). To prove the other direction observe that condition (4.73) implies the feasibility of allocation policy (4.74), and that under (4.74), $W_i = \frac{\rho-1}{\rho\overline{\mu}}$. ∎

Let us now provide some insights about the allocation policy (4.74). Although this policy is hard to implement, some useful lessons can be extracted from it. First, consider the completely symmetric case $\mu_i = \mu$. In this case, $u_i = \frac{\lambda_i^+}{\lambda^+}\lambda^-$ and the policy allocates service effort such that all patient classes experience the same utilization rate. This policy not only minimizes the queueing time inequity but it also minimizes the inequity in the queueing time until service completion, as well as the access inequity.

The asymmetric case in more interesting. In this case, the allocation rates deviate from the allocation rates for the symmetric case $\frac{\lambda_i^+}{\lambda^+}\lambda^-$. Specifically, when $\mu_i \geq \overline{\mu}$ the allocation rate is less than $\frac{\lambda_i^+}{\lambda^+}\lambda^-$. In contrast, when $\mu_i < \overline{\mu}$ the allocation rates $u_i > \frac{\lambda_i^+}{\lambda^+}\lambda^-$. Therefore, similarly to the "FCFS" policy, this policy gives preferential treatment to patients with the lowest death rates.

To further clarify the properties of the proposed policy (4.74) in the asymmetric case, consider the profile of the queueing time until transplantation. Equation (4.47) implies that in the heavy traffic regime the mean queueing time until transplantation is

$$W_j|T = \frac{1}{\mu_j}\ln\left(\frac{\lambda_j^+}{u_j}\right) \tag{4.78}$$

$$= \frac{1}{\mu_j}\ln\left(\frac{(1-\rho)\mu_j + \rho\overline{\mu}}{\rho\overline{\mu}}\right). \tag{4.79}$$

Equation (4.79) implies that $W_j|T$ is higher for patients with the higher death rate. However, equation (4.79) also shows that the proposed policy deviates from the "FCFS" policy. To see this, observe that although under FCFS $W_j|T$ is independent of j, under the proposed policy $W_j|T$ does depend on j. Despite that, the deviation of the proposed policy from the FCFS policy is a second order deviation. To show this, take a Taylor series expansion of (4.79):

$$W_j|T = -\frac{1}{\mu_j}\ln\left(1 - \frac{(\rho-1)\mu_j}{\rho\overline{\mu}}\right) \tag{4.80}$$

$$= \frac{\rho-1}{\rho\mu} + o(\frac{(\rho-1)\mu_j}{\rho\mu}). \tag{4.81}$$

This implies that the differences between the $W_j|T$ are of order $o(\frac{(\rho-1)\mu_j}{\rho\mu})$.

In conclusion, we have shown that the policy that minimizes the queueing time inequity imitates the "FCFS" policy. Although, this policy deviates from the"FCFS" policy, the deviations are second order. Therefore the queueing time inequity not only captures inequity in the total queueing time in the system, it also captures the first order effects in the inequity in the queueing time until transplantation. Consequently, it is expected that a policy that promotes queuing time inequity, it will also promote equity in the queueing time until transplantation. In the next chapter, we will use the queueing time inequity as a tool for balancing the trade-off between equity and medical benefit. Although the use of this metric is partly justified because it makes the mathematical analysis tractable, our previous discussion has demonstrated that this measure captures some of the hidden dimensions of equity, such as equity in queueing time until transplantation.

Chapter 5

The Fluid Population Model:

Our discussion, so far, has focused on modeling the kidney transplant waiting list and developing measures for equity. We now turn to the second objective of our investigation which is to develop a model that integrates the kidney transplant waiting list with the organ failure process and captures the evolution of the population of kidney transplant candidates. This Chapter is organized as follows: Section 5.1 presents the fluid model and some of its fundamental properties. Sections 5.2-5.4 describe the three routing problems and develop properties of the optimal policies and simple to implement heuristics. Finally, section 5.5 presents concluding remarks.

5.1 Description of the Fluid Model:

The model takes the form of a continuous time fluid model with K classes of patients and J classes of organs. Each patient class gives the patient's age, gender, race, health, tissue type, etc, and each organ class gives the organ's tissue type, the donor's race, gender, age, etc. The state of the system at time t consists of a K dimensional vector $x(t) = (x_1(t), ..., x_K(t))$, which gives the number of patients of each class. The dynamics of the system are as follows: Patients of class k flow into the system at a rate λ_k^+ per unit time. While in the system, class k customers can change classes at a rate $\mu_k x_k$ per

unit time, where x_k is the current level of class k patients. This generates a flow out of class k which is then transferred into other patient classes according to the flow matrix $P^+ = (p_{kl}^+)$, i.e. a fraction p_{kl}^+ of the patients flowing out of class k become patients of class l. Patients in the system can also change classes following organ allocation. Organs of class j enter the system at a rate λ_j^- per unit time and a fraction $v_{jk}(t)$ is *allocated* to the patients of class k that are currently in the system. This forces a flow of patients out of class k at a rate $u_{jk}(t) = \lambda_j^- v_{jk}(t)$. A fraction $p_{kl}^-(j)$ of these patients become patients of class l.

The dynamics of the system can be expressed more compactly using ordinary differential equations:

$$\begin{aligned}\frac{d}{dt}x_k(t) &= \lambda_k^+ - \mu_k x_k(t) + \sum_j \mu_j x_j(t) p_{jk}^+ - \\ &\quad \sum_{j=1}^{J} \lambda_j^- v_{jk}(t) + \sum_{j=1}^{J} u_{jl}(t) p_{lk}^-(j), k = 1, \ldots, K \end{aligned} \tag{5.1}$$

Equation (5.1) is subject to a set of initial conditions

$$x(0) = x_0, \tag{5.2}$$

and state constraints

$$x(t) \geq 0. \tag{5.3}$$

In addition, the organ allocation rates $u_{jk}(t)$ satisfy the following constraints:

$$\sum_{k=1}^{K} u_{jk}(t) \leq \lambda_j^- \tag{5.4}$$

$$0 \leq u_{jk}(t) \qquad 1 \leq j \leq J, 1 \leq k \leq K. \tag{5.5}$$

Constraint (5.4) states that we cannot allocate more organs than are available, and constraint (5.5) states that we cannot "borrow" organs from the patients that currently

populate the system..

The fluid equation (5.1) and the constraint (5.3) have a simple interpretation. The right hand side of (5.1) decomposes into five components. The first component gives the external arrival of patients of class k. The second component gives the natural flow rate of patients out of class k, $\frac{1}{\mu_k}$ is interpreted as the average amount of time a patient spends in class k. The third component gives the flow of patients into class k triggered by the natural flow out of other classes of customers. The fourth component gives the flow out of class k triggered by the allocation of organs. Lastly, the fifth component gives the flow into class k triggered by the allocation of organs into other classes of customers. Clearly, not all five components can be nonzero at the same time. For example, if class k corresponds to a patient in the waiting list, then the fifth component is zero; i.e. patients leave the waiting list after organ allocation. On the other hand, if class k corresponds to a patient with a functioning organ then components one and four are zero; i.e. patients can only join the system in the waiting list and organs are allocated only to patients in the waiting list.

Constraints (5.4)-(5.5), and the differential equations (5.1) can be expressed more compactly using matrix notation. To do that, define the vectors

$$u_j(t) = (u_{j1}(t), \ldots, u_{jK}(t))', \tag{5.6}$$

$$u(t) = (u_1(t)'| \ldots |u_J(t)')', \tag{5.7}$$

the K dimensional unit vector

$$e = (1, \ldots, 1), \tag{5.8}$$

and the matrices

$$B(j) = -I + P^-(j)', \tag{5.9}$$

$$B = (B(1)| \ldots |B(j)), \tag{5.10}$$

$$M = \operatorname{diag}(\mu_k : k = 1, \ldots, K), \tag{5.11}$$

and $D \in \Re^{J \times JK}$ such that

$$D = \begin{pmatrix} e & 0 & \dots & 0 \\ 0 & e & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}. \tag{5.12}$$

Then, equation (5.1) can be re-expressed as

$$\frac{d}{dt}x(t) = \lambda^+ - (I - P^+)'Mx(t) + Bu(t), \tag{5.13}$$

and the constraints (5.4)-(5.5) are re-expressed as

$$Du(t) \leq \lambda^- \tag{5.14}$$

$$u(t) \geq 0 \tag{5.15}$$

Before, we continue with the problem formulation, we find it convenient to introduce some general assumptions. These assumptions are partly motivated by the problem of kidney allocation and make the analysis simpler. We first state the assumptions and then mention their modeling implications.

Assumption 1 *The flow matrix P^+ has a spectral radius less than one.*

Assumption 2 *The flow matrix P^+ is feedforward.*

Assumption 3 *The death rates for the communicating classes of P^+ are different; i.e. if in the absence of negative arrivals, flow from class j can eventually enter class k, then $\mu_j \neq \mu_k$.*

Assumption 1 is common in the queueing literature and it implies that the fluid network is an open network. This assumption implies that no patient will stay in the system indefinitely, and that the *effective arrival rate* λ, which is defined as follows:

$$\lambda = \lambda^+ + \left(P^+\right)' \lambda, \tag{5.16}$$

exists and is finite. Assumption 2 is a natural assumption in the context of kidney transplantation because it implies that in the absence of organs, patients can only move to classes of worse health. Finally, assumption 3 is a technical assumption that is needed to prove Lemma 2. From a pratical point of view, this assumption does not appear to be restrictive.

Properties of the Fluid Model To better understand the dynamics of the fluid model we now develop four properties that are particularly insightful. The first property gives the asymptotic behavior of the system when a static allocation policy is used. The second property gives sufficient conditions for the non-negativity of the state space. The third property derives two conservation laws for the state of the fluid system. Finally, the last property gives the total arrival rate into the classes of the fluid model.

Property 1 *Consider a static allocation policy $u_{jk}(t) = u_{jk}$ that does not violate the nonnegativity constraints. Under such an allocation policy, the fluid model has a single fixed point*

$$x(t) = M^{-1}\left((I - P^{+})'\right)^{-1}\left(\lambda^{+} + Bu\right), \tag{5.17}$$

which is stable.

Proof. First, observe that (5.17) is indeed a fixed point. Second, notice that all the eigenvalues of $(I - P^{+})'M$ are positive. This is because P^{+} has a spectral radius less than one. This implies that the fixed point is stable. ■

Property 2 *If every time that $x_k(t) = 0$ for some $k \in \{1, \ldots, K\}$,*

$$\lambda_k^{+} - \sum_{j=1}^{J} u_{jk}(t) + \sum_{j=1}^{J}\sum_{l=1}^{K} u_{jl}(t) p_{lk}^{+}(j) \geq 0, \tag{5.18}$$

then $x(t) \geq 0$ for all t.

Proof. The proof follows from the statement that $x(t) \geq 0$ if and only if when $x_k(t) = 0$ then $\frac{d}{dt}x_k(rt) \geq 0$. ■

Property 3 *Let $K = \{1, \ldots, K\}$ denote the set of all patient classes, $WL = \{1, \ldots, K_w\}$ denote the set of patient classes in the waiting list, and $T = \{K_w + 1, \ldots, K\}$ denote the set of patient classes with functioning organ. Let u denote an arbitrary allocation policy, let $L^u = (L_1^u, \ldots, L_K^u)$ denote the steady-state queue lengths for classes $k = 1, \ldots, K$, and b^u denote the steady state organ allocation rate under this policy. For brevity of notation define $p_{k0}^+ = 1 - \sum_{j=1}^K p_{kj}^+$ and $p_{k,T}^+ = \sum_{j \in T} p_{kj}^+$; $p_{k,T}$ is the probability that a customer will remain in T after a transition out of k, and P_{k0} is the probability that a customer will leave the system immediately after a transition out of k. If for every $s \in WL$ and $j \in J$,*

$$\sum_{k \in T} p_{sk}^-(j) = 1, \tag{5.19}$$

then, the performance vector L^u satisfies the following conservation laws:

$$\sum_{k \in K} \mu_k p_{k0}^+ L_k^u = \lambda_K^+ \tag{5.20}$$

$$\sum_{k \in T} \mu_k (1 - p_{k,T}^+) L_k^u = b^u \tag{5.21}$$

Proof: The proof is based on the fundamental idea that, in steady state, the total flow into a set of patient classes is equal to the total flow out of this set.

First, consider the set K. In steady state, the total flow into K is λ_K^+, and the total flow out of K is equal to the total death rate:

$$\sum_{k \in K} \mu_k p_{k0}^+ L_k^u.$$

Therefore, in steady state

$$\sum_{k \in K} \mu_k p_{k0}^+ L_k^u = \lambda_K^+. \tag{5.22}$$

Similarly, we can prove equation (5.21). To do that, it suffices to observe that the total flow into set T is b^u and that the total flow out of set T is

$$\sum_{k \in T} \mu_k (1 - p^+_{k,T}) L^u_k. \tag{5.23}$$

■

The conservation law (5.21) can be further refined if we concentrate on *organ conserving* policies; a policy is called organ conserving if it always allocates organs when patients are available. Because the total patient arrival rate exceeds the organ arrival rate, it follows that for organ conserving policies

$$b^u \approx \lambda^-.$$

Therefore,

$$\sum_{k \in T} \mu_k (1 - p^+_{k,T}) L^u_k \approx \lambda^-. \tag{5.24}$$

This implies that, no matter how ingenious our allocation policy is, the weighted average of the number of patients with functioning organs is constant. Moreover, using this conservation law we can develop an upper bound for the mean steady state number of quality adjusted life years in the system. This bound is determined by the solution to the following linear program:

$$\begin{aligned} \max \quad & \textstyle\sum_{k \in K} h_k L^u_k \\ & \textstyle\sum_{k \in K} \mu_k p^+_{k0} L^u_k = \lambda^+_K \\ & \textstyle\sum_{k \in T} \mu_k (1 - p^+_{k,T}) L^u_k = \lambda^- \\ & L^u_k \geq 0; \end{aligned} \tag{5.25}$$

where $h_k \in [0, 1]$ is the Quality Adjustment Factor for patients of class k.

Property 4 *Let $\lambda_k(t,u)$ denote the instantaneous arrival rate into class k under allocation policy $u = (u(t))$ and let $\lambda(t,u) = \lambda_k(t,u)$. Then,*

$$\lambda(t,u) = \lambda^+ + \left(P^+\right)' Mx(t) + P^- u(t). \tag{5.26}$$

Furthermore, if as $t \to \infty$, $u(t) \to u$ then

$$\lambda(\infty,u) = \lambda^+ + P^- u + \left(P^+\right)' \left((I - P^+)'\right)^{-1} (\lambda^+ + Bu). \tag{5.27}$$

For brevity of notation, use $\lambda(u) = \lambda(\infty,u)$.

Proof. The proof is elementary and is omitted.

5.2 Formulation of the Control problem: Maximizing Clinical Efficiency

We now show how the fluid model of Section 5.1 can be combined with a mathematical objective function to give an optimal control problem. The solution to this problem will give the allocation policy that maximizes clinical efficiency.

The objective is

$$\int_0^T \sum_{k=1}^K h_k x_k(t) dt, \tag{5.28}$$

where $h_k \in [0,1]$ is the Quality of Life Adjustment Factor for class k patients; i.e. one life year of a class k patient is equivalent to h_k quality adjusted life years. For brevity of notation define the column vector $h = (h_k)$.

The optimal control problem is to find the allocation rates $u_{jk}(t); t \in [0,T]$ to

$$\text{maximize} \quad \int_0^T h'x(t)dt \tag{5.29}$$

$$\text{subject to} \quad \frac{d}{dt}x(t) = \lambda^+ - (I - P^+)'Mx(t) + Bu(t) \tag{5.30}$$

$$Du(t) \leq \lambda^{-} \tag{5.31}$$

$$x(t), u(t) \geq 0, \tag{5.32}$$

$$x(0) = x_0. \tag{5.33}$$

This problem is a linear control problem with state and control variable constraints. This problem can also be viewed as an infinite dimensional linear programming problem. Although efficient algorithms have been developed recently to solve this problem, see Luo (1995), closed form solutions to this problem are not available as of yet. However, it is possible to develop an efficient myopic policy if we relax the state constraint $x(t) \geq 0$. To do that, we must first formulate the maximum principle.

This is done next. We first formulate the maximum principle and show that the optimal allocation policy is a dynamic index policy. We then relax the nonnegativity constraints and solve the adjoint equations. This gives a dynamic index policy that solves the relaxed version of the problem and provides a simple heuristic for the original problem.

5.2.1 The Pontryagin Maximum Principle

In this subsection, we use Pontryagin's maximum principle to show that the optimal solution to the control problem (5.29)-(5.33) is a dynamic index policy. In the next subsection, we use this insight to develop a simple heuristic that solves the relaxed version of the control problem. For a nice review of the maximum principle and applications to operations research see Sethi and Thompson, (1981).

We start with the notation. Let $x^*(t)$ denote the optimal trajectory, $u^*(t)$ the optimal allocation policy and $V^*(x,t)$ the optimal reward from time t to time T when the state of the system at time t is x. The current value Hamiltonian function is defined as,

$$H(x,u,\pi) = \beta h'x - +\pi'(\lambda^{+} - (I - P^{+})'Mx + Bu).$$

The maximum principle states that there exist adjoint variables $\pi(t) = (\pi_k(t) : 1 \leq k \leq K)$ and Lagrange multipliers $\eta(t) = (\eta_k(t) : 1 \leq k \leq K)$ that solve the adjoint problem,

$$\frac{d}{dt}\pi(t) = -\beta h + M'(I - P^+)\pi(t) - \eta(t), \tag{5.34}$$

$$\pi(T) = 0, \tag{5.35}$$

and satisfy the complementarity conditions,

$$\eta(t)'x(t) = 0 \tag{5.36}$$

$$\eta(t) \geq 0. \tag{5.37}$$

The optimal allocation policy $u^*(t)$ solves,

$$\begin{aligned} \max \quad & H(x^*(t), u(t), \pi(t)) && (5.38) \\ \text{subject to} \quad & Du(t) \leq \lambda^- && (5.39) \\ & u(t) \geq 0, && (5.40) \end{aligned}$$

and

$$H(x^*(t), u^*(t), \pi(t)) = 0. \tag{5.41}$$

Applying the maximum principle, we obtain

$$H(x, u, \pi) = \beta h'x - \pi'(\lambda^+ - (I - P^+)'Mx) + \sum_{j=1}^{J}\sum_{k=1}^{K} G_{jk}(t)u_{jk}(t), \tag{5.42}$$

where

$$G_{jk}(t) = \sum_{l=1}^{K} p_{kl}^-(j)\pi_l(t) - \pi_k(t). \tag{5.43}$$

This implies that the optimal policy is characterized by the dynamic indices $G_{jk}(t)$: allocate organs of class j to the patients with the highest positive index $G_{jk}(t)$. The

index $G_{jk}(t)$ gives the marginal increase in the QALYs from allocating an organ of class j, to the patient of class k.

A striking property of the maximum principle is that it provides quantities with important physical interpretation. Here we provide such an interpretation for the adjoint variables $\pi(t)$, and Lagrange multipliers $\eta(t)$. A standard result from optimal control theory states that the adjoint variables are related to the optimal reward to go function $V^*(x,t)$ through

$$\pi_k(t) = \frac{\partial}{\partial x_k} V^*(x,t)\,. \tag{5.44}$$

This implies that $\pi(t)$ is the shadow price vector for the current state $x(t)$; i.e if $x(t)$ changes by Δx, then the optimal reward will change by $\pi(t)'\Delta x$. The interpretation of the Lagrangian multipliers $\eta(t)$ is slightly more involved. In theory, $\eta_k(t)$ is the shadow price for the non-negativity constraint $x_k(t) \geq 0$. This implies that if the non negativity constraint is relaxed to be $x_k(t) \geq -d$, where $d > 0$, then the optimal reward changes by $\eta_k(t)d$. Because the relaxed constraint states that it is possible to store kidneys for future transplantations to class k patients, it follows that $\eta_k(t)$ gives the marginal benefit from storing kidneys. Although storing kidneys is not technically feasibly today, we cannot exclude the possibility that the technology will become available in the future. In this case, $\eta_k(t)$ will provide estimates for the benefit from the implementation of such technology.

We conclude this section by outlining a strategy for developing efficient heuristics. Equation (5.43) suggests that to develop efficient allocation policies one can either attempt to solve the adjoint equations explicitly, or derive simple approximations. Solving equations (5.34)-(5.35) is a non trivial task because it requires knowledge of the Lagrangian multipliers $\eta(t)$. However, when the nonnegativity state constraint are relaxed, the Lagrangian multipliers become zero and the adjoint equations become simple linear ordinary differential equations which are easy to solve. This is done in the next subsection, where we use this approach to develop a class of efficient heuristics.

5.2.2 Solving the Relaxed problem

If we relax the nonnegativity constraint $x(t)$, then the adjoint equations become straightforward linear differential equations:

$$\frac{d}{dt}\pi(t) = -h + M'(I - P^+)\pi(t), \tag{5.45}$$

$$\pi(T) = 0. \tag{5.46}$$

Elementary results from the theory of linear ordinary differential equations show that

$$\pi(t) = (I - e^{-M'(I-P^+)(T-t)}) \left(M'(I - P^+)\right)^{-1} h, \tag{5.47}$$

where e^B is the exponential operator $\sum_{k=1}^{\infty} \frac{B^k}{k!}$ (Hirsch and Smale, chapter 5.4). This suggests the following policy: Calculate the dynamic indices $G_{jk}(t)$ by substituting (5.47) into (5.43). Then allocate organs of class j to the patient of class k that has the highest index $G_{jk}(t)$. If there are no patients of class k in the system, then allocate the organs to the patients with the second highest non negative index; this policy will be referred to as the efficiency policy.

At this point it is worthwhile to provide some insights about the efficiency policy and expression (5.47). To do that consider the evolution of a patient of class k that is populating a hypothetical system with no external flow of either organs or candidates. The state of this system at time t is given by the solution to the ordinary differential equations:

$$\frac{d}{dt}x(t) = -(I - P^+)'Mx(t) \tag{5.48}$$

$$x(0) = e_k. \tag{5.49}$$

It is an easy exercise to show that (5.47) is related to the solution $x(t)$ of (5.48)-(5.49)

through

$$\pi_k(t) = \int_0^{T-t} h' x(\tau) d\tau; \tag{5.50}$$

here $x(t)$ solves (5.48)-(5.49) Therefore, $\pi_k(t)$ gives the total number of quality of life years accumulated by a patient of class k over a time period of $T - t$. The index $G_{jk}(t)$ gives the increase in quality adjusted life years that is caused by allocating an organ of class j to a patient of class k and is a myopic policy.

5.3 Reformulation of the Control Problem I: Promoting Equity

The efficiency policy of section 5.2 suffers from a major weakness that makes it unacceptable both from a practical and an ethical point of view. Specifically, the policy allocates incoming organs only to those patients that are expected to gain the highest number of quality adjusted life years from transplantation. The remaining patients are refused access to transplantation, even though transplantation can provably improve their life expectancy and quality of life. To overcome this problem, we reformulate the control problem to incorporate a constraint on the fraction of organs allocated to each patient class. By doing that, we can ensure that each patient class is guaranteed a minimum number of transplants.

As an illustration of this approach, suppose that $\mathcal{S}$ is a collection of patient classes that have restricted access to transplantation under the efficiency policy. To counteract this problem, introduce the following constraint into the control problem (5.29)-(5.33),

$$\frac{\int_0^T \sum_{j=1}^J \sum_{k \in \mathcal{S}} u_{jk}(t) dt}{\sum_{k \in \mathcal{S}} \lambda_k^+} \geq C \tag{5.51}$$

This constraint requires that the fraction of incoming patients of group $\mathcal{S}$ that receive transplantation is no less than C. The resulting control problem becomes: Obtain the allocation rates $u_{jk}(t)$ to

$$
\begin{aligned}
\max \quad & \int_0^T h' x(t) dt \\
\text{subject to} \quad & \tfrac{d}{dt} x(t) = \lambda^+ - (I - P^+) M x(t) + B u(t) \\
& \frac{\int_0^T \sum_{j=1}^J \sum_{k \in \mathcal{S}} u_{jk}(t) dt}{\sum_{k \in \mathcal{S}} \lambda_k^+} \geq C \\
& D u(t) \leq \lambda^- \\
& x(t), u(t) \geq 0 \\
& x(0) = x_0.
\end{aligned}
\tag{5.52}
$$

Instead of solving this problem, we solve its Lagrangian relaxation,

$$
\begin{aligned}
\max \quad & \int_0^T h' x(t) dt + \delta \frac{\int_0^T \sum_{j=1}^J \sum_{k \in \mathcal{S}} u_{jk}(t) dt}{\sum_{k \in \mathcal{S}} \lambda_k^+} \\
\text{subject to} \quad & \tfrac{d}{dt} x(t) = \lambda^+ - (I - P^+) M x(t) + B u(t) \\
& D u(t) \leq \lambda^- \\
& x(t), u(t) \geq 0 \\
& x(0) = x_0.
\end{aligned}
\tag{5.53}
$$

The parameter δ can be interpreted as a subsidy for allocating organs to patients in the group $\mathcal{S}$; δ is also the shadow price for the access constraint (5.51).

A straightforward application of the maximum principle shows that the optimal allocation policy is, again, characterized by a dynamic index $\tilde{G}_{jk}(t)$. This index is given by

$$
\tilde{G}_{jk}(t) = \sum_{l=1}^{K} p_{kl}^-(j) \pi_l(t) - \pi_k(t) + \frac{\delta}{\sum_{k \in \mathcal{S}} \lambda_k^+} I(k \in \mathcal{S}); \tag{5.54}
$$

the adjoint variables $\pi(t)$ are derived from the adjoint equation and can be approximated by equation (5.47). This implies that the policy adjusts the efficiency policy by introducing the equity based subsidy $\frac{\delta}{\sum_{k \in \mathcal{S}} \lambda_k^+} I(k \in \mathcal{S})$; $I(.)$ is the indicator function. The policy that adopts the index in equation (5.54), together with the approximation (5.47), will be referred to as the *subsidized index policy.*

5.4 Reformulation of the Control Problem II: Promoting Equity

We have seen in the last section that by introducing constraints in the optimal control formulation of the kidney allocation problem it is possible to develop allocation policies that enhance medical utility and promote equity. An alternative approach is to adopt a bi-criteria objective that utilizes a weighted combination of medical benefit and equity. For example, using the queueing time inequity measure of Section 4.3 we can formulate the objective function

$$\int_0^T \left(\beta \sum_{k=1}^{K} h_k x_k(t) - (1-\beta) \sum_{k=1}^{K} \sum_{l=1}^{K} w_{kl} \left(\frac{x_k(t)}{\lambda_k(u;t)} - \frac{x_l(t)}{\lambda_l(u;t)} \right)^2 \right) dt.$$

The first part of the objective measures the total number of quality adjusted life years in the system, and the second part measures the total difference of the mean queueing time between different patient classes; by Little's law, $\frac{x_k(t)}{\lambda_k(u;t)}$ approximates the mean queueing time for class k patients at time t and $\sum_{k=1}^{K} \sum_{l=1}^{K} w_{kl} \left(\frac{x_k(t)}{\lambda_k(u;t)} - \frac{x_l(t)}{\lambda_l(u;t)} \right)^2$ is an approximation for the queueing time inequity measure. The weight β takes values between 0 and 1 and traces the trade-off between clinical efficiency and equity.

Using the bi-criteria objective, we obtain the following control problem Obtain the allocation policy $u = (u_{jk}(t))$ to

$$\begin{aligned}
\text{maximize} \quad & \int_0^T (\beta h' x(t) - (1-\beta) x(t)' Q(u,t) x(t))\, dt && (5.55)\\
\text{subjhect to} \quad & \frac{d}{dt} x(t) = \lambda^+ - (I - P^+)' M x(t) + B u(t) && (5.56)\\
& D u(t) \leq \lambda^- && (5.57)\\
& x(t), u(t) \geq 0, && (5.58)\\
& x(0) = x_0. && (5.59)
\end{aligned}$$

where $Q(u,t)$ is a $K \times K$ matrix with $[Q(u,t)]_{kl} = (-1)^{I(k \neq l)} \sqrt{w_{kl}} \frac{1}{\lambda_k(u;t)} \frac{1}{\lambda_l(u;t)}$.

Unfortunately, analyzing the control problem (5.55)-(5.59) appears to be an arduous task, primarily because the objective is nonconvex in u. To overcome this problem, we develop a simple approximation for $Q(u,t)$ and analyze the resulting optimal control problem. To develop the approximation for $Q(u,t)$ we first solve the following nonlinear optimization problem to obtain the steady state optimal allocation policy for (5.55)-(5.59)

$$\begin{aligned}
\text{maximize} \quad & (\beta h'x - (1-\beta)x'Q(u)x)\, d && (5.60)\\
\text{subject to} \quad & 0 = \lambda^+ - (I - P^+)'Mx + Bu && (5.61)\\
& Du \leq \lambda^- && (5.62)\\
& x, u \geq 0, && (5.63)\\
& && (5.64)
\end{aligned}$$

where $Q(u)$ is a $K \times K$ matrix with $[Q(u)]_{kl} = (-1)^{I(k \neq l)} \sqrt{w_{kl}} \frac{1}{\lambda_k(u)} \frac{1}{\lambda_l(u)}$.

The solution to the mathematical programming problem (5.60)-(5.64) gives the optimal steady state allocation policy u^* and the optimal steady-state point x^*. It is worth emphasizing that because the optimal control problem is undiscounted, x^* gives the optimal long run stationary equilibrium point for the infinite horizon version of (5.55)-(5.59) and u^* gives the optimal control when $x(t) = x^*$; see page 86 of Sethi and Thompson, 1981. This implies that if the planning horizon is sufficiently large and the system is operated under the optimal control policy, thenthe optimal trajectory $x^*(t)$ will converge to x^* and the optimal allocation policy will converge to u^*. From this it follows that an appropriate long-run approximation for $Q(u,t)$ is $Q = Q(u^*)$. Using this approximation, we can reformulate the optimal control problem (5.55)-(5.59) as

$$\begin{aligned}
\text{maximize} \quad & \int_0^T (\beta h'x(t) - (1-\beta)x(t)'Qx(t))\, dt && (5.65)\\
\text{subject to} \quad & \frac{d}{dt}x(t) = \lambda^+ - (I - P^+)'Mx(t) + Bu(t) && (5.66)\\
& Du(t) \leq \lambda^- && (5.67)\\
& x(t), u(t) \geq 0, && (5.68)
\end{aligned}$$

$$x(0) = x_0. \tag{5.69}$$

This is a Linear Quadratic Regulator with state and control constraints. In the absence of the constraints, the control problem is very well studied and the optimal control is a linear control that is available in closed form. However, in the presence of constraints, the problem becomes mathematically very difficult and little is known about the optimal policy.

This control problem can also be viewed as an infinite dimensional quadratic programming problem. Unlike infinite dimensional linear programs which are fairly well studied, infinite dimensional quadratic programs are poorly understood. Therefore, all the results that we will present in this paper are new in the literature.

5.4.1 The Pontryagin Maximum Principle

In this section, we use Pontryagin's maximum principle to show that the optimal policy for the control problem (5.65)-(5.69) is a dynamic index policy. Using this insight, we propose a strategy for developing efficient heuristics.

We start with the notation. Let $x^*(t)$ denote the optimal trajectory, $u^*(t)$ the optimal allocation policy and $V^*(x,t)$ the optimal reward from time t to time T when the state of the system at time t is x. The current value Hamiltonian function is defined as,

$$H(x,u,\pi) = \beta h'x - (1-\beta)x'Qx + \pi'(\lambda^+ - (I-P^+)'Mx + Bu).$$

The maximum principle states that there exist adjoint variables $\pi(t) = (\pi_k(t) : 1 \leq k \leq K)$ and Lagrange multipliers $\eta(t) = (\eta_k(t) : 1 \leq k \leq K)$ that solve the adjoint problem,

$$\frac{d}{dt}\pi(t) = -\beta h + 2(1-\beta)Qx^*(t) + M'(I-P^+)\pi(t) - \eta(t), \tag{5.70}$$

$$\pi(T) = 0, \tag{5.71}$$

and satisfy the complementarity conditions,

$$\eta(t)'x(t) = 0 \tag{5.72}$$

$$\eta(t) \geq 0. \tag{5.73}$$

The optimal allocation policy $u^*(t)$ solves,

$$\begin{aligned} \max \quad & H(x^*(t), u(t), \pi(t)) && (5.74) \\ \text{subject to} \quad & Du(t) \leq \lambda^- && (5.75) \\ & u(t) \geq 0, && (5.76) \end{aligned}$$

and

$$H(x^*(t), u^*(t), \pi(t)) = 0. \tag{5.77}$$

Applying the maximum principle, we obtain

$$H(x, u, \pi) = \beta h'x - (1-\beta)x'Qx + \pi'(\lambda^+ - (I - P^+)'Mx) + \sum_{j=1}^{J}\sum_{k=1}^{K} G_{jk}(t)u_{jk}(t), \tag{5.78}$$

where

$$G_{jk}(t) = \sum_{l=1}^{K} p_{kl}^-(j)\pi_l(t) - \pi_k(t). \tag{5.79}$$

This implies that the optimal policy is characterized by the dynamic indices $G_{jk}(t)$: allocate all organs of class j to the candidates with the highest positive index $G_{jk}(t)$. The index $G_{jk}(t)$ gives the marginal reward from allocating one organ of class j to a candidate of class k.

To complete the description of the allocation policy it remains to solve the adjoint equations and obtain $\pi_k(t)$. However, this does not appear to be feasible because solving the adjoint equations requires knowledge of the optimal trajectory $x^*(t)$. Instead, we use the policy iteration algorithm to approximate the index policy .

5.4.2 Analysis:

To obtain a dynamic policy for the control problem (5.65)-(5.69) we approximate $\pi_k(t)$ using the partial derivative of the reward to go function under an appropriate static policy. This is equivalent to performing one step of the policy iteration algorithm starting from this static policy. This approach was first used by Ott and Krishnan (1985) to develop routing policies in large scale stochastic telecommunication networks, and was used more recently by Wein, Nowak and Zenios (1996) to analyze a non linear control problem for dynamic AIDS drug treatment. The approach consists of three steps: First, obtain the trajectory $x(t)$ under a general static (stationary) allocation policy $u(t) = u$ and compute the reward to go function under the static policy. Second, obtain the static policy that maximizes the long run average reward. Denote the derived policy u^s, and compute the reward to go function $V^s(x,t)$ under u^s; $V^s(x,t)$ gives the reward from time t to time T under the optimal static policy when the state of the system at time t is x. Lastly, compute $\pi_k^s(t) = \frac{\partial}{\partial x_k} V^s(x,t)$ and calculate the dynamic index $G_{jk}(t)$ using $\pi^s(t)$ instead of $\pi(t)$. The computations involved are rather cumbersome but can be streamlined using matrix notation, and assuming that the planning horizon $T \to \infty$; the assumption $T \to \infty$ is not necessarily restrictive because in practise the planning horizon should be at least ten years.

Let us now carry out the three steps of the proposed approach. The reader who is not interested in the mathematical analysis, may wish to skip to subsection 5.4.3.

Static Policy:

To obtain the trajectory under a static policy u we solve the boundary value problem,

$$\frac{d}{d\tau}x(\tau) = \lambda^+ - (I - P^+)'Mx(\tau) + Bu, \tag{5.80}$$

$$x(t) = x. \tag{5.81}$$

Elementary results from the theory of ordinary differential equations show that,

$$x(\tau) = e^{-A(\tau-t)}\left(x - A^{-1}(\lambda^+ + Bu)\right) + A^{-1}(\lambda^+ + Bu), \tag{5.82}$$

where $A = (I - P^+)^T M$. Now, we can obtain the value function,

$$\begin{aligned} V(x,t;u) &= \int_t^T \left[\beta h' x(\tau) - (1-\beta)x(\tau)'Qx(\tau)\right] d\tau = \\ &= \left[\beta h' A^{-1}(\lambda^+ + Bu) - (1-\beta)\left(A^{-1}(\lambda^+ + Bu)\right)' QA^{-1}(\lambda^+ + Bu)\right](T-t) + \\ &\quad \beta\left[h' e^{-A(\tau-t)}\left(-A^{-1}\right)\left(x - A^{-1}(\lambda^+ + Bu)\right)\right]_t^T - \\ &\quad \left[2(1-\beta)\left(A^{-1}(\lambda^+ + Bu)\right)' Qe^{-A(\tau-t)}\left(-A^{-1}\right)\left(x - A^{-1}(\lambda^+ + Bu)\right)\right]_\tau^T - \\ &\quad (1-\beta)\int_t^T \left(x - A^{-1}(\lambda^+ + Bu)\right)' e^{-A'(\tau-t)} Q e^{-A(\tau-t)}\left(x - A^{-1}(\lambda^+ + Bu)\right) d\tau. \end{aligned} \tag{5.83}$$

For brevity of notation define $Q(t;T) = \int_t^T e^{-A'(\tau-t)} Q e^{-A(\tau-t)} d\tau$; later, we will see that we can calculate this integral explicitly. It follows from (5.83) that

$$\begin{aligned} V(x,t;u) &= \left[\beta h' A^{-1}(\lambda^+ + Bu) - (1-\beta)\left(A^{-1}(\lambda^+ + Bu)\right)' QA^{-1}(\lambda^+ + Bu)\right](T-t) \\ &\quad +\beta h'\left(I - e^{-A(T-t)}\right) A^{-1}\left(x - A^{-1}(\lambda^+ + Bu)\right) \\ &\quad -2(1-\beta)\left(A^{-1}(\lambda^+ + Bu)\right)' Q\left(I - e^{-A(T-t)}\right) A^{-1}\left(x - A^{-1}(\lambda^+ + Bu)\right) \\ &\quad -(1-\beta)\left(x - A^{-1}(\lambda^+ + Bu)\right)' Q(t;T)\left(x - A^{-1}(\lambda^+ + Bu)\right). \end{aligned} \tag{5.84}$$

Optimal Static Policy:

The optimal static policy maximizes the long run average cost

$$\lim\inf_{T\to\infty} \frac{1}{T} V(x,t;u).$$

This implies that u^s solves the quadratic programming problem

$$\begin{aligned} \max \quad & \beta h' A^{-1}(\lambda^+ + Bu) - (1-\beta)\left(A^{-1}(\lambda^+ + Bu)\right)' Q A^{-1}(\lambda^+ + Bu) && (5.85) \\ \text{subject to} \quad & Du \leq \lambda^- && (5.86) \\ & A^{-1}(\lambda^+ + Bu^s) \geq 0. && (5.87) \\ & u \geq 0. && (5.88) \end{aligned}$$

Constraint (5.87) states that under the static allocation rates u, the fixed point of (5.1) should be non-negative. The optimization problem (5.85)-(5.87) is a simple quadratic programming problem that can be solved numerically using any standard non linear solver such as MINOS.

Proposed Policy:

To obtain the proposed policy, substitute u^s into (5.84) to obtain $V^s(x,t)$, and differentiate $V^s(x,t)$ with respect to x to get that as $T \to \infty$,

$$\begin{aligned} \left(\frac{\partial}{\partial x} V^s(x,t)\right)' &= \beta h' A^{-1} - 2(1-\beta)\left(A^{-1}(\lambda^+ + Bu^s)\right)' Q A^{-1} \\ & \quad 2(1-\beta)\left(x - A^{-1}(\lambda^+ + Bu^s)\right)' Q(0;\infty). \end{aligned} \tag{5.89}$$

This implies that the proposed policy calculates the dynamic index $G_{jk}(t)$ by substituting

$$\begin{aligned} \pi^s(t)' &= \beta h' A^{-1} - 2(1 - \cdot\beta)\left(A^{-1}(\lambda^+ + Bu^s)\right)' Q A^{-1} && (5.90) \\ & \quad -2(1-\beta)\left(x(t) - A^{-1}(\lambda^+ + Bu^s)\right)' Q(0;\infty). && (5.91) \end{aligned}$$

into (5.43). For brevity of notation define the *proposed index*

$$G^s_{jk}(t) = \sum_{l=1}^{K} p^-_{kl}(j)\pi^s_l(t) - \pi^s_k(t). \tag{5.92}$$

Equations (5.91)-(5.92) give a closed form expression for a novel dynamic policy that balances the trade off between medical benefit and equity.

To complete the description of the policy, it remains to discuss the computation of the matrix $Q(t;T)$:

Lemma 2 *1. Matrix A is diagonalizable. This implies that there exists an invertible matrix V and a diagonal positive definite matrix S such that $AV = VS$, or equivalently, $A = VSV^{-1}$; the columns of V consist of the eigenvectors of A, and the diagonal entries of S are the corresponding eigenvalues (the ith diagonal entry of S is s_i, and the ith column of V is v_i).*

2. Let $\bar{Q}(t;T) \in \Re^{K\times K}$ such that $\bar{Q}_{ij}(t;T) = \frac{v_i' Q v_j}{s_i+s_j}\left(1 - e^{-(s_i+s_j)(T-t)}\right)$. Then,

$$Q(t;T) = \left(V^{-1}\right)' \bar{Q}(t;T) V^{-1}.$$

Proof. The proof relies heavily on the assumptions in Section 5.1. Specifically, A is diagonalizable because P^+ is feedforward, and the communicating classes of P^+ have different death rates. To see this, observe that because P^+ is feedforward and M is diagonal, the matrix $A = (I - P^+)M$ can be expressed in an upper triangular form. Furthermore, if the matrix P^+ consists of more than one communicating classes, then A can be expressed as a block diagonal matrix, with each block being upper triangular. Now, the assumption that the communicating classes of P^+ have different death rates, implies that the diagonal entries of each block are distinct. From this, it follows that each block matrix is diagonalizable and consequently, that A itself is diagonalizable. ■

5.4.3 Interpretation of the Policies:

In this subsection we interpret the proposed policy, illuminating the mechanism through which this policy balances the two conflicting objectives of equity and benefit; the derived

policy will be reffered to as the *dynamic priority* policy. To do that, we find it convenient to decompose $\pi^s(t)$ and $G^s_{jk}(t)$ as follows: Let

$$\pi^0(t) = h'A^{-1} \tag{5.93}$$

$$\pi^1(t) = -2\left(A^{-1}(\lambda^+ + Bu^s)\right)' QA^{-1} - 2\left(x(t) - A^{-1}(\lambda^+ + Bu^s)\right)' Q(0;\infty) \tag{5.94}$$

$$G^0_{jk}(t) = \sum_{l=1}^{K} p^-_{kl}(j)\pi^0_l(t) - \pi^0_k(t) \tag{5.95}$$

$$G^1_{jk}(t) = \sum_{l=1}^{K} p^-_{kl}(j)\pi^1_l(t) - \pi^1_k(t). \tag{5.96}$$

Equations (5.93)-(5.96) imply that

$$G^s_{jk}(t) = \beta G^0_{jk}(t) + (1-\beta)G^1_{jk}(t). \tag{5.97}$$

The decomposition (5.97) implies that the proposed policy adopts a weighted combination of two distinct policies: $G^0_{jk}(t)$ and $G^1_{jk}(t)$. The index policy $G^0_{jk}(t)$ coincides with the index policy derived in section 5.2.2 when the planning horizon $T \to \infty$. On the other hand, the index policy $G^1_{jk}(t)$ is obtained by performing one step of the policy iteration algorithm starting from the optimal static policy u^s, and taking $\beta = 0$ in the objective. Therefore, the proposed policy attempts to balance the conflicting objectives by taking the average of two indices: one that emphasizes medical benefit, and a second one that focuses on waiting time inequity.

It is worth emphasizing the relationship between the proposed policy and the UNOS policy. Similarly to the UNOS policy, the proposed policy allocates priority points based on tissue matching, $G^0_{jk}(t)$, and additional points based on waiting time criteria, $G^0_{jk}(t)$. However, in contrast to the UNOS policy, the proposed dynamic policies adopt indices that explicitly capture the dynamics of the system and *anticipate* the impact of an allocation decision to future equity and medical benefit.

5.5 Concluding Remarks:

In the last two chapters we have developed two models for the kidney transplant waiting list. The first model is a queueing system with reneging that was used to provide new insights about the performance of the kidney transplant waiting list and about factors that generate inequity between different patient classes. The second model is a fluid model that integrates the kidney transplant waiting list with the process of organ rejection. Embedded in this model are an objective that captures medical benefit, a constraint that predetermines the fraction of organs allocated to each patient class, and a bi-criteria objective that combines the medical benefit objective with the queueing time inequity metric motivated by the reneging model. Using an optimal control theoretic approach, we develop allocation policies that optimize each of these objectives subject to the dynamics of the system and the constraints on the fraction of the transplanted organs. The analysis of the optimal control problems gives rise to three simple heuristics: the efficiency policy, the subsidized index policy and the dynamic priority policy. A modified form of the subsidized index policy is is tested in Chapter 7 using the simulation model of Chapter 6.

Chapter 6

The Kidney Transplantation Model

In order to evaluate the current allocation policy and the index policies of Chapter 5 we develop the Kidney Transplantation Model (KTM). This is a modular computer simulation program that captures the dynamics of the ESRD population both before and after transplantation, and it has the ability to simulate different allocation policies under alternative environmental variables. To ensure that KTM realistically captures the dynamics of ESRD, we have used data from more than 40,000 kidney transplantations and detailed summary statistics about ERSD.

In this Chapter, we describe KTM and the sources of data used for its development. The chapter is organized as follows: Section 6.1 describes the data used for this study. Section 6.2 presents a general conceptualization of KTM. Section 6.3 describes the various components of KTM. General remarks that synthesize some of the lessons learned from the model building exercise are given in Section 6.4.

6.1 Data

We use data provided by the United Network of Organ Sharing (UNOS) and the United States Renal Data System (USRDS). The data include transplant survival data (UNOS Public Use Data Set [48]), summary statistics about the size of the waiting list (UNOS

Annual Report, 1995 [49]) and summary statistics about the incidence and prevalence of ESRD and the mortality of ESRD patients (USRDS Annual Report, 1994 [50]).

UNOS Public Use Data: The UNOS public use data set contains records of 40,114 kidney transplants performed in the United States between October 1987 and June 1991. For each transplant case, the data set gives right censored graft survival times, patient and donor demographics, and patient and donor medical characteristics; see Table 6.1 for a detailed description.

We use this data set to estimate the parameters for the kidney transplant model and to cross validate the model. To do that we subdivide the set into two subsets: the *training* set which contains the first 30,000 cases, and the validation set which contains the remaining 10,114.

UNOS Annual Report: This report gives annual statistics about the size of the waiting list for kidney transplantation, the number of new transplants and the number of deaths in the waiting list for each year between 1988 and 1994. These statistics are used to estimate the arrival rate of new patients and organs.

USRDS Annual Report: The USRDS maintains the most comprehensive data set about ESRD. This data set contains information about all patients treated under Medicare's ESRD program since 1982; this accounts for more than 93% of the total ESRD population. In our analysis we utilize detailed summary statistics that are presented in the USRDS annual report. These statistics include information about the *incidence* and *prevalence* of ESRD, the *mortality* and *causes of death* of ESRD patients, methods of treatment, the kidney transplant process and patient survival. Because we are interested in data that complement the data from UNOS, we focus on the incidence, prevalence and mortality statistics.

The *incidence* statistics give the number of patients starting ESRD therapy by age, gender, race and primary disease for the calendar years 1982 until 1992 . The prevalence statistics give point prevalence of ESRD on December 31 by year, age, gender, sex and primary disease. The *mortality* statistics give death rates during 1990-92 by gender, race

hicode	encrypted patient identifier
doncode	encrypted donor identifier
tx_date	transplant date (plus or minus a random offset)
provcode	hospital identifier
abo	patient blood type
descrip	patient functional status at the time of transplant
don_type	cadaver or living related donor
previous	previous transplants
race	patient race and ethnicity
sex	patient sex
age	patient age
pretx_tf	number of pretransplant transfusions
cisch_ti	cold ischemic time
g_status	graft status at time of last reported follow-up date
g_time	number of days from transplant to graft failure or last follow-up date
p_status	patient status at time of last reported follow-up
p_time	number of days from transplant to last follow-up
mult1,mult2	simultaneous multiple organ transplants
dabo	donor blood type
drace	donor race and ethnicity
dsex	donor gender
ndcod	cause of death
dage	donor age
cpra	most recent panel reacting cytotoxic antibody (PRA)
ppra	peak PRA
a1,a2	recipient HLA A type
b1,b2	recipient HLA B type
dr1,dr2	recipient HLA DR type
da1,da2	donor HLA A type
db1,db2	donor HLA B type
ddr1,ddr2	donor HLA DR type
height	patient height
weight	patient weight
iddm	patient insulin dependent diabetes
t_csa	cyclosporine therapy

Table 6.1: Description of the variables included in the UNOS public use data set.

and age, both for patients in dialysis and patients with functioning transplant.

6.2 The Conceptual Model

The kidney transplantation model is a *compartmental* simulation model with two major compartments: the *waiting list* compartment, and the *functioning graft* compartment; see Figure 6-1. The dynamics of the model are as follows: new ESRD patients join the waiting list compartment. From this compartment, patients can either leave the system because of death, or join the functioning transplant compartment following kidney transplantation. From the functioning transplant compartment the patients can either exit the system due to death, or rejoin the waiting list right after organ failure. The state of the model consists of two vectors: the waiting list vector and the post transplant vector. These vectors maintain the characteristics of the individuals in the two model compartments. Age, gender, race, tissue type, blood type, body surface area (bsa), peak panel reactive antibodies (ppra), previous transplantations and time since last entry to the current compartment are the characteristics common to both vectors. In addition to these, the post transplant vector maintains a one dimensional variable that reflects the quality of the transplanted kidney; this variable is extracted from the prognostic index of the graft survival model developed in subsection 6.3.6. To calculate this variable, the model uses the organ characteristics: donor age, donor gender, donor race, donor tissue type.

In the heart of KTM we have the following submodels:

1. Candidate Stream Model.

2. Donor Stream Model.

3. Waiting List Mortality Model.

4. Organ Allocation Model.

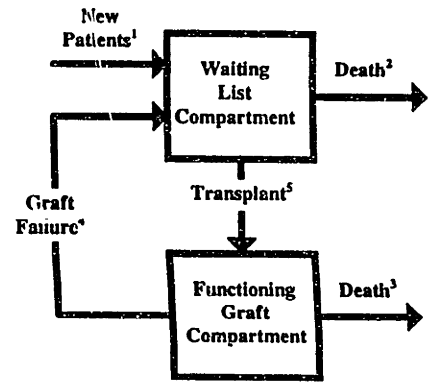

Figure 6-1: The kidney transplant model; (1) New patient arrivals are generated by the candidate stream model; (2) Deaths in the waiting list are generated by the waiting list mortality model; (3) Deaths of patients with functioning graft are generated by the post transplant mortality model; (4) Organ failures are generated by the post transplant failure model; (5) New transplants are generated by the donor stream model, and are allocated by the organ allocation model.

5. Post Transplant Mortality Model.

6. Post Transplant Graft Failure and Relisting Model.

These submodels are the backbone of our model (see figure 6-1) and are used to update the state of the system in periodic (monthly) intervals. This is done according to the following sequence of events: At the beginning of each simulated month the candidate stream model generates new patient arrivals. Following that, the waiting list mortality model generates deaths in the waiting list. Next, the post transplant graft failure model generates organ failures and decides whether the candidates rejoin the waiting list or withdraw from the system. Finally, the donor stream model generates organ arrivals and the organ allocation model allocates these kidneys to patients on the transplant waiting list. Summary statistics about the evolution of the system are collected at the end of the time period. This sequence of events is described in Figure 6-2.

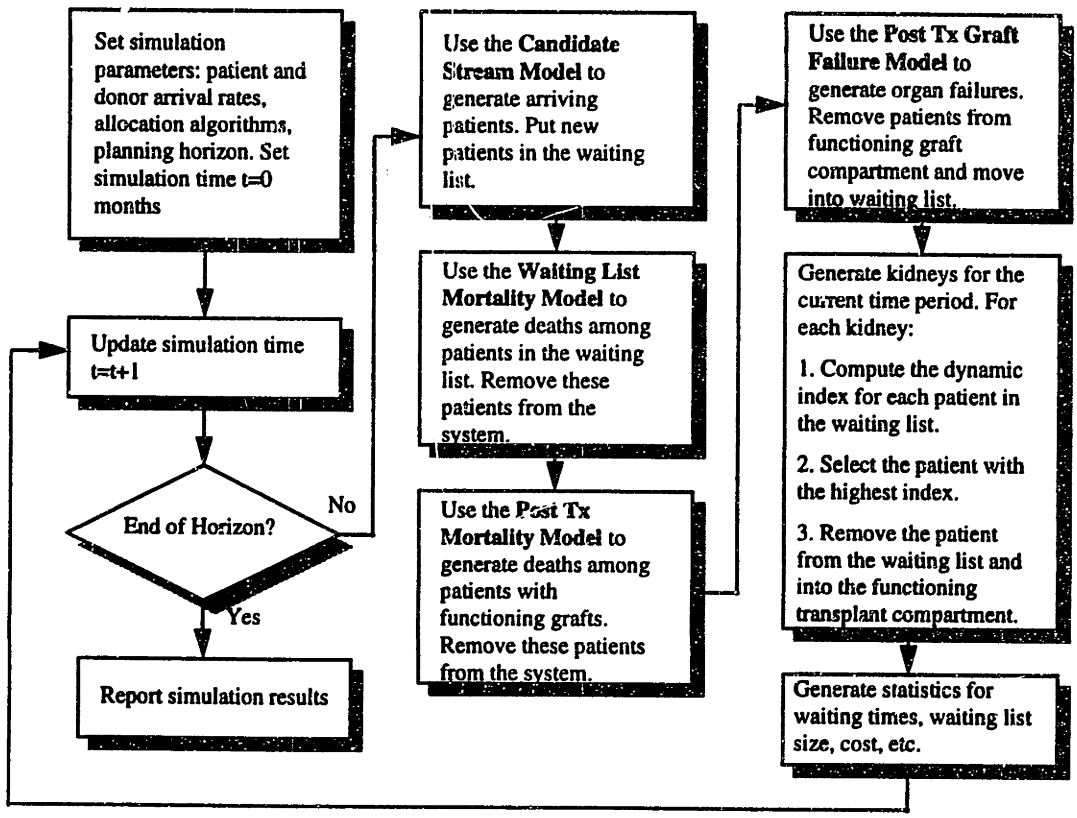

Figure 6-2: Flowchart for KTM; Tx is a commonly used abbreviation for transplantation.

6.3 Description of the Submodels

The six submodels are critical to the development of the kidney transplant model. Specifically, it is through a very careful selection of these models that we expect to realistically capture the dynamics of the waiting list and develop a highly reliable simulation model. To develop these submodels we went through several iterations of exploratory data analysis. In this section, we present the final results.

6.3.1 Candidate Stream Submodel

The candidate stream submodel generates new transplant candidates and their characteristics; recall that the characteristics of new candidates are age, gender, race, tissue type, blood type, peak pra and body surface area. To do that the model assumes that patients arrive according to a non homogeneous Poisson process and that their characteristics are independent of the arrival process. To generate the characteristics, the model uses historical data about the incidence of ESRD and the distribution of tissue types, ppra and body surface area in the ESRD population.

Age, Gender and Race:

To generate the age, gender and race of new candidates, we use data about the incidence of ESRD that are reported in the USRDS annual report. The data give the count of new ESRD cases by age, sex and race for the years 1987-1989, and 1990-1992. From these data we can estimate the fraction of new ESRD patients by age, gender and race. The results are described in Tables 6.2-6.4. To generate the age, gender and race of new patients, we adopt the following hierarchical algorithm: First, generate the gender of the candidate from Table 6.2. Next, use Table 6.3 to generate the race, and finally use Table 6.4 to generate the age.

A weakness of this approach is that the demographics of the simulated candidate stream are expected to resemble the demographics of the new ESRD but not the demo-

Female	Male
0.3891	0.6109

Table 6.2: Fraction of new ESRD patients by gender.

Female		Male	
African American	Caucasian	African American	Caucasian
0.3279	0.6721	0.2784	0.7216

Table 6.3: Fraction of new ESRD patients by race *given gender.*

graphics of the actual transplant candidates. In particular, because older ESRD patients are less likely to enroll for kidney transplantation, their proportion is higher among new ESRD cases than among new transplant candidates. This implies that the simulated candidate stream is expected to *overestimate* the fraction of older patients joining the waiting list. The only way to overcome this shortcoming is to utilize data about the actual waiting list registrations. However, because no such data are available, we will pretend that the incidence data from USRDS are sufficient.

Tissue Types

The tissue type is a combination of six proteins: two at loci HLA-A, two at loci HLA-B and two at loci HLA-DR. Therefore, to estimate the tissue type frequencies for new patients it is first necessary to estimate the protein frequencies for each of the six loci. If the proteins at the six loci are independent, then we can generate different tissue types directly from the individual protein frequencies. Indeed, the six loci appear to be independent so it suffices to identify the protein frequencies by loci and race; see Barnes and Miettinen, 1972.

To estimate the HLA protein frequencies, we use data from the UNOS public use data set. This data set gives information about the tissue types of *all* organs donated between October 1987 and June 1991. If we assume that stratified by race, the frequency of tissue types in the donor pool is the same as the frequency of tissue types in the new patient pool, then we can use the UNOS data set to estimate the frequency of tissue

Age	Female		Male	
	African American	Caucasian	African American	Caucasian
20-24	0.0197	0.02061	0.0237	0.0148
25-29	0.0309	0.0353	0.0403	0.0267
30-34	0.0418	0.0396	0.0614	0.0387
35-39	0.0458	0.0436	0.0907	0.0463
40-44	0.0526	0.0483	0.0972	0.0550
45-49	0.0685	0.0527	0.0914	0.0584
50-54	0.0872	0.0675	0.0954	0.0649
55-59	0.1160	0.0905	0.1014	0.0791
60-64	0.1435	0.1306	0.1119	0.1110
65-69	0.1555	0.1578	0.1085	0.1505
70-74	0.1143	0.1367	0.0820	0.1504
75-79	0.0754	0.1046	0.0567	0.1193
80-84	0.0338	0.0532	0.0273	0.0615
85+	0.0150	0.0189	0.0121	0.0233

Table 6.4: Fraction of new ESRD patients by age *given gender and race.*

types in new patients. The derived estimates appear in Tables 6.5-6.7. The columns of the tables are numbered (1)-(4) and are as follows: column (1) gives the protein type, column (2) gives the total number of donors that had this protein type, column (3) gives the frequency of this protein type, and column (4) gives the cumulative frequency.

To demonstrate how these tables can be used to obtain the frequencies of different tissue types, let us obtain the frequency of tissue type (23,30,7,70,3,11) among African Americans; this tissue type should be read as follows: HLA-A1 protein 23, HLA-A2 protein 30, HLA-B1 protein 7, etc. From Tables 6.5-6.7, the probability that an African American individuals has this tissue type is $(0.128)(0.185)(0.177)(0.275)(0.190)(0.144) = 3.2 \times 10^{-5}$.

Blood Type:

To generate the blood type of new candidates, we use the frequencies in Table 6.8. These frequencies were extracted from the UNOS Public use data base and give the frquencies of the four blood types among African American and Caucasian donors. It is assumed

Caucasians								African Americans							
HLA-A1				HLA-A2				HLA-A1				HLA-A2			
(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
2	12957	0.402	0.402	3	4337	0.159	0.159	2	1236	0.337	0.337	30	570	0.185	0.185
1	9585	0.297	0.700	24	3653	0.134	0.293	23	469	0.128	0.465	33	429	0.139	0.324
3	4295	0.133	0.833	2	3179	0.117	0.410	3	441	0.120	0.585	23	291	0.094	0.419
24	1590	0.049	0.882	11	2515	0.092	0.502	1	372	0.101	0.686	28	265	0.086	0.505
11	1380	0.043	0.925	32	2152	0.079	0.581	30	292	0.080	0.766	34	247	0.080	0.585
26	480	0.015	0.940	29	1950	0.072	0.653	28	214	0.058	0.824	3	171	0.056	0.641
23	390	0.012	0.952	28	1754	0.064	0.717	33	102	0.028	0.852	29	141	0.046	0.686
28	341	0.011	0.963	31	1612	0.059	0.777	24	92	0.025	0.877	36	118	0.038	0.725
25	339	0.011	0.973	26	1561	0.057	0.834	26	87	0.024	0.901	31	106	0.034	0.759
29	286	0.009	0.982	30	1303	0.048	0.882	29	72	0.020	0.921	74	102	0.033	0.792
30	190	0.006	0.988	25	945	0.035	0.917	11	69	0.019	0.939	2	101	0.033	0.825
32	143	0.004	0.992	23	773	0.028	0.945	34	47	0.013	0.952	68	100	0.032	0.857
31	136	0.004	0.997	33	552	0.020	0.965	36	30	0.008	0.960	32	87	0.028	0.886
68	29	0.001	0.997	68	435	0.016	0.981	68	28	0.008	0.968	24	86	0.028	0.914
33	27	0.001	0.998	1	104	0.004	0.985	32	24	0.007	0.975	26	85	0.028	0.941
10	21	0.001	0.999	34	102	0.004	0.989	31	22	0.006	0.981	11	48	0.016	0.957
9	13	0.000	0.999	69	54	0.002	0.991	25	19	0.005	0.986	19	31	0.010	0.967
34	6	0.000	1.000	19	49	0.002	0.992	74	16	0.004	0.990	69	26	0.008	0.975
36	6	0.000	1.000	97	49	0.002	0.994	19	13	0.004	0.994	25	25	0.008	0.983
19	5	0.000	1.000	10	47	0.002	0.996	10	9	0.002	0.996	66	23	0.007	0.991
69	2	0.000	1.000	98	27	0.001	0.997	9	6	0.002	0.998	97	11	0.004	0.994
74	2	0.000	1.000	74	25	0.001	0.998	69	6	0.002	0.999	10	7	0.002	0.997
				9	24	0.001	0.999	66	2	0.001	1.000	9	6	0.002	0.999
				36	18	0.001	0.999					1	2	0.001	0.999
				66	15	0.001	1.000					98	2	0.001	1.000

Table 6.5: HLA-A frequencies.

Caucasians								African Americans							
HLA-B1				HLA-B2				HLA-B1				HLA-B2			
(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
7	7687	0.239	0.239	44	4526	0.153	0.153	7	648	0.177	0.177	53	482	0.15	0.15
8	5944	0.185	0.423	62	3428	0.116	0.27	35	347	0.095	0.271	70	399	0.124	0.275
44	4166	0.129	0.552	60	2908	0.099	0.368	44	315	0.086	0.357	58	295	0.092	0.367
35	2453	0.076	0.629	35	2753	0.093	0.461	53	265	0.072	0.43	57	227	0.071	0.438
18	1557	0.048	0.677	51	1702	0.058	0.519	8	228	0.062	0.492	35	220	0.069	0.506
27	1408	0.044	0.721	57	1421	0.048	0.567	45	225	0.061	0.553	42	154	0.048	0.554
51	1357	0.042	0.763	8	1218	0.041	0.609	42	219	0.06	0.613	44	153	0.048	0.602
14	1272	0.039	0.802	18	1198	0.041	0.649	18	174	0.047	0.66	45	136	0.042	0.645
13	988	0.031	0.833	27	1181	0.04	0.689	14	147	0.04	0.7	49	107	0.033	0.678
62	708	0.022	0.855	55	744	0.025	0.714	57	119	0.032	0.733	63	95	0.03	0.708
39	521	0.016	0.871	49	708	0.024	0.738	17	118	0.032	0.765	51	80	0.025	0.733
57	505	0.016	0.887	14	698	0.024	0.762	51	117	0.032	0.797	60	74	0.023	0.756
60	484	0.015	0.902	39	683	0.023	0.785	58	95	0.026	0.823	62	72	0.022	0.778
38	465	0.014	0.916	61	598	0.02	0.805	49	84	0.023	0.846	18	69	0.022	0.8
37	350	0.011	0.927	37	565	0.019	0.825	39	52	0.014	0.86	72	62	0.019	0.819
49	290	0.009	0.936	38	549	0.019	0.843	13	50	0.014	0.874	17	53	0.017	0.836
17	241	0.007	0.944	13	473	0.016	0.859	52	50	0.014	0.887	50	48	0.015	0.851
55	190	0.006	0.949	50	373	0.013	0.872	27	40	0.011	0.898	8	47	0.015	0.865
41	170	0.005	0.955	41	338	0.011	0.883	70	37	0.01	0.908	52	42	0.013	0.878
52	168	0.005	0.96	45	311	0.011	0.894	41	34	0.009	0.918	39	40	0.012	0.891
45	148	0.005	0.965	7	309	0.01	0.904	62	33	0.009	0.927	65	35	0.011	0.902
50	144	0.004	0.969	58	302	0.01	0.914	63	30	0.008	0.935	27	34	0.011	0.912
5	127	0.004	0.973	52	297	0.01	0.925	5	28	0.008	0.942	14	30	0.009	0.922
22	109	0.003	0.976	56	276	0.009	0.934	60	26	0.007	0.95	7	24	0.007	0.929
56	88	0.003	0.979	40	229	0.008	0.942	50	25	0.007	0.956	41	24	0.007	0.937
58	88	0.003	0.982	53	223	0.008	0.949	37	23	0.006	0.963	48	23	0.007	0.944
53	74	0.002	0.984	65	217	0.007	0.957	38	22	0.006	0.969	61	22	0.007	0.951
40	71	0.002	0.986	17	210	0.007	0.964	65	22	0.006	0.975	55	21	0.007	0.957
61	71	0.002	0.989	63	201	0.007	0.971	56	10	0.003	0.977	71	20	0.006	0.963
16	49	0.002	0.99	22	190	0.006	0.977	22	9	0.002	0.98	37	19	0.006	0.969
63	49	0.002	0.992	70	154	0.005	0.982	40	9	0.002	0.982	5	14	0.004	0.974
65	48	0.001	0.993	47	104	0.004	0.986	48	9	0.002	0.985	15	12	0.004	0.978
15	45	0.001	0.994	48	74	0.003	0.988	64	9	0.002	0.987	13	9	0.003	0.98
12	42	0.001	0.996	64	53	0.002	0.99	15	8	0.002	0.989	56	9	0.003	0.983
47	31	0.001	0.997	21	51	0.002	0.992	72	8	0.002	0.992	40	7	0.002	0.985
21	29	0.001	0.998	15	50	0.002	0.993	12	7	0.002	0.993	22	6	0.002	0.987
42	25	0.001	0.998	42	35	0.001	0.995	16	7	0.002	0.995	64	6	0.002	0.989
48	19	0.001	0.999	16	23	0.001	0.995	55	4	0.001	0.996	54	5	0.002	0.991
64	10	0	0.999	5	21	0.001	0.996	61	4	0.001	0.998	21	4	0.001	0.992
70	10	0	1	97	20	0.001	0.997	59	2	0.001	0.998	38	4	0.001	0.993
				54	16	0.001	0.997	71	2	0.001	0.999	75	4	0.001	0.994
				98	15	0.001	0.998	21	1	0	0.999	97	4	0.001	0.996
				72	14	0	0.998	46	1	0	0.999	98	4	0.001	0.997
				73	13	0	0.999	47	1	0	0.999	47	3	0.001	0.998
				12	11	0	0.999	54	1	0	1	16	2	0.001	0.998
				75	7	0	0.999	75	1	0	1	77	2	0.001	0.999
				67	6	0	1					12	1	0	0.999
												46	1	0	1

Table 6.6: HLA-B frequencies.

Caucasians								African Americans							
HLA-DR1				HLA-DR2				HLA-DR1				HLA-DR2			
(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
2	6475	0.203	0.203	7	5304	0.204	0.204	2	753	0.207	0.207	6	437	0.144	0.144
4	6224	0.196	0.399	4	4099	0.157	0.361	3	688	0.190	0.397	11	437	0.144	0.289
1	5746	0.181	0.580	6	3481	0.134	0.495	1	460	0.127	0.524	7	392	0.130	0.418
3	5382	0.169	0.749	11	2534	0.097	0.592	7	310	0.085	0.609	13	253	0.084	0.502
7	2242	0.070	0.819	3	1809	0.069	0.662	6	284	0.078	0.688	8	238	0.079	0.580
6	1669	0.052	0.872	13	1538	0.059	0.721	11	264	0.073	0.760	5	208	0.069	0.649
11	1031	0.032	0.904	8	1405	0.054	0.775	4	236	0.065	0.825	3	180	0.059	0.709
5	998	0.031	0.935	5	1274	0.049	0.824	5	204	0.056	0.882	15	176	0.058	0.767
15	582	0.018	0.954	15	1022	0.039	0.863	8	91	0.025	0.907	12	137	0.045	0.812
13	421	0.013	0.967	2	823	0.032	0.895	13	84	0.023	0.930	9	116	0.038	0.850
8	361	0.011	0.978	12	725	0.028	0.922	15	70	0.019	0.949	4	107	0.035	0.886
12	161	0.005	0.983	9	596	0.023	0.945	9	60	0.017	0.966	10	95	0.031	0.917
9	128	0.004	0.987	14	450	0.017	0.963	12	54	0.015	0.980	2	83	0.027	0.945
17	112	0.004	0.991	10	408	0.016	0.978	10	34	0.009	0.990	14	77	0.025	0.970
10	111	0.003	0.994	17	238	0.009	0.987	14	20	0.006	0.995	18	48	0.016	0.986
14	109	0.003	0.998	1	123	0.005	0.992	18	9	0.002	0.998	17	19	0.006	0.992
16	36	0.001	0.999	18	76	0.003	0.995	17	5	0.001	0.999	16	8	0.003	0.995
18	31	0.001	1.000	97	49	0.002	0.997	16	2	0.001	1.000	97	8	0.003	0.997
				16	47	0.002	0.999	97	1	0.000	1.000	1	7	0.002	1.000
				98	31	0.001	1.000	52	0	0.000	1.000	98	1	0.000	1.000
								98	0	0.000	1.000	52	0	0.000	1.000

Table 6.7: HLA-DR frequencies.

Blood Type	African American	Caucasian
A	0.2506	0.4419
B	0.2100	0.1037
AB	0.0348	0.0284
O	0.5046	0.4560

Table 6.8: Blood type frequencies for new candidates and donors.

that the donor pool is an *unbiased* sample of the general population.

Peak Panel Reacting Antibodies (ppra):

For simplicity we assume that the ppra can take two values: low (when ppra< 60%) and high (when ppra > 60%); patients with low ppra are referred to as being *non-presensitized*, whereas patients with high ppra are referred to as being *presensitized*. To estimate the fraction of new candidates with high ppra we use data from the UNOS public use data system. Unfortunately, this data set gives information about the ppra of transplant recipients and not transplant candidates. To overcome this problem, we combine estimates about the ppra of transplant recipients, with bounds about the ppra of candidates that die in the waiting list. We use these bounds to obtain bounds on the proportion of high ppra among patients leaving the waiting list; this includes patients that leave the waiting list due to transplantation or death. Because in steady state the

Caucasians			
Male		Female	
Low	High	Low	High
16867	2863	10299	2837
African Americans			
Male		Female	
Low	High	Low	High
4294	765	2415	774

Table 6.9: Count of kidney transplant recipients with high and low ppra.

Caucasians	
Male	Female
0	0
African Americans	
Male	Female
1529.71	1523.29

Table 6.10: Estimated number of deaths in the waiting list by gender and race.

characteristics of the stream leaving the waiting list resemble the characteristics of the stream entering the waiting list, it follows that the the derived bounds are applicable to the ppra frequencies in the candidate stream.

Table 6.9 gives the count of kidney transplants with high ppra by gender and race between October 1987 and June 1991 (these numbers do not distinguish between cadaveric and living donors) and Table 6.10 gives estimates for the number of deaths by gender and race; for more details about the derivation of these estimates see the discussion at the end of this subsubsection. To get the lower bounds presented in Table 6.11, assume that all deaths are from low ppra candidates . On the other hand, to obtain the upper bounds of Table 6.11 assume that all the deaths are from patients with high ppra. KTM, uses the average of the two bounds.

To complete the discussion, it remains to outline the details behind the estimates in Table 6.10. First, we use data from the UNOS Annual Report to estimate the total number of deaths between October 1987 and June 1991. This is done as follows: The UNOS report gives the total number of deaths in the waiting list for each year since 1988.

Lower Bounds				Upper Bounds			
Caucasians				Caucasians			
Male		Female		Male		Female	
Low	High	Low	High	Low	High	Low	High
0.8549	0.1451	0.7840	0.2160	0.8549	0.1451	0.7840	0.2160
African Americans				African Americans			
Male		Female		Male		Female	
Low	High	Low	High	Low	High	Low	High
0.8839	0.1161	0.8357	0.1643	0.6517	0.3483	0.5125	0.4875

Table 6.11: Lower and Upper Bounds for ppra frequencies.

We assume that the number of deaths in 1987 is approximately the same as the number reported for 1988, and that the total number of deaths from October to December 1987 is $\frac{3}{12}$ths of the total number of deaths for that year. Similarly, the number of deaths between January and June 1991 is $\frac{1}{2}$ of the total number of reported deaths. The final estimate is 3053 deaths. Next, we break down this number by gender and race. To do that, we use equation (4.53) from Chapter 5. This equation states that if the demand for organs of type j is λ_j^+ and the supply is λ_j^- then the fraction of class j patients that receive transplant is $\frac{\lambda_j^-}{\lambda_j^+}$ and the observed death rate for class j patients is $\lambda_j^+ - \lambda_j^-$. Using these results, we can estimate the number of deaths by gender and race. Specifically, let $\lambda_{m,c}^+$ be the arrival rate for male Caucasians candidates, $\lambda_{f,c}^+$ the arrival rate for female Caucasians, $\lambda_{m,a}^+$ the arrival rate for male African Americans, and $\lambda_{f,a}^+$ the arrival rate for female African Americans. Similarly, let $\lambda_{m,c}^-, \lambda_{f,c}^-, \lambda_{m,a}^-, \lambda_{f,a}^-$ be the organ allocation rates by gender and race, and let λ^+ be the total candidate arrival rate, and λ^- the total organ arrival rate. The parameters $\lambda_{m,c}^-, \lambda_{f,c}^-, \lambda_{m,a}^-, \lambda_{f,a}^-$ can be estimated directly from the data of Table 6.9 and are as follows:

$$\begin{aligned} \lambda_{m,c}^- &= 19249 \text{ per time period} \\ \lambda_{f,c}^- &= 12817 \text{ per time period} \\ \lambda_{m,a}^- &= 4936 \text{ per time period} \\ \lambda_{f,a}^- &= 3112 \text{ per time period;} \end{aligned} \tag{6.1}$$

here the time period is 45 months (i.e. October 1987-June 1991). To estimate $\lambda^+_{m,c}, \lambda^+_{f,c}$, $\lambda^+_{m,a}, \lambda^+_{f,a}$, we first estimate λ^+ and then obtain $\lambda^+_{m,c}, \lambda^+_{f,c}, \lambda^+_{m,a}, \lambda^+_{f,a}$ using the following relations:

$$\begin{aligned}
\lambda^+_{m,c} &= P(R = \text{Caucasian}, G = \text{Male})\lambda^+ \\
\lambda^+_{f,c} &= P(R = \text{Caucasian}, G = \text{Female})\lambda^+ \\
\lambda^+_{m,a} &= P(R = \text{African American}, G = \text{Male})\lambda^+ \\
\lambda^+_{f,a} &= P(R = \text{African American}, G = \text{Female})\lambda^+ .
\end{aligned}$$

The estimate for $\lambda^+ = 40114 + 3053$ per time period; 40114 is the total number of transplant from Oct 1987 to June 1991, and 3053 is the total number of deaths in the same period. The final estimates for the patient arrival rates are:

$$\begin{aligned}
\lambda^+_{m,c} &= 19019 \text{ per time period} \\
\lambda^+_{f,c} &= 11288.77 \text{ per time period} \\
\lambda^+_{m,a} &= 7341.60 \text{ per time period} \\
\lambda^+_{f,a} &= 5507.5 \text{ per time period.}
\end{aligned} \tag{6.2}$$

Unfortunately, there is a discrepancy between the estimates in (6.1) and (6.2): the estimates in (6.2) imply that 30307.77 Caucasian candidates joined the waiting list between October 1987 and June 1991, whereas, the estimates in (6.1) imply that 32066 Caucasian candidates received kidney transplants in the same period. There are two possible explanations for this discrepancy: either the data are not reliable, or the queueing system is not in steady state.

Despite this discrepancy, we can still derive rough estimates about the total number of deaths by age and gender. More specifically, if we pretend that the discrepancy does not exist, we can use the estimates in (6.1) - (6.2) to conclude that $\frac{4936}{7341.60}$ of the African American male candidates receive transplant, and $\frac{3112}{5507.5}$ of the African American female candidates receive transplant. In contrast, almost all Caucasian candidates should expect

to receive kidney transplant; because the estimates for $\lambda^-_{m,c}$ and $\lambda^-_{f,c}$ exceed the estimates for $\lambda^-_{m,c}$ and $\lambda^-_{f,c}$. From this it follows that the death rates are : $\left(1 - \frac{4936}{7341.60}\right) 7341.60 = 2435.09$ deaths per time period for African American males, $\left(1 - \frac{3112}{5507.5}\right) 5507.5 = 2395.5$ for African American females, and 0 for Caucasians.

These numbers are also inconsistent with the current data: they imply that approximately 4848.59 deaths should have been observed between October 1987 and June 1991; this number is larger than the observed 3053. Nevertheless, we can loosely apply the lessons learned from these numbers to break down the observed deaths by gender and race. In particular, the estimates for the death rates imply that $\frac{2435.09}{2435.09+2395.5} = 0.5041$ of the observed deaths are African American males, and the remaining 0.4959 are African American females. Therefore, it is estimated that $(0.5041)(3053) = 1529.71$ of the deaths observed from October 1987 to June 1991 are African American males, and the remaining 1523.29 are African American females; see Table 6.10.

New Arrival Rate

We adopt two different estimates for the arrival rate; one that captures the "average" OPO and a second one that reflects a highly congested OPO (the New England Organ Bank). To obtain the estimates for the "average" OPO, we utilize data from the UNOS Annual Report and we assume that the arrival process is a non homogeneous Poisson process with linearly increasing trend. To obtain estimates for the NEOB we have contacted Jim Bradley from the NEOB.

Average OPO: To obtain the arrival rate into the "average" OPO we scale the national arrival rate by $\frac{1}{73}$; this reflects the intuition that the average OPO handles $\frac{1}{73}$ of the national ESRD cases. The national arrival rate is estimated to be 9778.35 patients per year (ppy) for year 1993, and the arrival rate into the "average" OPO is 133.95 ppy for year 1993; see the discussion after this paragraph. To forecast future arrival rates, we project the past trend in the arrival data. Specifically, the arrival rate appears to be increasing linearly for the five years between 1988 and 1992; see Table 6.12. The slope is

Year	1988	1989	1990	1991	1992	1993
New Arrivals	10302	9424	10167	11731	11340	11973

Table 6.12: New wating list arrivals by year.

estimated to be 327 candidates per (year)^2. This implies, that the future arrival rate into the average OPO is expected to increase linearly with a slope of $\frac{327}{73} = 4.48$ candidates per (year)^2.

To estimate the arrival rate of new patients, λ, we use the following data from the UNOS annual report: the size of the waiting list at the end of the year, x_t, the number of deaths, y_t, and the number of new cadaveric transplants, z_t, in years $t = 1988..1994$. From these data, we can estimate the number of new waiting list registrations λ_t by $t = 1988..1994$ using the flow conservation equation

$$x_t = x_{t-1} + \lambda_t - z_t - y_t; \tag{6.3}$$

(this equation assumes that recipients of living related organs do not contribute to the waiting list). The estimates are reported in Table 6.12. Unfortunately, we cannot estimate the new patient arrival rate directly from the numbers in Table 6.12 because they contain *both* new patients *and* old patients that rejoin the waiting list following graft failure. Therefore, we must first adjust these numbers to reflect *only* the new patients. We have estimated that the adjustment factor is 0.8167. This implies that the national arrival rate was 9778.35 ppy for year 1993.

To estimate the adjustment factor 0.8167 we use the following equation

$$\text{Fraction of First Time Candidates=} \\ \frac{\text{Number of First Time Tx + Number of Deaths w/out Previous Tx}}{\text{Number of Tx + Number of Deaths}}; \tag{6.4}$$

Tx is a commonly used abbreviation for "transplants".

The numbers for equation (6.4) can obtained from the USRDS Annual report and

	1988	1989	1990	1991	1992	Total
Total Number of Tx (UNOS)	7230	7086	7783	7732	7697	37528
Total Number of Tx (USRDS)	6441	6487	7093	7147	7057	34225
Total Number of First Time Tx (USRDS)	5471	5525	6027	6117	6082	29222
Deaths in the Waiting List (UNOS)	721	749	915	975	1046	4406

Table 6.13: Total Number of Transplants and Deaths per year.

the UNOS Annual report. The USRDS Annual Report gives the total number of transplants and total number of first time transplants that were reported to the Health Care Financing Administration for each year between 1988 and 1992 , and the UNOS report gives the total number of deaths that are reported to UNOS. Because we are interested in cadaveric donation, we assume that recipients of living related organs do not contribute to the waiting list. The relevant data are given in Table 6.13. Because the USRDS includes only the cases reported to the HCFA, its numbers are consistently lower than the numbers from UNOS. To adjust for this bias, we multiply the USRDS numbers by $\frac{37528}{34225}$. This implies that the total number of cadaveric transplants is 37528, the total number of first time transplants is $\frac{37528}{34225}(29222) = 32042.17$ and the total number of deaths in the waiting list is 4406.

So far, we have obtained three of the four quantities need to evaluate the right hand side of equation (6.4). To estimate the last quantity, which is the fraction of new waiting list registrations that are first time candidates, we need that total number of deaths from first time candidates. Although, we have no data about this number, we can consider two scenarios that lead to upper and lower bounds: the first scenario assumes that all deaths are first transplant candidates, and the second scenario assumes that all deaths are second or higher time transplant candidates. Using these scenarios in equation (6.4), we obtain the upper bound 0.8692 and the lower bound 0.7641. In the simulation study we use $\frac{0.8692+0.7641}{2} = 0.8167$.

New England Organ Bank: In 1995, 787 new transplant candidates joined the NEOB waiting list; this number includes both first time candidates and candidates for retransplantation. Using the factor developed before to adjust for retransplantation, we

conclude that that arrival rate of first time candidates into the NEOB is approximately $0.8167(787) = 642.74$ ppy for year 1995. To project future arrival rates, we assume that the NEOB arrival rate increases at the same linear rate as the national arrival rate. This implies that the arrival rate is expected to increase by 21.50 patients per $(\text{year})^2$.

Body Surface Area (bsa):

To generate bsa values for incoming candidates, we develop models for the probability density of the bsa values by gender and age. Specifically, let Y_{ij} be the bsa for an individual of age $i = 0..8$ and gender $j = 1..2$; age is discretized in increments of 10. Our model assumes that

$$\log(Y_{ij}) \sim N(\alpha + \mu_i + \eta_j, \sigma^2). \tag{6.5}$$

This model assumes that the log bsa values are normally distributed with mean that depends on the age and gender of the individual and constant variance. To estimate α, μ_i, η_j and σ, we fit equation (6.5) to the transplant data from the UNOS public use data base. This will give us estimates for the population of transplant recipients. Although this population is different from the population of transplant candidates, we assume that both populations have the same distribution of bsa values. Therefore, we can use the derived estimates to generate the bsa of new candidates. The results are given in Table 6.14

To test the validity of model (6.5) we plot the traditional regression diagnostics; see Figure 6-3. The first two plots of the top row give the residual plots and the third plot gives the regression line; the unusual clustered pattern in the regression plots is caused by the inclusion of categorical variables in the model. The first plot of the second row gives the normal plot of residuals, the second plot gives the residual fit spread plot, and the last plot gives the Cook's distance plot. Although the regression line captures the trend of the data and the residual plots are null, the normal plot suggests that the residuals may not be normally distributed. Nevertheless, the deviations from normality appear

Parameters	Value
Intercept	-0.420 ± 0 .016
Gender: male	0.121 ± 0.004
Age: 11-20	0.693 ± 0.002
Age: 21-30	0.881 ± 0.002
Age: 31-40	0.921 ± 0.002
Age: 41-50	0.948 ± 0.002
Age: 51-60	0.952 ± 0.002
Age: 61-70	0.942 ± 0 .017
Age: 71-80	0.936 ± 0 .027
Age: 81-90	0.872 ± 0.288
Std. Dev: σ	0.1471

Table 6.14: Estimates for the coefficients of bsa model. To obtain these estimates, we fit a linear regression model to the data from the UNOS public use data base.

to be rather small. Attempts to reduce the deviation from normality using additional covariates and different transformations of the bsa data were unsuccessful. Therefore, in KTM, we will use equation (6.5) with the parameters of Table 6.14.

6.3.2 Donor Stream Submodel:

The donor stream submodel generates new donors and their characteristics. The model assumes that new donors arrive according to a homogeneous Poisson process and that each donor provides two identical kidneys. The characteristics of the donors are independent of the arrival process and include the donor's gender, race, age and tissue type. To estimate the arrival rate and the frequency of the donor characteristics we use the UNOS public use data set. Because we are interested in cadaveric transplants, we only consider the 31513 transplantations that were performed using cadaveric organs. It is worth emphasizing that unlike the candidate arrival process, the donor arrival process is time homogeneous. This is justified because there are evidence that while the demand for organs is expected to increase, the supply of organs will remain fairly constant (Suthanthiran et al, 1994 [52]).

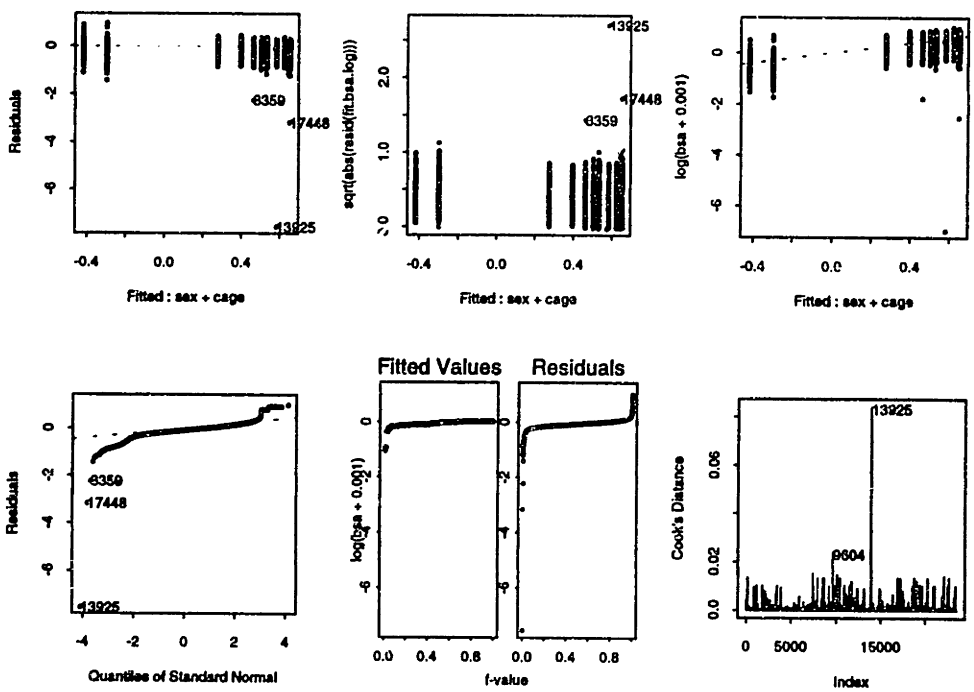

Figure 6-3: Regression diagnostics for the bsa model.

Female	Male
0.3677	0.6323

Table 6.15: Fraction of new donors by gender.

Female		Male	
African American	Caucasian	African American	Caucasian
0.0691	0.9309	0.1122	0.8878

Table 6.16: Fraction of new donors by race *given* gender.

Age, Gender and Race:

Tables 6.15-6.17 give the fraction of previous donors that are of a particular gender, race and age. By randomly sampling from these Tables, we can generate the characteristics of new donors.

Tissue Types and Blood Types:

To generate the tissue type of a simulated donor, we first generate the donor's race and then use Tables 6.5-6.7. To generate the blood type we use Table 6.8.

It is worth emphasizing that the characteristics of the simulated donor stream are expected to resemble the characteristics of previous donors. This is because Tables 6.15-6.17 and Tables 6.5-6.7 were derived using data about the actual kidney donations between 1987 and 1991.

	Female		Male	
Age	African American	Caucasian	African American	Caucasian
0-10	0.0742	0.0856	0.0688	0.0680
11-20	0.2400	0.2242	0.2118	0.2228
21-30	0.2568	0.2342	0.2520	0.2461
31-40	0.1506	0.1683	0.1768	0.1743
41-50	0.1338	0.1474	0.1494	0.1439
51-60	0.0987	0.1021	0.1056	0.1070
61-70	0.0437	0.0357	0.0334	0.0363
71-80	0.0020	0.0021	0.0021	0.0015

Table 6.17: Fraction of new donors by age *given* gender and race.

Arrival Rates:

As before, we need arrival rates for the "average" OPO and the NEOB. To obtain the estimate for the average OPO, we first estimate the national arrival rate and then divide it by 73. To obtain the NEOB rate we have contacted Jim Bradley.

Average OPO: From the UNOS data set we estimate that 31,513 cadaveric kidney transplants were performed from October 1987 to June 1991. This implies that the average number of cadaveric organ donations per year is 8220.78. Because from each donor two almost identical kidneys are harvested, it follows that the average number of new donors per year is 4110.39. Therefore, the donor arrival rate for the average OPO is 56.30 donors per year. It is worth emphasizing that using data from the UNOS annual report we have reached the slightly lower estimate of 52.92 donations per year. In KTM we use the first estimate.

New England Organ Bank: In 1995, 338 kidneys were recovered by the NEOB. This implies that the donor arrival rate into the NEOB is approximately $\frac{338}{2} = 169.0$ donors per year.

6.3.3 Waiting List Mortality Submodel:

The waiting list for kidney transplants is dynamic. Patients can either join the waiting list for the first time, or following graft failure. Patients can also leave the waiting list either to receive kidney transplantation or due to death. The waiting list mortality model simulates the process by which patients exit the waiting list due to death.

To estimate the death rates in the waiting list, we have utilized data from the USRDS annual report about the number of deaths in dialysis per 1000 patient year at risk by age, gender and race. From these data, we can obtain the probability of death in a given calendar year. The estimates are given in Table 6.18. Assuming that the general dialysis population resembles the population in the waiting list, we can use Table 6.18 to simulate the waiting list mortality process. It is worth emphasizing that because not all dialysis patients are eligible for transplantation, the estimates in Table 6.18 may overestimate

the mortality risk in the waiting list. However, we hope that by stratifying the estimates by age, we can partially correct for this bias.

6.3.4 Organ Allocation Model:

The organ allocation model simulates the process by which organs are allocated to transplant recipients. This involves the following steps:

1. Identify all zero mismatched candidates of the same blood type as the donor (both locally and nationally).

2. Compute the priority points for these candidates and offer the organ to the candidate with the highest points.

3. If the candidate cross reacts negative and accepts the offer, then perform the transplantation. Otherwise offer the kidney to the next candidate.

4. If there is no zero mismatched candidate or all zero mismatched candidates are exhausted, then allocate the kidney to the local candidate of the same blood type with the highest priority points.

5. If the candidate cross reacts negative and accepts the offer, then perform the transplantation. Otherwise offer the kidney to the next candidate.

To capture this sequence of events KTM uses the following simulation modules:

Mandatory Sharing of Kidneys: To capture the mandatory sharing of kidneys, we assume that there exists a *static* national waiting list of known demographic composition. Using the static waiting list, we can calculate the probability that there exists a zero mismatched candidate in the national list for each locally procured kidney. We can then use this probability to simulate the process of offering zero mismatched kidneys to these "hypothetical" candidates. In addition, because the local OPO is also likely to receive zero mismatched kidneys from the national donor pool, it is necessary to simulate the

Caucasian Males					
Age	20-24	25-29	30-34	35-39	40-44
Probability	0.0506	0.0728	0.1024	0.1251	0.1475
Age	45-49	50-54	55-59	60-64	65-69
Probability	0.1614	0.1963	0.2418	0.2896	0.3238
Age	70-74	75-79	80-84		
Probability	0.3881	0.4467	0.5330		
Caucasian Female					
Age	20-24	25-29	30-34	35-39	40-44
Probability	0.0546	0.0713	0.1056	0.1134	0.1269
Age	45-49	50-54	55-59	60-64	65-69
Probability	0.1557	0.1750	0.2159	0.2619	0.3119
Age	70-74	75-79	80-84		
Probability	0.3663	0.4302	0.4977		
African American Male					
Age	20-24	25-29	30-34	35-39	40-44
Probability	0.0557	0.0881	0.1075	0.1226	0.1155
Age	45-49	50-54	55-59	60-64	65-69
Probability	0.1209	0.1367	0.1542	0.1912	0.2414
Age	70-74	75-79	80-84		
Probability	0.3027	0.3637	0.4066		
African American Female					
Age	20-24	25-29	30-34	35-39	40-44
Probability	0.0603	0.0847	0.0801	0.0972	0.1048
Age	45-49	50-54	55-59	60-64	65-69
Probability	0.1078	0.1338	0.1454	0.1791	0.2301
Age	70-74	75-79	80-84		
Probability	0.2709	0.3336	0.3942		

Table 6.18: Probability of death in a calendar year by age, gender and race for all dialysis patients.

Characteristics	Total Number
Presensitized African Americans	3844
Non Sensitized African Americans	9889
Presensitized Caucasians	2471
Non Sensitized Caucasians	11251

Table 6.19: Composition of the static waiting list; the numbers are rounded to the closes integer.

national donor pool using the estimates in the donor arrival stream, and for each national donor test for the presence of a zero mismatched local candidate.

However, it is possible that this sharing scheme may generate a deficit of kidneys between the local OPO and the national list. To maintain a balance of flow, we adjust for kidney deficit every month; i.e. if at the end of each month the local OPO has received more kidneys from the national pool than it has sent to the national pool, then it will send the next few kidneys that it procures to the national list until the kidney deficit is balanced. Similarly, if the deficit is in the opposite direction, the OPO will receive from the national pool the next few kidneys until the deficit is balanced. It is worth emphasizing that UNOS requires its members to adopt similar practices to adjust the balance in the flow of kidneys between regions.

Static Waiting List: To simulate the mandatory sharing of zero mismatched kidneys, we must specify the size of the national waiting list, the fraction of Caucasian and African American candidates in the list and the proportion of presensitized candidates in each race. These numbers can be extracted from the UNOS 1995 Annual Report as follows: The waiting list at the end of 1994 was 27,455, and 50% of the registrants were minorities (assumed African Americans). The proportion of presensitized patients are 18% for Caucasians and 28% for African Americans (these numbers are averages of the bounds reported in Table 6.11). The composition of static waiting list is summarized in Table 6.19.

Kidney Offer and Acceptance: The kidney acceptance process is simulated as follows. The organ is first offered to the highest priority candidate. If the candidate

accepts the offer (an event of probability $p = 0.42$), then the model proceeds to test for positive crossmatching. If the candidate is presensitized, then the crossmatching test is positive with probability 0.854 (this is the average ppra among all cadaveric transplant recipient in the UNOS data set that have high ppra). Otherwise, if the candidate is not presensitized, the crossmatching test is positive with probability 0.092 (this is the average ppra among all cadaveric transplant recipients in the UNOS data set that have low ppra). If the candidate accepts the offer and crossmatches negative, the candidate leaves the waiting list to join the functioning graft compartment. Otherwise, the kidney is offered to the next candidate and the availability and crossmatching tests are repeated. If the kidney is still not transplanted after the second attempt, it is then offered to the third candidate which is assumed to accept the offer and crossmatch negative. Figure 6-4 gives a flow chart for this process.

Simulating the process of accepting a kidney offer is one of the biggest hurdles in the development of a kidney allocation model; see Kaufman et al., 1996 [53]. KTM bypasses this problem by using data from the UNOS and utilizing the mechanism behind the evaluation of ppra. To simulate the kidney acceptance offer , KTM utilize data from UNOS, 1991 [46] which state that only 35% of zero mismatched kidneys allocated between October and December 1991 were transplanted to the initially designated candidate. This implies that the probability that a candidates accepts an offer of a zero mismatched kidney is $\frac{0.35}{0.82} = 0.42$; the acceptance rate of 35% quoted by UNOS is the product of two probabilities: the probability that a caniddate accepts the offer, and the probability that a candidate crossmatches negative. We estimate from the UNOS public use data set that, on average, the probability of a negative crossmatch is 0.82 (the average peak pra for all transplants included in the UNOS data set is 0.18). This implies that the probability of acceptace is $\frac{0.35}{0.82} = 0.42$. In general, it is expected that the acceptance probability will be different from 0.42 for mismatched kidneys, and that the exact value will depend on such candidate characteristics as age, race, gender and socioeconomic conditions. This is because the kidney acceptance process incorporates the possibility that the candidate

may not be medically suitable for surgery at the time of the organ offer because of some other condition. For example, older candidates are more likely to suffer from pneumonia when an organ is offered and that precludes them from accepting the offer. Therefore, the acceptance probability will be smaller for older patients. Nevertheless, to avoid making assumptions that we cannot justify, in KTM we use a uniform acceptance probability $p = 0.42$ and we will perform sensitivity analysis using alternative values for $p = 0.56$, 0.70.

Simulating the crossmatching process is not as difficult. Recall that to determine the pra of a candidate, the candidate's blood is crossmatched with a large panel of donors selected at random, and pra is the fraction of positive crossmatches. This implies that pra gives the probability that a candidate will crossmatch positive with a potential donor. KTM uses the average peak pra values among non presensitized and presensitized patients to simulate this process. If instead of using the peak pra, KTM uses the current pra, then the probability of positive crossmatch for presensitized patients should become 0.4862 (average current pra among presensitized patients), and for non presensitized patients should be 0.0395 (average current pra among non presensitized patients). It is not clear which of the two sets of values should be preferred. However, statistical analysis in Chertow et al suggest that the peak pra is a better predictor of presensitization. Therefore, it appears that 0.854 and 0.092 should be preferred. The values from the current pra will be used for sensitivity analysis.

Point Scheme: All the allocation policies tested in KTM allocate policies based on an index that is a function of the characteristics of the organ and the candidate. These policies include the UNOS point system policy and the dynamic index policies of Chapter 5. More details about the priority indices will be given in the next chapter.

6.3.5 Post-Transplant Mortality Model:

To estimate the death rates for patients with functioning transplants we use mortality data from the USRDS annual report. These data give the number of deaths per 1000

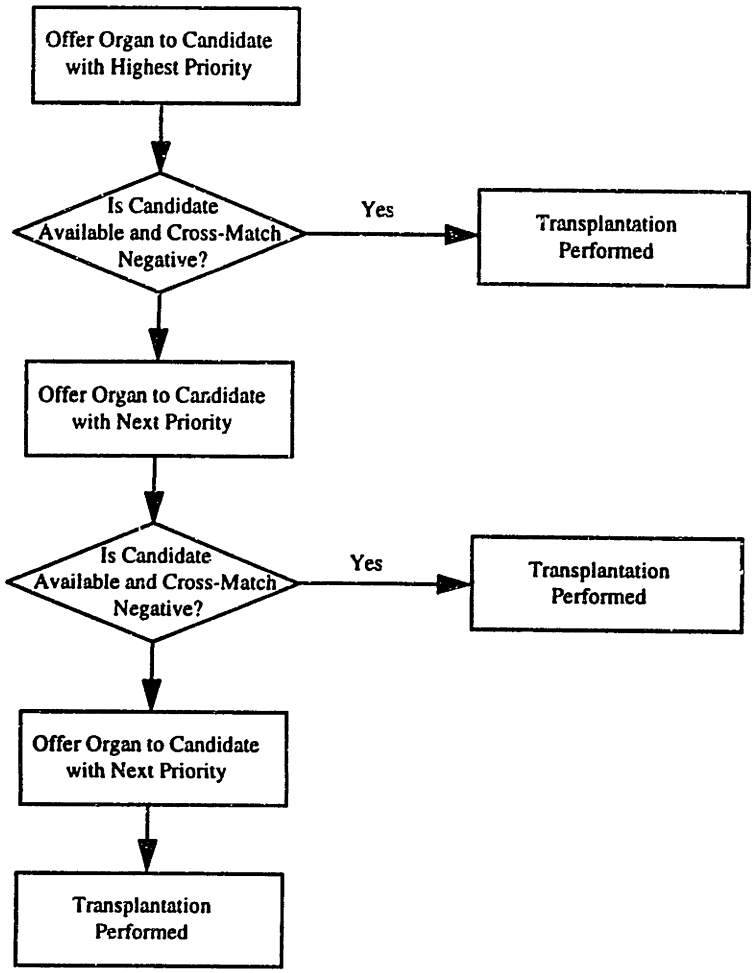

Figure 6-4: Flow chart diagram for the kidney acceptance process: The organ is offered to the first candidate. If the candidate is available (with probability 0.42) and the crossmatch is negative (with probability 0.146 for presensitized candidates, and probability 0.908 for non sensitized candidates), the perform the transplantation. Otherwise offer to the next candidate. If the organ is not transplanted to the next candidate, then transplant immediately to the third candidate.

patient years at risk by age, gender and race. From these data, we can obtain the probability of death in a given calendar year. The estimates are given in Table 6.20. We have attempted to fit a Weibull model to the data in Table 6.20 and our results (not shown) suggest that a such a model is also appropriate.

6.3.6 Post Transplant Graft Failure Model

In order to capture the different factors that predict the short and long term success of a transplant operation, we develop a statistical model for graft failure. The model takes the form of a proportional hazard model with non-parametric baseline. To develop this model, we rely heavily on techniques from survival analysis. Therefore, before we present our model, we will summarize some of the key results from survival analysis.

Survival Analysis

Survival analysis is the field of statistics that is concerned with the distribution of lifetimes, often of humans but also components and machines. Most of the applications of survival analysis come from the fields of medicine and biology. However, several important applications also occur in industrial engineering and manufacturing. Cox and Oakes (1984) [54] is the classical reference and provides the traditional treatment of the subject using applied statistics. Fleming and Harrington (1991) [55] take a mathematically rigorous approach to the subject using point processes and martingales. Here, we follow a treatment close to Cox and Oakes.

The basic model for survival analysis is the following. Let T denote the lifetime random variable; in our case this can denote the lifetime of the transplanted organ. Assume that T is a non-negative continuous random variable with density f and cumulative distribution function F. In survival analysis it is more convenient to use the *survivor function* $S(t) = 1 - F(t)$, and the *hazard function* $h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T \leq t+\Delta t | T \geq t)}{\Delta t}$. In our case, $S(t)$ gives the probability that the graft will fail no earlier than time t, and $h(t)\Delta t$ is the probability that a graft that survived until time t will fail in the next Δt time

Caucasian Males					
Age	20-24	25-29	30-34	35-39	40-44
Probability	0.0110	0.0102	0.0156	0.0207	0.0301
Age	45-49	50-54	55-59	60-64	65-69
Probability	0.0380	0.0465	0.0469	0.0736	0.0855
Age	70-74	75-79	80-84		
Probability	0.1000	0.0940	0.2424		
Caucasian Females					
Age	20-24	25-29	30-34	35-39	40-44
Probability	0.0058	0.0094	0.0130	0.0186	0.0256
Age	45-49	50-54	55-59	60-64	65-69
Probability	0.0328	0.0318	0.0393	0.0434	0.0751
Age	70-74	75-79	80-84		
Probability	0.0862	0.0617			
African American Males					
Age	20-24	25-29	30-34	35-39	40-44
Probability	0.0120	0.0186	0.0186	0.0186	0.0342
Age	45-49	50-54	55-59	60-64	65-69
Probability	0.0500	0.0568	0.0603	0.0948	0.1268
Age	70-74	75-79	80-84		
Probability	0.0700	0.1019			
African American Females					
Age	20-24	25-29	30-34	35-39	40-44
Probability	0.0107	0.0222	0.0170	0.0256	0.0271
Age	45-49	50-54	55-59	60-64	65-69
Probability	0.0295	0.0455	0.0464	0.0706	0.0727
Age	70-74	75-79	80-84		
Probability	0.1507				

Table 6.20: Probability of death in a given calendar year by age, gender and race for patients with functioning transplants.

units. The purpose of survival analysis is to obtain estimates for the hazard function and survival curve.

The most common parametric survival models are the *exponential* model with hazard λ and survivor function $S(t) = \exp(-\lambda t)$, and the *Weibull* model with hazard $\lambda\alpha\,(\lambda t)^{\alpha-1}$ and $S(t) = \exp\left(-\,(\lambda t)^{\alpha}\right)$. The exponential model assumes a constant hazard and is very restrictive. By contrast, the Weibull model assumes a monotonic hazard and appears to be a natural model for most failure processes. Moreover, because the Weibull model is very simple and flexible has been used with great success in several statistical studies.

The distinguishing characteristic of survival analysis is *censoring*. For example, in a clinical trial an individual case is not observed for its whole lifetime but only until the end of the trial. Similarly, transplant recipients are not observed until the time of graft failure but rather until the end of the follow up period. To make the analysis of censored data feasible, we must place some restrictions on the mechanism of censoring. The most common restriction is that censoring is independent of the lifetime variable. To see that this is a necessary assumption, observe that removal of observations right before failure can lead to misleading conclusions. A weaker assumption, called the *independent censoring* assumption, is that the hazard at time t conditional on the *whole* history of the process until time t depends only on the survival of that individual at time t.

To illustrate the statistical analysis of censored data, consider the following example. Suppose there are n observations; the first $m\,(< n)$ observations are uncensored, whereas the last $n - m$ observations are censored. Let t_i denote the failure time for $i = 1...m$, and the censoring time for $i = m + 1...n$. Under the independent censoring assumption, the likelihood for the parameters of f are

$$L = \prod_{i=1}^{m} f(t_i) \prod_{i=m+1}^{n} S(t_i).$$

The unknown parameters of f can be estimated by maximizing L.

In general, survival analysis has two objectives: to obtain estimates of the survivor

curve, and to identify how survival varies between groups. There are distinct tools that can be applied to study each of these questions. For example, to estimate baseline survivals we can use both non-parametric techniques and parametric techniques. To identify how survival depends on covariates we use the *proportional hazard* model.

Estimators of survivor curves:

To estimate the survivor curve we can either use nonparametric or parametric methods. The most widely used nonparametric estimator is the Kaplan-Meier estimator which is obtained by adjusting the hazard rate based on the number of cases at risk in each time. The estimator is as follows: Suppose that the time axis is divided into intervals of the form $I_i = [t_i, t_{i+1})$, and let d_i be the number of deaths in interval I_i, and $r(t)$ be the number of cases at risk just before time t. Then, the probability of surviving interval i is

$$p_i = \frac{r(t_i) - d_i}{r(t_i)},$$

and the probability of surviving until t_i is

$$S(t_i) \approx \prod_{j=0}^{i-1} \frac{r(t_j) - d_i}{r(t_j)}.$$

This can be refined further to give the Kaplan-Meir estimator

$$\hat{S}(t) \approx \prod_{\{j:t_j \leq t\}} \frac{r(t_j) - d_i}{r(t_j)}.$$

This estimator coincides with one minus the empirical distribution function when the data are uncensored. Moreover, the estimate becomes constant after the largest observed failure time and the points at the right hand end of the survivor curve are very variable; see Venables and Ripley (1994) for additional information.

Parametric estimates of the survivor curve can be obtained using maximum likelihood estimation. We will not discuss this here. Rather, we will describe a simple parametric estimator for Weibull models which is, to the best of our knowledge, new. This estimate

is based on the following property of the Weibull model: the survivor is linear on a complementary log-log plot (cloglog), i.e. the plot of $\log(-\log(S(t))$ versus t is linear, the slope of the cloglog plot gives α, and the intercept with the y-axis gives $\alpha \log \lambda$. We can utilize this property to obtain estimates of α and λ as follows: First, obtain the Kaplan-Meir survivor estimator. Next, plot the cloglog plot of the survivor curve. To that fit a linear model and estimate the slope and intercept. To obtain more precise estimates use weighted least squares where the weights are obtained from the standard error of the survivor curve.

Proportional Hazard Model:

To identify how survival varies between different groups, Cox (1972) introduced the *proportional hazard model.* This model assumes a baseline hazard function $h_0(t)$ which is modified multiplicative by covariates. The hazard function is

$$h(t) = h_0(t) \exp \beta^T x,$$

where x is the covariates vector. The interest is in the multiplicative factors rather than the baseline survival.

To estimate the parameter β, Cox developed the concept of the *partial likelihood.* The rationale behind the partial likelihood is that most of the information about β is contained in the conditional probability that case i died given that a death occurred at time t_j for all i and j. Now, conditional on the event that a death occurred at time t_j the probability that case i died is

$$\frac{h_0(t) \exp \beta^T x_i}{\sum_j I(t_j \geq t) h_0(t) \exp \beta^T x_j} = \frac{\exp \beta^T x_i}{\sum_j I(t_j \geq t) \exp \beta^T x_j}.$$

The partial likelihood is the product of such terms over all observed deaths. It is worth emphasizing that using the partial likelihood instead of the full likelihood leads to loss of efficiency. However, the loss is very small under some fairly weak assumptions (see Cox and Oakes).

Statistical Model

The statistical model for graft failure is a Cox proportional hazard model with non parametric baseline. The prognostic variables for the model are the recipient and donor gender, recipient and donor race, recipient and donor age, number of mismatches at the HLA-A, -B and -DR loci, peak pra, body surface area and previous transplantations. The continuous variables are categorized as follows: age and bsa into intervals of equal width; ppra into presensitized (ppra>60%) and non presensitized. The model does not take into account any interaction terms except from the interaction between donor and recipient sex; evidence in the literature suggest that male candidates that receive transplants from female donors have lower survival (see Brenner et al., 1992 [19]). Graft survival models similar to ours are described in Chertow et al., 1996 [18], and Houwelingen and Thorogood, 1995 [56].

The data used for the analysis are extracted from the UNOS public use data set. Not all patients in the data set were observed until graft failure time. Those patients that were lost to follow up, or died with a functioning organ or were with a functioning organ on the last day of follow up, are assumed to contribute a right-censored observation. To assess the validity of the model, the data set is subdivided into a training set (containing the first 30,000 patients), and a validation set (containing the last 10,114 patients). Furthermore, patients that received organs from living related donors are omitted from the analysis; the size of the reduced data set is 31,513. The training set is used to develop the model, and the validation set to cross validate. Missing values in the data set are treated as follows: patients with missing failure or censoring time are omitted from the analysis; patients with missing covariates are included in the analysis and missing values are treated as new categories. The analysis can also be repeated using a missing at random assumption and omitting all patients with missing values, but the conclusions remain the same.

Model estimation: The regression coefficients and the non-parametric baseline survivor can be obtained using the S-Plus system. The results are presented in Table 6.21

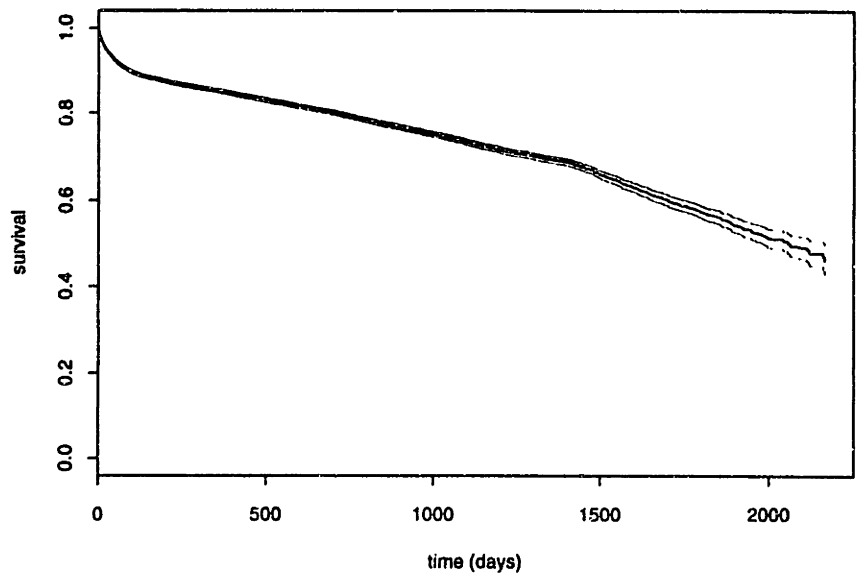

Figure 6-5: Survivor curve and 95% confidence bounds for a hypothetical individual with "average" prognostic index.

and Figure 6-5. The first column of Table 6.21 gives the prognostic factor, the second column gives the categories of the factor, the third column gives the regression coefficients with the 95% confidence intervals, and the fourth column gives the conditional risk relative to the baseline; this is the same as the exponential of the regression coefficient. Figure 6-5 gives the survivor curve for a patient with mean *prognostic index*; the prognostic index is defined by $PI = \sum_i \hat{\beta}_i X_i$, where X_i are the covariates and $\hat{\beta}_i$ the regression coefficients in Table 6.21; Chertow et al. [18] analyzed the data in the UNOS public use data set using a Cox model similar to ours and their estimates are consistent with the estimates in Table 6.21. It should be emphasized, here, that to reach the estimates in Table 6.21 we went through several iterations of exploratory data analysis in which we have utilized both the data in the training set and previous studies that utilized the same data as our data. For this reason, the estimates in Table 6.21 should be treated cautiously.

Factor	Categories	Regression Coefficient	Conditional Relative Risk
Donor and recipient sex	male to male or male to female or female to female	0 (baseline)	1
	female to male	0.1134 ± 0.0059	1.1201
Recipient race	Non African American	0 (baseline)	1
	African American	0.4205 ± 0.0055	1.5227
Donor race	Non African American	0 (baseline)	1
	African American	0.1653 ± 0.0823	1.1797
Recipient Age	0-10	0 (baseline)	1
	10-20	0.0828 ± 0.2081	1.0863
	20-30	-0.1640 ± 0.2077	0.8488
	30-40	-0.2586 ± 0.2065	0.7772
	40-50	-0.3414 ± 0.2077	0.7107
	50-60	-0.4141 ± 0.2107	0.6609
	60-70	-0.4670 ± 0.2225	0.6269
	70-80	-0.2562 ± 0.3706	0.7740
Donor Age	0-10	0 (baseline)	1
	10-20	-0.4667 ± 0.0098	0.6271
	20-30	-0.5019 ± 0.0097	0.6053
	30-40	-0.3592 ± 0.1011	0.6982
	40-50	-0.2167 ± 0.1027	0.8051
	50-60	-0.0142 ± 0.1072	0.9859
	60-70	0.1812 ± 0.1392	1.1987
	70-80	-0.4036 ± 0.6588	0.6679
Peak pra	presensitized	0 (baseline)	1
	non-presensitized	-0.3854 ± 0.0686	0.6679
Body Surface Area	0-0.50	0 (baseline)	1
	0.50-1.00	-0.0351 ± 0.1292	0.9655
	1.00-1.50	-0.0449 ± 0.1231	1.0459
	1.50-2.00	0.0786 ± 0.1239	1.0818
	2.00-2.50	0.1733 ± 0.1329	1.1892
	2.50-3.00	0.3266 ± 0.1609	1.3862
Previous transplants	0	0 (baseline)	1
	>0	0.2528 ± 0.0654	1.2876
HLA-A mismathces	0	0 (baseline)	1
	1	0.0921 ± 0.1031	1.0965
	2	0.1221 ± 0.1049	1.13
HLA-B mismatches	0	0 (baseline)	1
	1	0.1901 ± 0.1145	1.21
	2	0.2636 ± 0.1143	1.30
HLA-DR mismatches	0	0 (baseline)	1
	1	0.0994 ± 0.0954	1.10
	2	0.2495 ± 0.0957	1.28

Table 6.21: Results from fitting a cox proportional hazard model to the training set of 30,000 kidney transplants.

Model Validation: To establish the validity of the model, we perform three tests. The first test compares the distribution of the PI in the validation set, to the PI in the training set; the two distributions are virtually indistinguishable (results not shown). The second test fits a cox proportional hazard model to the validation data set using PI as the single prognostic variable. If the model is correct, then the derived coefficient for PI should be approximately 1; the actual value is indeed 1 with standard error 0.058. The final test compares the Kaplan-Meir survivor curve of the validation set to the mean survivor curve predicted by our model; see Figure 6-6(a). If the proposed model is correct, the mean survivor function is an unbiased estimator of the exact survivor. The test also categorizes the PI for the validation set into 4 groups: (each group contains about 25% of the data), and for each group, computes the Kaplan-Meir survivor and compares it to the mean survivor curve predicted by the Cox proportional hazard model; see Figure 6-6(b). The figures show that the mean survivor functions predict with precision the Kaplan Meier curves. This establishes the validity of our model.

Other Models:

To test whether it is possible to further improve our proposed model, we have considered several alternative models. The alternative models incorporated additional prognostic factors such as cold ischemia time, functional status of patients (employed or unemployed), health of patient (diabetic or not), etc. Although the coefficients for all these factors were significant, the predictions from these models were indistinguishable from our initial model. Therefore, for the sake of simplicity we decided not to incorporate these additional factors into our model. We have also considered models with several interaction terms, and in particular interactions between donor and recipient race. However, when these interaction terms were included, the fitting algorithm failed to converge to a solution. Therefore, although it is possible to further improve the proposed graft survival model, any improvement is marginal and for most practical purposes irrelevant.

Discussion:

The estimates in Table 6.21 and the plots in Figures 6-5 and 6-6 provide a wealth

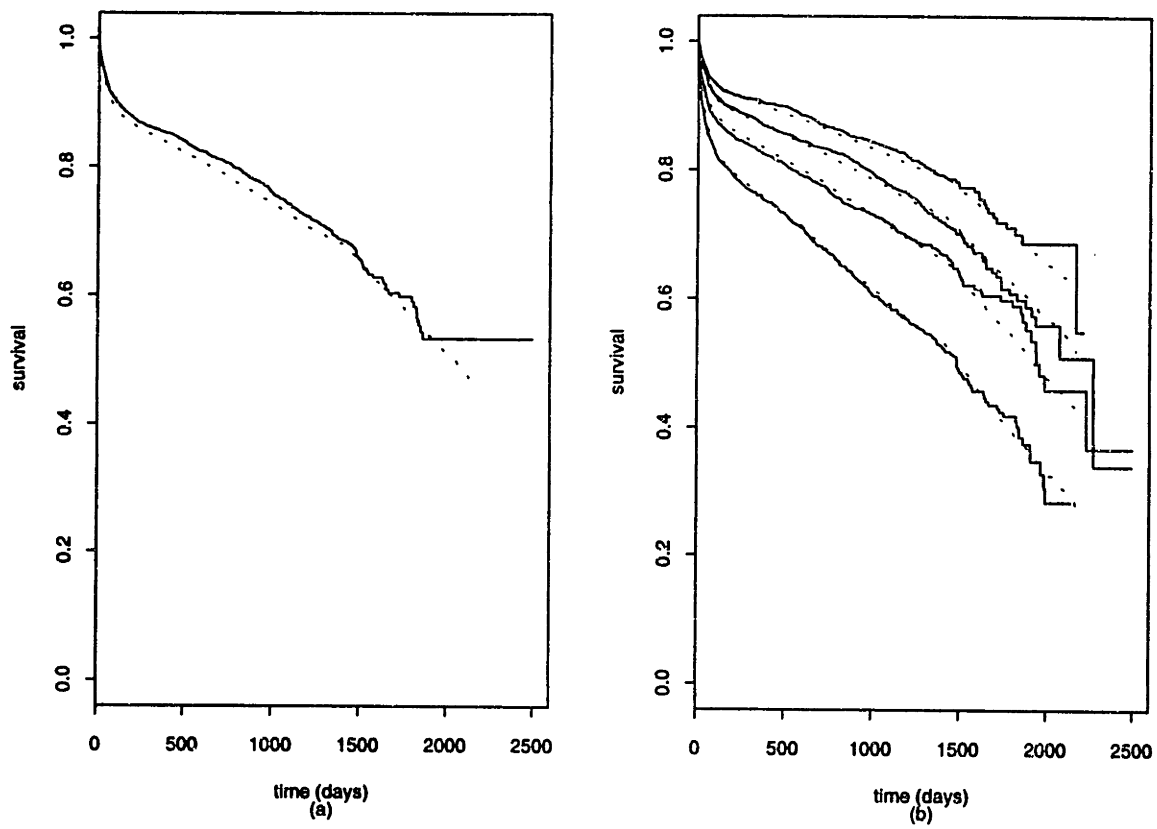

Figure 6-6: Validation of the graft survival model. Panel (a) gives the Kaplan-Meier survivor (solid line) and the mean survivor predicted by our model (dotted line). Panel (b) gives the Kaplan-Meier survivor (solid lines) and the mean survivor predicted by our model (dotted line) for four groups: $-0.5410 \leq PI < -0.3052$, $-0.3052 \leq PI < -0.0141$, $-0.0141 \leq PI < 0.1414$, and $PI > 0.1414$.

of information about the process of graft failure. Let us start first with the survivor plots. We observe that the graft failure process is characterized by three periods. The first period starts on the day of transplantation and ends after, approximately, 90 days. This period is characterized by a steep decline in the survivor curve and a high risk of immediate failure; approximately 10% of transplants fail within these first three months. The clinical phenomenon that gives rise to this steep decline in survival is known as *acute organ rejection* and is well studied; see [46]. The acute rejection period is followed by a period of medium and long term survival. The medium term survival period extends from day 90 until day 1500 and approximately 25% of the performed transplants fail within this period. During that period, the risk of graft failure is lower than before and the survivor curve drops at a linear rate. The medium term survival period is followed by the long term survival period after day 1500. This period is marked by a sudden drop in the survivor curve and a higher risk for organ rejection. Although we do not have any explanation for the sudden drop in survival after day 1500, it is well known that *chronic organ rejection* is the cause of graft failure after the first 90 days [46].

Let us now turn to the estimates in Table 6.21. Several observations are in place here:

1. African American transplant recipients have a higher failure rate than their Caucasian counterparts. The reasons for this finding are unclear. A study by Kasiske et al. (1991) [9] proposed that patient noncompliance, socioeconomic factors and disparities in known and unknown histocompatibility antigens are possible explanations. In a more recent study, Brenner et al. (1992) [19] suggested the hyperfiltration hypothesis to explain this effect. According to this hypothesis, African Americans have higher demand for renal function. This forces the transplanted organ to work harder than usual and causes premature failure. None of these hypotheses are confirmed.

2. Male recipients of female organs have a lower graft survival. Similarly to the previous observation, Brenner et al used the hyperfiltration hypothesis to explain this observation as well. Specifically, female organs are on average 17% smaller than

male organs and do not have the capacity to meet the demand for renal function imposed by the male body.

3. Recipients of organs from African American donors have a lower graft survival rate. Again, Brenner et al attempted to explain this observation through the hyperfiltration hypothesis: African Americans have smaller kidney mass than Caucasians. This smaller mass of the donated organs may reduce the organ's capability to meet the demand for renal function imposed by the human body. Although this hypothesis is not accepted by the medical community, it is the only hypothesis that attempts to explain this observation.

4. There is a non-linear relationship between recipient age and graft survival. The highest risk is observed for candidates of age 10 to 20. There is a gradual reduction in risk as age increases and the minimum risk is observed for candidates of age 60 to 70. A possible explanation for this trend is that younger patients are more likely to be non-compliant than the older patients and this increases their risk of graft failure.

5. Donor age is also strongly associated with graft failure. The ideal donor's age is between 10-40, and kidneys donated from younger or older donors are associated with high graft failure risk. Brenner's hyperfiltration hypothesis is proposed as a possible explanation for this observation.

6. Patients with high ppra have an increased graft failure risk. Because these patients are presensitized to a big majority of the population, they are more likely to produce antibodies that can cause premature organ rejection.

7. Body surface area is also associated with graft failure. In particular, there is a gradual increase in risk as bsa increases. Bremmer's hyperfiltration hypothesis provides a particularly attractive explanation for this phenomenon. This also suggests that the bsa of the donor should also be a factor that is associated with graft failure.

The bsa of the donor is a surrogate measure of the organ's size and, thus, of its ability to meet the renal demand imposed by the recipient's body. Unfortunately, the UNOS data set does not include any data about the donor's bsa.

8. Candidates with a history of previous transplantation have lower graft survival rate.

9. HLA compatibility is also strongly associated with graft survival. It also appears that compatibility at the B and DR loci is more strongly associated with graft failure, than compatibility at the A loci. What is particularly intriguing is that the magnitude of the relative risk associated with HLA mismatching is lower than the risk associated with several other non-immunological factors such as donor's and recipient's age. However it should be emphasized that the estimates from the Cox model may underestimate the risk associated with HLA matching. This is because approximately 25% of the HLA-DR types are thought to be inaccurately classified.

In conclusion, the results from the proportional hazard analysis suggest that several non-immunological factors are important determinants of graft survival. More importantly, these factors appear to be more strong predictors of long term graft survival than the traditional immunological factors. This suggests that it may be possible to improve the current organ allocation policies by utilizing some of these non-immunological factors. For example, the results in Table 6.21 cast some doubts on the utility of an allocation policy that allocates organs from female donors the male recipients. These results also suggest a "hypothetical" allocation policy that allocates organs to the candidates with the lowest prognostic index. Although such a policy is clearly unacceptable because it is racially and age biased, it can provide a useful benchmark to which the current allocation policy and the index policies of Chapter 5 can be compared. In the next Chapter, we present results from a simulation of the prognostic index policy.

6.3.7 Relisting:

The graft survival model is only one component of the Graft Failure/Relisting submodel. The second component of the submodel simulates the relisting process: Following graft failure, patients can either rejoin the waiting list or choose to leave the transplant system. Although there are no direct data about the fraction of patients that rejoin the waiting list following transplantation, indirect evidence suggest that this number is at least 50% (personal communication, Glen Chertow). KTM will use 75% as a baseline and will also consider several other possibilities for sensitivity analysis.

6.3.8 Initial Waiting List and Model Outputs:

So far, we have described the main functional components of KTM. These components provide the backbone of our model. However, to complete the description of the simulation program, we must specify the initial waiting list and the model outputs.

Initial Waiting List: Ideally, we would like to start the simulation from a "snapshot" of the actual waiting list. However, because no such data are available, we adopt a procedure that will generate a "hypothetical" snapshot. For the "average" OPO, we start the simulation from an empty system. We then simulate the old UNOS policy until the size of the waiting list reaches a level that is comparable to the current size of the waiting list in the "average" OPO. At that point, we freeze the system, collect data about the current state of the system, and use that state as our initial condition for the waiting list. A realistic value for the current size of the "average" waiting list is $\frac{27698}{73} \approx 380$; 27,698 was the size of the national waiting list at the end of 1994. For the NEOB, we adopt a slightly different procedure. Instead of fixing the initial waiting list size, we fix the duration of the start-up period. However, to maintain consistency between the NEOB simulation and the "average" OPO simulation, we use the latter to determine the duration of the start-up period for the former. Specifically, we simulate the "average" OPO 40 times and obtain the average time until the OPO waiting list reaches the threshold 380. We then use this number as the duration for the NEOB start-up period.

Model Outputs: The model will simulate the waiting list for ten years after the starting point. The starting point is chosen to resemble the kidney transplant environment in the United States at the beginning of 1995. The outputs of the model are reported at the end of the simulation period and are the following (all outputs are reported by age, gender, race and blood type).

- Median and mean waiting time until transplantation, and lower and upper 25% quartiles.
- Total number of people receiving transplants.
- One, two and five year survivals (in waiting list, post-tx, from entry point).
- Total number of deaths pre- and post-transplantation.
- Size of waiting list
- Deaths per eligible patients.
- Transplantations per eligible patients
- Number of life years in waiting list and post -transplant.

This concludes the description of KTM.

6.4 Discussion

In this Chapter we have presented the Kidney Transplant Model. This is a computer simulation model that was developed and validated using data from the UNOS and the USRDS. The results from the statistical analysis are particularly insightful. We have seen that the Post Transplant Graft Survival Model provides valuable insights about the graft failure process and has some important implications on organ allocation policies. Other important insights that follow from the statistical analysis are the following:

1. There is a dramatic shortage of organs for transplantation. If we ignore the demand for retransplantation, the average demand for first time transplants exceeds the average supply by 19% (i.e. the utilization rate is 1.19). The situation is more critical in the New England Organ Bank, one of the most highly congested organ banks in the nation. There, the demand exceeds the supply by 90% (utilization rate 1.90). Furthermore, the actual excess in demand is more dramatic than these numbers suggest. First, these numbers do not consider the demand for retransplantation; in the next Chapter we will see that demand for retransplantation is substantial and can stress the limited organ supply even more. Second, these numbers do not take into account future forecasts that predict a linearly increasing candidate arrival rate and a fairly constant donor arrival rate.

2. The African American population has a more diverse distribution of HLA antigens than the caucasian population; see Tables 6.5-6.7. This, combined with the disparity between the arrival rate of African American candidates (29.7% of all new candidate arrivals) and African American donors (9.6% of all new donor arrivals), generates an environment where the access to well matched kidneys is much lower for African Americans than for caucasians. This also forces the waiting time for African Americans to be longer than caucasians.

3. The mortality rates in the waiting list are lower for African Americans than caucasians; see Tables 6.18. This can further contribute to the difference in the waiting time between the two races. Recall that equation (4.52) from Chapter 4 states that the waiting time until transplantation is linearly proportional to the life expectancy and logarithmically proportional to the utilization rate.

4. Female candidates have higher ppra values than male candidates. Because ppra determines the probability that a candidate will accept a transplant offer, this implies that female candidates are more likely not to accept a transplant offer and thus wait longer than male candidates. This can partly explain the gap in the

waiting times between the two genders. Furthermore, mortality differences between the two genders further contribute to the waiting list gap.

These insights are particularly useful and provide a better understanding about the dynamics of the kidney transplant waiting list. In the next chapter, we will quantify these insights providing results from several simulation experiments.

Chapter 7

Simulation Study

In this chapter, we use the Kidney Transplant Model to simulate three allocation policies: the new UNOS policy, the First Come First Transplanted policy (FCFT), and a surrogate for the subsidized index policy of Chapter 5. The purpose of the simulation study is not to provide policy recommendations about the allocation of kidneys in the United States, but to illustrate how the methodology and tools developed in this thesis can provide qualitative and quantitative insights that can contribute to the debate about kidney allocation.

This chapter is organized as follows: Section 7.1 describes the policies tested in the simulation study and section 7.2 presents and analyzes the results from the simulation. Concluding remarks and directions for future research are presented in section 7.3.

7.1 Description of the Simulated Policies:

The new UNOS policy and the First Come First Transplanted policy were described in Chapters 3 and 5. In this section, we describe the third of the tested policies which will be referred to as the *prognostic index* policy. We start with the simplest prognostic index policy which will be denoted by PI(0.0). This policy is derived directly from the prognostic index of the proportional hazard model and allocates priorities to transplant

candidates as follows: For each incoming organ, compute the prognostic index for all eligible (i.e. blood compatible) candidates, and offer the kidney to the candidate with the lowest index; recall that the prognostic index reflects the relative risk associated with a vector of covariates, candidates with high prognostic index are expected to have shorter graft survival.

The prognostic index policy generalizes some aspects of the point policy of UNOS. Specifically, although the UNOS policy allocates priority points based on presensitization and tissue matching, the prognostic index policy allocates priority points based on a combination of all factors that are known to be associated with graft survival. However, a weakness of the priority index policy is that it is inherently unfair. For exampl;e, because the relative graft failure risk for African American candidates is 1.5227 (see Table 6.21), the prognostic index policy allocates, on average, lower priority to African American than to Caucasian candidates. To counteract this effect, we modify PI(0.0) to include a *subsidy* for African Americans; recall our discussion in Chapter 5 about the significance of a subsidized index policy. The modified prognostic index policy will be denoted by PI(α) and allocates priorities based on the following formula: Let X_i be the vector of covariates for transplant candidate i; these are the covariates used for the proportional hazard model of Chapter 6 and include characteristics of the candidate and donor (see Table 6.21). In addition, let $\hat{\beta}$ be the vector of regression coefficients in Table 6.21. The policy gives higher priority to the candidate with the smallest *subsidized prognostic index*

$$\hat{\beta}X_i - 0.4205\alpha I(\text{African American candidate}); \tag{7.1}$$

it is worth emphasizing that because 0.4205 is the regression coefficient for African American candidates in the Cox model, the prognostic index policy PI(1.0) does not discriminate based on race.

We conclude this section by establishing the relationship between the prognostic index policy PI(α) and the subsidized index policy of Chapter 5. Recall that the subsidized

index policy allocates incoming organs based on the index

$$\tilde{G}_{jk}(t) = \sum_{l=1}^{K} p_{kl}^{-}(j)\pi_l(t) - \pi_k(t) + \frac{\delta}{\sum_{k\in\mathcal{S}} \lambda_k^{+}} I(k \in \mathcal{S}); \tag{7.2}$$

Comparing (7.2) with (7.1) we observe that (7.1) approximates $\sum_{l=1}^{K} p_{kl}^{-}(j)\pi_l(t) - \pi_k(t)$ by $\hat{\beta}X_i$. To see that this is a natural approximation, recall that $\sum_{l=1}^{K} p_{kl}^{-}(j)\pi_l(t) - \pi_k(t)$ gives the expected increase in quality adjusted life years from allocating an organ of class j to a patient of class k., while $\hat{\beta}X_i$ reflects the expected increase in fuction years following organ transplantation. Therefore, $\hat{\beta}X_i$ provides a simple crude approximation for $\sum_{l=1}^{K} p_{kl}^{-}(j)\pi_l(t) - \pi_k(t)$.

7.2 Results:

We have used the Kidney Transplant Model to simulate the average OPO under seven policies: FCFT, UNOS, PI(0.0), PI(0.5), PI(1.0), PI(1.5) and PI(2.0). The results from the simulation are presented in Tables 7.1(I)-7.2(II). Table 7.1(I) describes the total number of transplantations performed over the ten year simulation period. Table 7.1(II) presents the mean waiting time until transplantation; this is a time average computed over the ten year simulation period. Table 7.1(III), presents the proportion of eligible candidates receiving transplantation over the ten year simulation period; to obtain the numbers in this Table divide the numbers in Table 7.1(II) by the total number of new candidates that arrive during the simulation period. Table 7.2(I) presents the total number of life years in the waiting list over the ten year simulation period, and Table 7.2(II) presents the total number of life years in the functioning graft compartment over the same period. In all Tables, the rows present the results by blood type, gender, race and age, while the columns present the results for the seven policies. The first column under each policy gives the average of the performance measure, where the average is computed using results from 40 independent simulation runs, and the second column gives its standard error (this can be used to compute confidence intervals).

TABLE I

	Number of Transplants											
	FCFT		UNOS		P(0.0)		P(0.5)		P(1.0)		P(1.5)		P(2.0)	
A	545.85	5.82	537.45	4.68	541.95	5.24	546.65	4.10	547.15	4.49	536.10	3.71	549.83	3.80
B	159.00	9.09	154.65	6.17	159.43	6.07	158.15	8.59	158.60	8.95	155.75	7.84	161.05	10.58
AB	39.10	3.87	37.30	3.24	40.83	4.34	39.38	4.09	37.80	4.12	38.38	4.18	41.38	4.14
O	630.65	17.12	634.80	16.99	627.90	14.70	632.35	17.91	636.53	18.29	630.48	15.39	634.45	15.16
Male	816.90	19.72	789.55	18.12	833.38	18.28	817.40	17.56	807.83	18.69	782.50	13.37	804.38	18.27
Female	557.50	14.40	574.65	13.31	536.73	11.53	559.13	10.33	570.25	15.12	568.28	12.48	582.33	12.76
Caucasian	891.18	21.35	904.10	17.27	1326.78	19.63	1194.05	16.20	1000.68	17.41	777.93	18.14	813.13	19.17
Afr. Amer.	483.23	13.98	460.10	13.66	43.33	6.55	182.48	10.85	377.40	12.07	582.85	13.62	773.58	17.38
Age: 20-30	76.68	4.70	72.93	5.73	46.53	3.29	36.73	2.97	26.63	3.32	26.80	2.90	32.03	2.77
Age: 30-40	182.58	6.80	174.65	7.14	109.75	6.70	102.25	6.13	108.75	7.02	109.65	4.86	115.00	5.12
Age: 40-50	240.05	7.74	229.63	8.42	175.73	7.77	173.65	6.20	206.20	8.25	207.83	6.27	222.10	9.82
Age: 50-60	274.98	9.16	266.18	9.96	265.38	9.82	287.50	10.20	299.10	8.77	314.68	9.07	327.03	11.30
Age: 60-70	319.20	11.47	313.15	10.15	446.03	12.87	487.85	9.32	490.40	9.89	499.80	11.96	488.13	12.19
Age: 70-80	220.50	9.80	238.40	10.00	248.80	7.22	222.20	8.07	191.05	10.37	158.10	7.86	159.85	8.73
Age: 80-90	61.43	4.52	68.83	4.52	77.85	5.26	66.35	4.15	55.95	4.54	44.73	3.51	42.78	4.94

TABLE II

	Waiting Time											
	FCFT		UNOS		P(0.0)		P(0.5)		P(1.0)		P(1.5)		P(2.0)	
A	24.08	0.49	20.74	0.58	7.24	0.38	8.92	0.37	8.99	0.30	8.94	0.32	8.58	0.40
B	37.08	0.92	33.74	1.03	13.54	0.53	15.41	0.63	17.04	0.48	15.47	0.60	15.40	0.81
AB	30.88	1.69	25.04	1.59	14.98	1.51	15.24	1.26	17.08	1.34	17.26	1.18	11.74	1.49
O	29.70	0.58	25.57	0.49	8.53	0.28	11.02	0.27	11.65	0.30	11.81	0.30	10.23	0.37
Male	28.20	0.43	23.37	0.46	8.86	0.25	10.85	0.24	11.93	0.25	11.59	0.25	10.57	0.32
Female	28.45	0.49	25.41	0.62	8.58	0.37	10.72	0.29	10.51	0.29	10.86	0.33	9.61	0.38
Caucasian	27.81	0.35	24.58	0.47	8.00	0.21	8.92	0.24	10.26	0.19	12.31	0.19	13.90	0.22
Afr. Amer.	29.58	0.59	24.51	0.58	30.67	0.69	23.24	0.41	14.29	0.30	9.82	0.41	7.24	0.39
Age: 20-30	28.22	0.41	26.04	0.45	16.59	0.53	18.08	0.53	27.35	0.59	18.78	0.63	13.02	0.58
Age: 30-40	28.50	0.51	28.16	0.60	14.29	0.54	14.89	0.53	23.05	0.38	19.55	0.45	15.58	0.70
Age: 40-50	28.87	0.74	27.00	0.74	11.61	0.74	11.50	0.65	16.88	0.51	15.40	0.35	13.40	0.43
Age: 50-60	28.77	0.88	25.90	0.71	9.32	0.87	7.76	0.53	9.87	0.47	11.08	0.58	10.10	0.48
Age: 60-70	28.16	0.54	24.08	0.68	8.50	0.38	8.51	0.31	8.58	0.30	8.98	0.32	7.34	0.35
Age: 70-80	27.89	0.70	20.89	0.71	7.00	0.27	8.16	0.22	11.44	0.22	12.32	0.28	10.12	0.51
Age: 80-90	27.09	1.09	20.52	0.95	7.34	0.50	9.00	0.53	11.87	0.54	12.02	0.69	9.43	0.91

TABLE III

	Fraction of Eligible Candidates Receiving Transplantation											
	FCFT		UNOS		P(0.0)		P(0.5)		P(1.0)		P(1.5)		P(2.0)	
A	0.91	0.03	0.89	0.02	0.90	0.03	0.91	0.02	0.91	0.02	0.89	0.02	0.91	0.02
B	0.75	0.04	0.73	0.04	0.75	0.03	0.75	0.04	0.74	0.04	0.74	0.04	0.76	0.05
AB	0.83	0.08	0.79	0.07	0.86	0.09	0.83	0.09	0.80	0.09	0.81	0.09	0.87	0.09
O	0.86	0.02	0.86	0.02	0.85	0.02	0.86	0.02	0.87	0.02	0.86	0.02	0.86	0.02
Male	0.86	0.02	0.83	0.02	0.87	0.02	0.86	0.02	0.85	0.02	0.83	0.01	0.84	0.02
Female	0.92	0.02	0.94	0.02	0.88	0.02	0.92	0.02	0.94	0.02	0.93	0.02	0.96	0.02
Caucasian	0.81	0.02	0.82	0.02	1.21	0.02	1.09	0.01	0.91	0.02	0.71	0.02	0.58	0.02
Afr. Amer.	1.04	0.03	0.98	0.03	0.09	0.01	0.39	0.02	0.81	0.03	1.25	0.02	1.60	0.04
Age: 20-30	1.01	0.06	0.97	0.08	0.62	0.04	0.49	0.04	0.36	0.04	0.36	0.04	0.43	0.04
Age: 30-40	1.21	0.05	1.16	0.05	0.73	0.04	0.68	0.04	0.72	0.05	0.72	0.03	0.76	0.03
Age: 40-50	1.24	0.04	1.19	0.04	0.91	0.04	0.90	0.03	1.06	0.04	1.07	0.03	1.15	0.05
Age: 50-60	1.07	0.04	1.04	0.04	1.03	0.04	1.12	0.04	1.17	0.03	1.23	0.04	1.27	0.04
Age: 60-70	0.77	0.03	0.75	0.02	1.07	0.03	1.17	0.02	1.18	0.02	1.20	0.03	1.17	0.03
Age: 70-80	0.61	0.03	0.66	0.03	0.69	0.02	0.62	0.02	0.53	0.03	0.44	0.02	0.44	0.02
Age: 80-90	0.57	0.04	0.64	0.04	0.72	0.05	0.61	0.04	0.52	0.04	0.41	0.03	0.40	0.05

Table 7.1: Results from the simulation study: number of transplants performed, mean waiting time until transplantations, and fraction of performed transplantations.

Table I

	Total Number of Life Years in the Waiting List													
	FCFT		UNOS		PI(0.0)		PI(0.5)		PI(1.0)		PI(1.5)		PI(2.0)	
A	17548.70	264.46	17440.88	255.32	17376.88	249.62	16996.83	235.99	16566.33	251.34	16900.03	193.96	16337.88	202.98
B	8144.55	162.40	8554.23	162.20	8327.58	139.96	8308.50	158.13	7920.23	142.78	8232.25	141.23	8098.40	164.13
AB	1728.13	65.15	1724.93	68.49	1696.15	68.53	1662.30	73.54	1728.85	77.70	1707.35	72.17	1598.50	71.82
O	25472.65	266.92	25636.45	233.03	25705.00	259.34	24947.55	315.74	24357.75	261.36	24409.78	269.45	24566.80	273.84
Male	31508.15	228.85	32519.23	246.29	31436.15	213.02	31095.95	264.57	30645.50	233.60	31089.65	248.59	30652.35	317.80
Female	21385.88	208.84	20837.15	196.68	21750.45	153.31	20819.23	169.07	19927.65	234.72	20159.75	235.85	19749.23	166.93
Caucasian	34755.75	277.25	34240.90	263.68	23911.35	240.42	26474.90	276.98	31025.18	275.90	36730.25	270.83	40060.10	310.22
Afr. Amer.	18138.28	179.76	19115.48	180.78	29284.25	189.30	25440.28	191.11	19547.98	201.83	14519.15	190.83	9741.48	156.08
Age: 20-30	2624.38	44.12	2677.35	56.36	3091.50	75.18	3402.68	54.91	3716.50	68.97	3815.93	67.63	3705.38	79.77
Age: 30-40	6215.18	95.49	6375.18	86.76	7637.98	129.49	7730.70	102.86	7775.53	108.00	7955.53	114.24	7858.20	99.76
Age: 40-50	8390.83	103.21	8702.50	105.72	10151.35	111.04	10112.80	128.43	8978.25	112.11	9140.95	103.56	8800.65	122.41
Age: 50-60	9944.50	95.72	10298.50	112.84	10169.38	134.64	9199.10	129.63	8476.00	114.81	8127.78	114.00	7950.50	120.40
Age: 60-70	12790.50	114.68	12735.00	131.32	9834.73	86.30	8731.53	98.33	8225.38	102.83	7938.20	89.96	8150.93	108.99
Age: 70-80	9964.88	85.36	9796.43	99.45	9461.43	93.25	9814.85	91.68	10341.83	105.11	10972.25	112.76	10881.70	120.05
Age: 80-90	2933.68	41.77	2771.43	43.68	2849.25	52.94	2923.65	47.99	3059.68	51.76	3292.78	51.96	3246.23	58.37
Total	52864.03	380.41	53356.38	392.59	53195.60	294.81	51915.18	371.40	50573.15	379.80	51249.40	380.32	50601.58	397.81

Table II

	Total Number of Life Years in the Functioning Graft Compartment													
	FCFT		UNOS		PI(0.0)		PI(0.5)		PI(1.0)		PI(1.5)		PI(2.0)	
A	25732.43	1542.40	25880.88	1696.85	27460.88	1548.86	27760.55	1172.05	27886.75	1560.49	26962.58	1391.43	26836.43	1286.79
B	6988.83	835.38	6580.50	547.83	7442.83	645.91	7405.90	939.55	7287.00	752.10	7196.33	869.64	6929.15	1033.39
AB	1778.80	411.86	1691.60	329.10	1991.70	441.10	1861.83	384.31	1778.45	379.59	1745.68	397.74	1929.75	361.27
O	28674.43	1569.15	29500.55	1447.94	31345.85	1317.14	31460.15	1642.09	31262.88	1358.54	30517.08	1730.07	29883.80	1463.09
Male	36977.75	1603.37	36843.60	1664.19	40402.90	1496.06	39865.58	1716.09	39566.63	1662.08	38256.43	1269.20	37822.80	1535.21
Female	26196.53	1469.65	26809.83	1226.09	27838.35	1149.46	28622.85	1137.60	28668.45	1521.38	28165.23	1533.99	27756.33	1127.32
Caucasian	44734.85	1740.03	45471.28	1763.07	52044.10	1754.69	57973.25	1740.77	51582.18	1735.22	43399.98	1593.55	37214.43	1973.95
Afr. Amer.	18439.43	1082.67	17682.25	883.66	6197.15	683.18	10515.18	993.66	16652.90	988.86	23021.68	1269.30	28364.70	1238.45
Age: 20-30	2294.05	327.00	2154.45	355.30	1898.45	315.86	1527.00	247.89	1154.90	263.88	1180.95	249.23	1408.23	233.36
Age: 30-40	7310.03	642.47	7194.00	658.19	6008.73	512.20	5700.33	522.81	5581.73	664.73	5261.83	503.92	5344.95	473.96
Age: 40-50	10665.88	677.06	10848.38	710.91	9493.30	701.00	9618.23	652.13	10543.95	778.42	10480.65	598.70	10352.75	776.70
Age: 50-60	12510.68	735.71	12449.98	904.57	13242.75	801.50	13867.33	762.84	14428.23	897.62	14529.35	846.40	14568.83	1037.29
Age: 60-70	14534.50	939.87	14603.05	847.30	19514.43	1030.48	20870.20	870.02	21050.53	838.16	21032.78	1044.23	20231.15	999.19
Age: 70-80	11841.20	916.25	12424.95	715.99	14151.48	751.18	13235.45	829.83	12144.05	930.72	11037.93	728.04	10860.80	814.09
Age: 80-90	3717.95	411.55	3978.73	401.97	3932.13	391.91	3869.90	356.51	3333.70	395.46	2918.18	374.53	2814.43	379.96
Total	63174.28	2240.45	63653.53	2119.96	68241.25	1936.91	68488.43	1893.55	68235.08	2058.07	66421.65	2088.13	65579.13	2114.07

Table 7.2: Results from the simulation study: Life years in the waiting list and in the functioning graft compartment.

Several important observations are in place here:

1. From studying the First Come First Transplanted Policy, we can partly understand the multifaceted nature of equity. Let us first start our discussion, by stating the "axiom" that the FCFT policy is inherently fair (or equitable). This intuition is certainly confirmed by the results in Table 7.1(II). We observe that under the FCFT the mean waiting time until transplantation appears to be independent of the candidate's age, gender and race; there is only a small difference (1.97 months) in the waiting time between Caucasian and African American. The difference in the waiting times can be attributed to the higher fraction of presensitized African American candidates. However, the results in Table 7.1(III) appear to contradict the axiom that FCFT is fair. Specifically, Table 7.1(III) presents the ratio of transplantations performed to transplantations requested. Intuitively, this quantity gives the number of transplants each candidate is expected to receive. For example, Table 7.1(III) suggests that each Caucasian candidate expects to receive 0.81 transplants under the FCFT policy, whereas each African American candidate expects to receive 1.04 transplantation; values greater than one indicate that candidates receive second transplants. Therefore, although when we compare waiting times we conclude that the FCFT is inherently fair, when we compare fractions of performed transplants we deduce that Caucasians receive on average less transplantations than African American. This implies that the hard-to-quantify concept of equity can be completely described by a multidimensional performance measure. Policies that appear to be fair when analyzed using one dimension of equity, they may well turn out to be inherently unfair under a second, equally plausible, equity measure. In fact, achieving the right balance between the different dimensions of equity is a nontrivial task. This point will be further amplified in the remainder of our discussion.

2. The comparison between FCFT and UNOS is also particularly insightful. As expected, the UNOS policy is clinically more efficient than the FCFT policy (we say

that a policy is clinically more efficient than a second policy if it achieves a higher number of life years with functioning organs): By using UNOS, we gain 462.35 function years in the waiting list. Although this number is not statistically significant let us attempt to put it in perspective by comparing it to a well established benchmark. This benchmark is the estimated number of life years gained by the introduction of cyclosporine which is approximately 9000 function years per 100,000 patient years; see Wujciak and Opelz (1993) [26]By comparing this improvement to the improvement from using UNOS, which is 462.35 function years per 63174.28 patient years, or 731.86 function years per 100,000 patient years, we conclude that improvement from the UNOS policy is approximately 7.3% of the improvement from using cyclosporine.

3. It is also worth emphasizing that under the new UNOS policy the mean waiting time appears to be independent of race; a highly desirable outcome from the point of view of UNOS. However, despite that, the UNOS policy generates some differences in the waiting times between the different age groups and between genders. In particular, older and male patients experience much shorter waiting times. This effect can be attributed to differences in the mortality rates between the different age groups and genders (see Table 6.20); older and male candidates tend to have higher mortality rates.

4. Let us now compare the UNOS policy to the prognostic index policy. As expected, several function years are gained by using the PI(0.0) policy . The exact magnitude of the improvement is $\left(\frac{68241.25-63174.28}{63174.28}\right) 100000 = 8020.6$ function years per 100,000 patient years. This is of the same order of magnitude as the improvement from cyclosporine. However, the PI(0.0) policy is unacceptable, both from an ethical and a legal point of view. Specifically, the PI(0.0) generates a huge access gap between the two races: Although Caucasians receive, on average, 1.21 transplants, African Americans receive only 0.09 transplants! Moreover, the mean waiting time

until transplantation is 8.00 months for Caucasians, but 30.67 months for African Americans.

5. The introduction of the race subsidy in the prognostic index policy counteracts the racial bias generated by PI(0.0). This is best seen by comparing PI(1.0) to PI(0.0). As expected, under PI(1.0) Caucasians and African Americans enjoy comparable access to transplantation: The mean waiting time is 10.25 months for Caucasians and 14.29 months for African Americans. Moreover, African Americans receive, on average, 0.81 transplants, while Caucasians receive 0.91 transplants. Although this is not ideal, it represents a considerable improvement over the PI(0.0) policy. Moreover, this improvement is achieved at a negligible loss of clinical efficiency: A total number of $\frac{68235.08-63174.28}{63174.28}100000 = 8010.9$ life years are now gained compared to the FCFT policy. However, to further improve the equity ratio between the two races, it is necessary to incrementally sacrifice more clinical efficiency: Comparing PI(1.5) to FCFT, we see that life years gained is now only $\frac{66421.65-63174.28}{63174.28}100000 = 5140.3$.

6. The mean waiting time until transplantation is significantly reduced under the prognostic index policy as compared to the FCFT policy and the UNOS policy.

The trade-off between clinical efficiency and equity is better illustrated through several trade-off diagrams (Figures 7-1 - 7-4); the vertical axes in the trade-off diagrams give the Quality Adjusted Life Months (QALMs) per eligible patient, the horizontal axis varies between figures; to compute the QALMs we assume that one year in the waiting list is equivalent to 0.62 quality adjusted years, and one year with a functioning graft is equivalent to 0.75 quality adjusted years (see Evans, 1993 [36]). Figure 7-1 presents the trade-off diagram between efficiency and access to transplantation by race. The horizontal axis gives the ratio of the fraction of eligible Caucasian candidates receiving transplantation over the fraction of eligible African American candidates receiving transplantations; panel (a) present the results in a linear scale, and panel (b) in a logarithmic

scale. The solid line gives the "efficient" frontier that is traced by the prognostic index policies. We observe that the frontier clearly dominates both the UNOS and the FCFT policy. Furthermore, it is possible to adopt a policy that achieves the same level of access to transplantation between the two races as the UNOS policy but at a higher level of clinical efficiency. The possible improvement in efficiency is estimated to be approximately 2067 function years per 100000 patient years. Figure 7-2, presents the trade-off between clinical efficiency and equity in waiting time by race; the horizontal axis gives the absolute difference between the mean waiting time until transplantation by race. We observe, that there is no policy that clearly dominates the UNOS policy. In fact, the UNOS policy appears to be the only policy that achieves inequity close to the ideal value of zero. By contrast, the prognostic index policies suffer from large differences in waiting times and PI(0.0) appears to be extremely inequitable. It is worth emphasizing that the FCFT policy is dominated by the efficient frontier. This implies that there exists a policy that adopts a weighted combination of PI(1.0) and UNOS which clearly dominates the FCFT policy; the dominant policy gains approximately 1950 function years per 100,000 patients years. Such a policy can be derived using the dynamic index policy of Chapter 5.

So far we have focused on the trade-off between clinical efficiency and equity, or fairness, between races. However, the results in Tables 7.1-7.2 suggest that different policies can also generate inequity between genders or age groups. To investigate these two forms of equity, we plot the trade-off diagrams in Figures 7-3- 7-4.

Figure 7-3 presents inequity between genders. The horizontal axis in panel (a) gives the ratio of the fraction of eligible Male candidates receiving transplantation over the fraction of eligible Female candidates receiving transplantations. Panel (b) gives the absolute difference between the mean waiting time until transplantation by gender. Panel (a) show that UNOS is dominated by PI(0.0); both more efficient and equitable. The potential improvement from adopting the dominant policy is approximately 8020.6 function years per 100,000 patient years. Panel (b) also shows that UNOS is dominated by

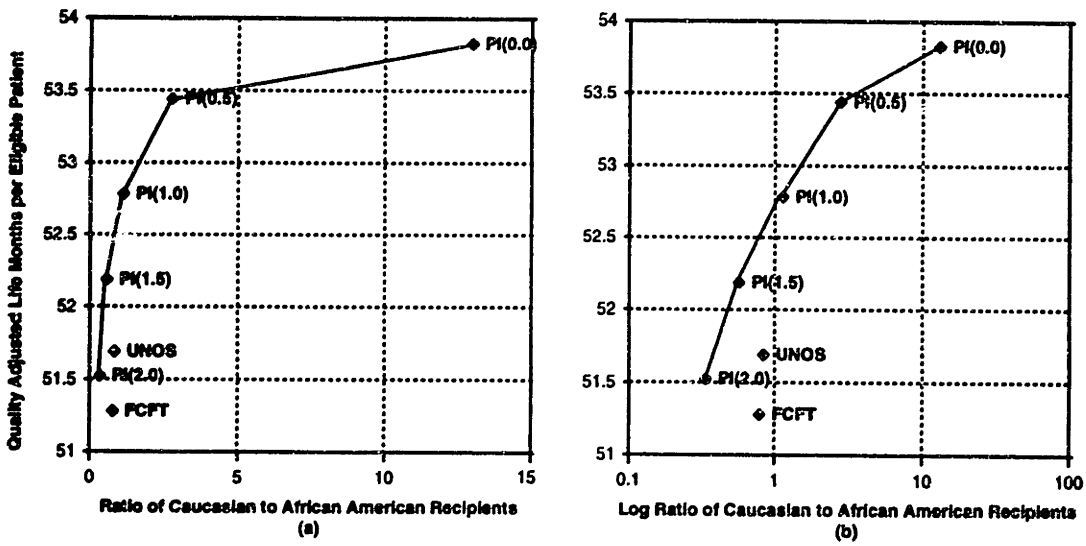

Figure 7-1: Efficiency versus fairnes: Here fairness is measured by the fraction of eligible Caucasian candidates that received transplantation during the simulation period over the fraction of eligible African American candidates that received transplantations

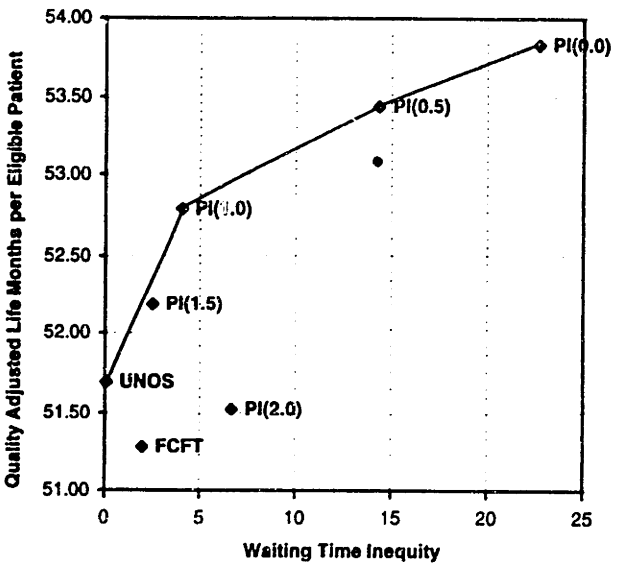

Figure 7-2: Efficiency versus fairnes: Here fairness is measured by the absolute difference between the mean waiting time for Caucasian and African American candidates.

the prognostic index policies.

Figure 7-4 presents inequity between age groups. The horizontal axis in panel (a) gives the square root of the pairwise square differences between the fraction of eligible candidates receiving transplantation by age group. The horizontal axis in panel (b) gives the square root of the pairwise square difference between the mean waiting time until transplantation by age group. Both panels show that UNOS fails to achieve equity between age groups. By contrast, there exists policies that are obtained by combining the existing prognostic index policy with the FCFT policy that uniformly dominate the UNOS policy. Specifically, examining plot (a) we conclude that there exists a weighted combination of PI(0.0) and PI(0.5) that dominates UNOS. This policy improves clinical efficiency by approximately 5158 function years per 100,000 patient years. Similarly, examining panel (b) we conclude that there exists a weighted combination of PI(0.0) and FCFT that dominates UNOS. The improvement from adopting the dominant policy is approximately 2837 function years per 100000 patient years. It is worth emphasizing

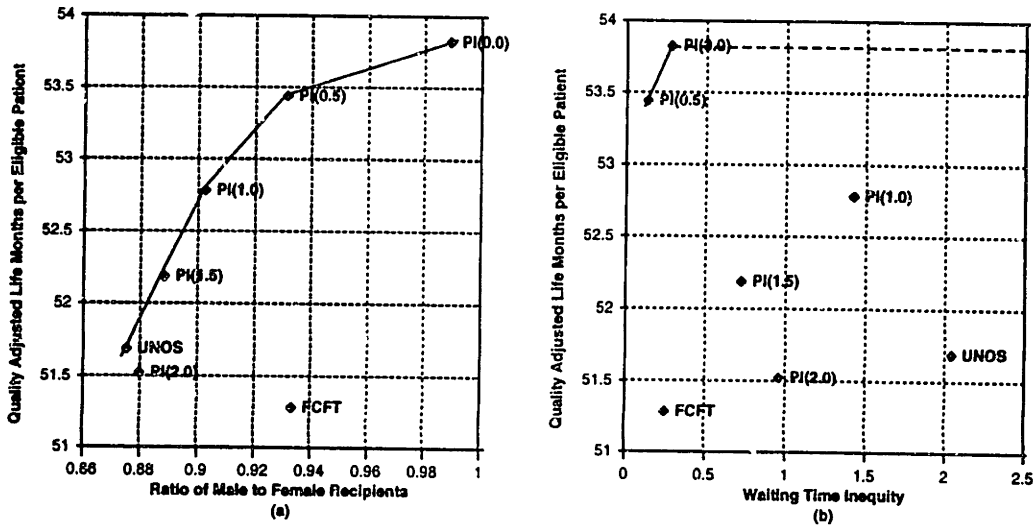

Figure 7-3: Clincial efficiency versur fairness between genders. The horizontal axis in panel (a) gives the ratio between the fraction of eligible female candidates that received transplantation over the fraction of eligible male candidates that received transplantation; policies closer to 1 are more fair. The vertical axis in panel (b) give the absolute difference between the mean waiting time until transplantation for the two genders; values closer to zero reflect a more fair policy.

here that panel (a) does not confirm the existence of a trade-off between age equity and clinical effectiveness; rather, the panel suggest that the most effective policy is also the most equitable once. However, this results should not be viewed in isolation, but in combination with the other trade-off diagrams which verify the presence of trade-off.

In conclusion, the trade-off diagrams demonstrate that although UNOS minimizes the difference in the waiting time between African American and Caucasian candidates it does so at the expense of clinical efficiency and equity between age groups and genders. Our analysis, suggests that a policy that combines the prognostic index policy with a time dependent priority policy is expected to better balance the several dimension of equity and clinical efficiency. A likely candidate for such a policy is the dynamic index policy of Chapter 5.

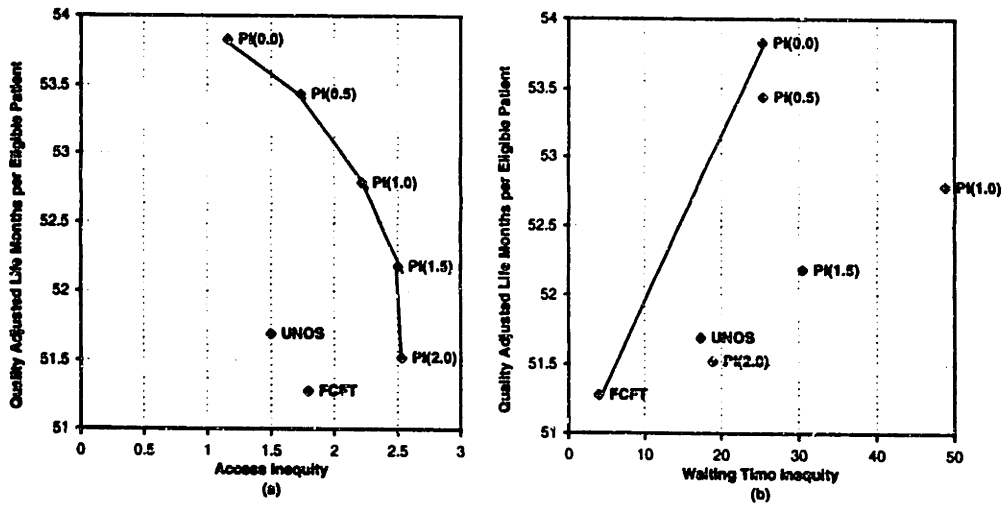

Figure 7-4: Efficiency versus fairness between age groups. The vertical axis in panel (a) gives the pairwise square difference between the fraction of eligible candidates receiving transplantation in the 7 age groups ; the vertical axis in panel (b) gives the pairwise square difference between the waiting times until transplantation in the 7 age groups (values closer to 0 reflect a more fair policy).

To conclude this section, we demonstrate how the results from the simulation study can be used to project the future size of the transplant waiting list under the UNOS policy. Figure 7-5 gives the average size of the waiting list over time; the average is computed from 40 independent simulation runs. The initial size of the waiting list is 383.09 ± 2.03, and the final size of the waiting list is 513.90 ± 10.35. Therefore, the waiting list is expected to increase by approximately 34.15% over the next ten years. The output from the simulation study can also be used to study the trend in the waiting time until transplantation. Although we do not show the detailed statistics, the mean waiting time until transplantation for all transplants performed in a period of twelve months increases from 25.17 at the end of the first year to 27.15 at the end of the last year.

7.3 Conclusions:

In this part of the thesis we have analyzed the problem of allocating kidneys to patients on the transplant waiting list. Our objective was twofold: First, to understand the notion of fairness and its relation to medical benefit, and second, to propose organ allocation policies that balance the two conflicting objectives of medical benefit and equity. To do that, we have developed two mathematical models: A queueing model with reneging that provides insights about the performance of the transplant waiting list, and a fluid model that integrates the kidney waiting list with the process of organ rejection. Our analysis of the reneging model motivates several mathematical metrics that reflect equity. We embedded these metrics into the fluid model to formulate three optimal control problems. In all problems, the objective is to find the optimal allocation policy. Using several approximations we have developed three heuristic allocation policies: the efficiency policy, the subsidized index policy, and the dynamic priority policy.

To supplement our mathematical models we have also developed a state of the art computer simulation model. Although the conceptual structure of the simulation model is the same as that of the fluid model, the simulation model is more complex and cap-

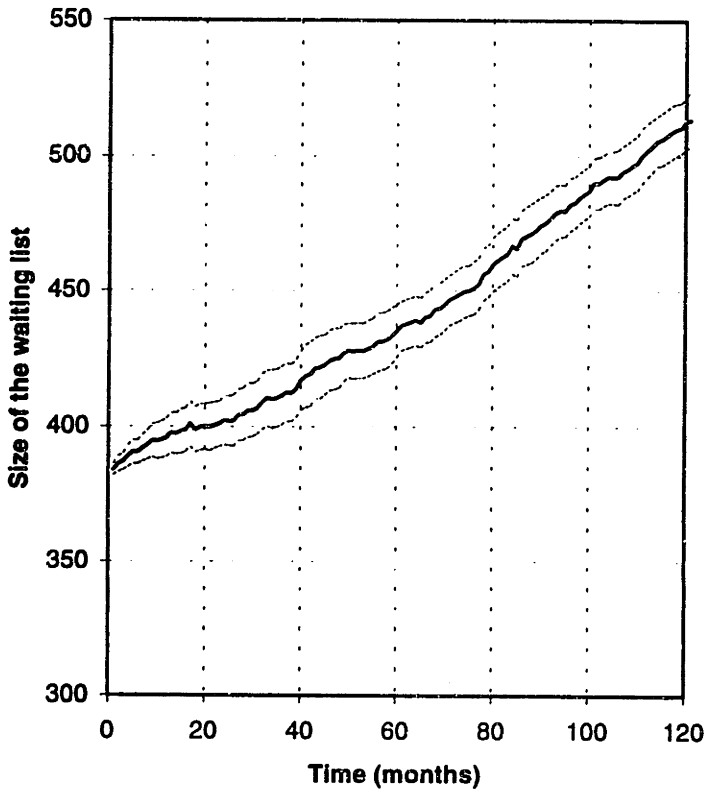

Figure 7-5: Average size of the waiting list over time; the waiting list is simulated under the new UNOS policy. The solid line gives the average over 40 independet simulation runs, and the dotted lines give the lower and upper 95% confidence intervals.

tures aspects of the organ allocation process that cannot be captured by the analytically tractable fluid model.

To demonstrate the derived methodology we have used the computer simulation model to simulate a hypothetical waiting list under several policies, including the FCFT policy, the new UNOS policy and several subsidized index policies. Our results show that these policies trade-off medical benefit and equity in a policy-specific manner. For example, although under the UNOS policy Caucasian and African American candidates have approximately the same mean waiting time until transplantation, male and female candidates have different mean waiting time. Analysis of the trade-off diagrams suggest possible ways for improving the current policy.

In general, the results from this thesis demonstrate that a combination of an analytical fluid model and a computer simulation model provide a powerful tool for ananlyzing the performance of different kidney allocation policies. However, the problem of kidney allocation is complex and multifaceted. Therefore, the results from this thesis are not expected to resolve the problem. Rather, our results provide a methodological way for comparing different allocation policies. Using this methodology, we can provide quantitative answers to some important policy questions. Answers to these questions can contribute to a more informed decision making.

Bibliography

[1] R. D. M. Allen, and J. R. Chapman. (1994). A Manual of Renal Transplantation. Edward Arnold, London, 1994.

[2] P. I. Terasaki, et al. (1983). Improving Success Rates of Kidney Transplantation. *Journal of American Medical Association,* **250: 1065-1088.**

[3] D. W. Ghjertson, et al. (1991). National Allocation of Cadaveric Kidneys by HLA Matching. *New England Journal of Medicine,* **324: 1032-1036.**

[4] D. Mathieu. (1988). *Organ Substitution Technology: Key Issues.* In Organ Substitution Technology: Ethical, Legal, and Public Policy Issues, eds. D. Mathieu. Westview Press, Boulder.

[5] M. A. Baily. (1988). *Economic Issues in Organ Substitution Technology.* In Organ Substitution Technology: Ethical, Legal, and Public Policy Issues, eds. D. Mathieu. Westview Press, Boulder.

[6] F. P. Sanfilippo, et al. (1992). Factors Affecting the Waiting Time of Cadaveric Kidney Transplant Candidates in the United States. *Journal of American Medical Association,* **267: 247-252.**

[7] D. S. Gaylin, et al. (1993). The Impact of Comorbid and Sociodemographic Factors on Access to Renal Transplantation. *Journal of American Medical Association,* **269: 603-608.**

[8] R. S. Gaston, et al. (1993). Racial Equity in Renal Transplantation. *Journal of American Medical Association*, **270: 1352-1356.**

[9] B. L. Kasiske et al. (1991). The Effect of Race on Access and Outcome in Transplantation. *New England Journal of Medicine*, **324: 302-307.**

[10] Office of Inspector General. (1991). *The Distribution of Organs for Transplantations: Expectations and Practices.* US Dept. of Health and Human Services publication OE1-01-89-00550.

[11] M. C. Weinstein, et al. (1987). Forecasting Coronary Heart Disease Incidence, Mortality and Cost: The Coronary Heart Disease Policy Model. *American Journal of Public Health*, **77: 1417-1426.**

[12] M. C. Weinstein. (1989). Methodological Issues in Policy Modeling for Cardiovascular Disease. *Journal of the American College of Cardiologists*, **14: 38A-43A.**

[13] Y. Yuan, et al. (1994). Development of a Central Matching System for the Allocation of Cadaveric Kidneys: A Simulation of Clinical Effectiveness. *Medical Decision Making.* **14: 124-136.**

[14] R. W. Evans. (1993). Organ Procurement Expenditures and the Role of Financial Incentives. *Journal of American Medical Association*, **269: 3113-3118.**

[15] P.L.Barton and J. Kallich. (1994). The Rationing of Kidneys for Transplantation: US Distribution Models for Cadaveric Renal Organs. *Transplantation Proceedings*, **26(6): 3685-3692.**

[16] S. Takemoto, et al. (1994). Equitable Allocation of HLA Compatible Kidneys for Local Pools and For Minorities. *New England Journal of Medicine*, **331(12): 760-764.**

[17] P.J.Held, et al. (1994). The Impact of HLA Mismatches on the Survival of First Cadaveric Kidney Transplants. *New England Journal of Medicine*, **331(12): 765-770.**

[18] G.M. Chertow, et al. (1996). Antigen Independent Determinants of Cadaveric Renal Allograft Failure. Working paper, Harvard Medical School.

[19] B.M. Brenner, et al. (1992). In Renal Transplantation, One Size May Not Fit All, *Journal of American Society of Nephrology*, **3: 162-169.**

[20] P.I. Terasaki, et al. (1994). The Hyperfiltration Hypothesis in Human Renal Transplantation, *Transplantation*, **57: 1450-1454.**

[21] V. Miles, et al. (1996). The Effect of Kidney Size on Cadaveric Renal Allograft Outcome, *Transplantation*, **61: 894-897.**

[22] R.S.Gaston, et al. (1993). Impact of Donor/Recipient Size Matching on Outcomes in Renal Transplantation. *Transplantation*, **61: 383-388.**

[23] R.S.N. Kalil, et al. (1991). Patients with Low Income Have Reduced Renal Allograft Survival. *American Journal of Kidney Diseases*, **XX(1): 63-69.**

[24] M.D. Ellison, et al. (1993). Blacks and Whites on the UNOS Renal Waiting List: Waiting Times and Patient Demographics Compared. *Transplantation Proceedings*, **25(4): 2462-2466.**

[25] T.E. Starzl et al. (1987). A Multifactorial System for Equitable Selection of Cadaver Kidney Recipients, *Journal of American Medical Association*, **257(22); 3073-3075.**

[26] T. Wujciak, and G. Opelz (1993). Computer Analysis of Cadaver Kidney Allocation Procedures, *Transplantation*, **55: 516-521.**

[27] G. Opelz and T. Wujciak (1995). Cadaveric Kidneys Should Be Allocated According to HLA Matching, *Transplantation Proceedings*, **27(1): 93-99.**

[28] F. Sanfilippo. (1993). Organ Allocation: Current Problems and Future Issues. *Transplantation Proceeding*, **25(4): 2467.**

[29] R. Righter. (1989). A Resource Allocation Problem in a Random Environment, *Operations Research*, **37: 329-338.**

[30] I. David and U. Yechiali (1985). A Time Dependent Stopping Problem With Application to Live Organ Transplants. *Operations Research*, **33: 491-504.**

[31] I. David and U. Yechiali (1995). One Attribute Sequential Assignment Match Processes in Discrete Time. *Operations Research*, **43: 879-884.**

[32] B. Shah, et al. (1988). Current Experience With Renal Transplantation in Older Patients. *American Journal of Kidney Diseases*, **12: 516-523.**

[33] UNOS. (1991). *The Feasibility of Allocating Organs on the Basis of a Single National List*, UNOS, VA

[34] UNOS (1995). *UNOS Policies for Organ Allocation.* UNOS, VA.

[35] R. W. Evans, et al. (1993). Is Retransplantation Cost Effective? *Transplantation Proceedings*, **25: 1694-1696.**

[36] R. W. Evans, et al. (1993). A Cost-Outcome Analysis of Retransplantation: The Need for Accountability. *Transplantation Reviews*, **7: 1-13.**

[37] G.F.Carrier, M. Krook, C.E.Pearson, (1983). *Functions of Complex Variables.* New York: Hod Books.

[38] E.G.Coffman, A.A. Puhalskii, M.I.Reiman, and P.E.Wright, (1994). Processor shared buffers with reneging. *Performance Evaluation*, 19: 25-46.

[39] A. Federgruen, and H. Groenvelt, (1988). $M/G/c$ queueing systems with multiple customer classes. *Managmenent Science*, 34: 1121-1138.

[40] M.W. Hirsch, and S. Smale, (1974). *Differential Equations, Dynamical Systems, and Linear Algebra.* San Diego: Academic Press.

[41] L. Kleinrock, (1975). *Queueing Systems, Volume 1,* New York: John Wiley and Sons.

[42] J.D.C. Little (1961). A proof of the queueing formula $L = \lambda W$. *Operations Research,* 9: 383-387.

[43] X-D. Luo, (1995). *Continuous Linear Programming: Theory, Algorithms and Applications.* Unpublished Ph.D Thesis, Operations Research Center, MIT.

[44] S. P. Sethi, (1981). *Optimal Control Theory,* Boston: Martinus Nijhoff Publishing.

[45] J. G. Shantikumar and D. D. Yao (1992). Mutliclass queueing systems: Polymatroidal structure and optimal scheduling control, *Operations Research,* 40: 293-299.

[46] UNOS. (1991). *The Feasibility of Allocating Organs on the Basis of a Single National List,* UNOS, VA

[47] L.M. Wein, S.A. Zenios and M. Nowak, (1996). Dynamic Multidrug Treatment for HIV: A Control Theoretic Approach. Working Paper, Sloan School of Management.

[48] UNOS (1994). *UNOS Public Use Data Tape Documentation,* United Network of Organ Sharing, Richmond, VA.

[49] UNOS (1995). *1995 Annual Report of the U.S. Scientific Registry for Transplant Recipients and the Organ Procurement and Transplant Data: 1988-94.* UNOS, Richmond, VA.

[50] USRDS (1995). *1995 Annual Report.* web address: `http://www.med.umich/usrds/`

[51] Barnes, B.A. and Miettinen, O.S., (1972). The search for an HLA and ABO compatible cadaver organ for transplantation. *Transplantation,* **13: 592-598.**

[52] Suthanthiran, M. and Strom, T.B., (1994). Renal Transplantation. *The New England Journal of Medicine*, **331(6): 365-376.**

[53] Kauffman, H.M. et al., 1996. Facing Organ Shortage, Should We Revisit Organ Allocation? *Transplantation Proceedings*, **28(1): 34-35.**

[54] Cox, D.R. and Oakes, D., (1984). *Analysis of Survival Data*, London: Chapman and Hall.

[55] Fleming, T.R. and Harrington, D.P. (1991). *Counting Process and Survival Analysis*, New York: John Wiley and Sons.

[56] Van Houwelingen, H.C., and Thorogood, J. (1995). Construction, Validation and Updating of a Prognostic Model for Kidney Graft Survival. *Statistics in Medicine*, **14: 1999-2008.**