Title
Stem Cell Donor Matching for Patients of Mixed Race

Journal
B.E. Journal of Economic Analysis, 12(1)

Authors
Bergstrom, Ted C
Garratt, Rod
Sheehan-Connor, Damien

Publication Date
2012

STEM CELL DONOR MATCHING FOR PATIENTS OF MIXED RACE

Theodore C. Bergstrom*
Rodney J. Garratt†
Damien F. Sheehan-Connor‡§

April 28, 2011

Abstract

The plight of mixed-race leukemia patients unable to find stem cell donors with matching immunity types has received much media attention. Because of small samples, direct estimation of the distribution of immunity types for persons of mixed race has not been possible. However, because the alleles that control immunity type are located close together on a single chromosome, it is possible to estimate the distribution of types for mixed races, by using existing estimates of haplotype distributions for single races. We provide such estimates and calculate probabilities that persons of specific mixed-race combinations can find a match in the existing registry of potential donors. We also estimate probabilities that adding mixed-race donors to the registry will result in a life-saving match for a patient in need of a transplant and we compare the benefits and costs of adding new registrants of specified race or mixed race to the registry.

*Department of Economics, University of California Santa Barbara
†Department of Economics, University of California Santa Barbara
‡Department of Economics, Wesleyan University
§This material is based upon work supported by the National Science Foundation under Grant No. 0851357. We are grateful to Martin Maiers of the NMDP and Thomas Bart of the Swiss Bone Marrow Donor Registry for helpful comments.

1 Introduction

Patients with leukemia and other blood diseases stand a good chance of recovery and a return to normal life if they receive a stem cell transplant from a living donor. In the absence of a transplant, their survival prospects are grim. For a transplant to be successful, the immunity responses of the donor and recipient must be controlled by matching human leukocyte antigen (HLA) systems. Finding a match is often difficult because the distribution of HLA types in the human population is extremely diffuse. Type distributions differ between races. While the distributions of different races have some overlap, the probability that two randomly selected individuals will match is higher if they are of the same race. Approximately half of the population of European descent belong to types with frequency less than one in one hundred thousand, while twenty percent belong to types with a frequency of less than one in a million. The distribution of types for persons of Asian and African descent is even more diffuse.

Registries of potential stem cell donors have been established to deal with this problem. The largest of these registries is the National Marrow Donor Program (NMDP) which includes more than 7 million registrants from the United States, Germany, Scandinavia, the Netherlands, and Israel. Because the number of persons of European ancestry in the NMDP registry is more than 10 times that of the other races, whites are much more likely to find a match than persons of other races.

New registrants contribute a saliva sample from which their HLA type is determined and state their willingness to donate to any patient for whom they are a match. Stem cells may be transferred from donors either by means of bone marrow transplants or by peripheral blood stem cell (PBSC) transfer. Both procedures involve temporary discomfort for the donor, but normally donors return to full health within two to three weeks.

The plight of multiracial leukemia patients who are unable to find matching stem cell donors has received much attention in the news media. These accounts present wrenching stories of multi-racial individuals who have been unable to find a match and claim that for persons of mixed race, chances of finding a match are slim. They do not, however, offer supporting statistical evidence. As a recent *Time Magazine* [35] story reports,

> "It's difficult to ascertain the exact chances of finding a match for a mixed race person because the different combinations have different success rates, and the U.S.-based National Marrow Donor Program (NMDP) ... does not have statistics on the success rates of mixed race patients."

Several studies, [4], [14], [6], [32], [1], [12], have estimated matching probabilities for persons of single races. But as far as we know, there are no published estimates of matching probabilities for persons of mixed race. Because the relevant sample sizes are small, direct estimation of the distribution of types for mixed-race populations is not reliable. We will demonstrate, however, that the distribution of HLA types for persons of mixed race can be estimated indirectly, using existing estimates of the distribution of haplotypes in single-race populations. To do so, we apply simple principles of probability and the combinatorics of diploid reproduction. These estimates are made possible by the fact that the alleles that determine HLA compatibility are located in

close proximity on the same chromosome, and conesequently genetic crossover is rare.

2 Methods

2.1 Genetic Background and Data Sources

The human leukocyte antigen (HLA) system is a complex of genes related to the immune system. The immune system uses the HLA genes to differentiate self cells and non-self cells, a task that is central to defense against disease. Cells that do not match the body's own HLA type are treated as invaders. In order for a stem cell transplant to be successful, the HLA genes of the donor and recipient must be a close match. Human HLA types are determined by the pairs of alleles found in a small number of genetic loci located on the same chromosome. The traditional medical standard for a suitable match for stem cell transplant has been that donor and recipient share the same six alleles in three specific loci, known as HLA-A, HLA-B, and HLA-DRB1. [1] The combination of six alleles that an individual has in these three loci is referred to as his or her *phenotype*.

The critical six alleles are inherited in the form of two strings of three, one from each parent. These three-allele strings are known as *haplotypes*. The haplotype that each parent passes to a child is randomly chosen from the parent's two haplotypes. Because the HLA alleles are located in close proximity on the same chromosome, recombination between these loci is very infrequent. Therefore, with rare exceptions, HLA haplotypes remain intact from generation to generation.

The largest available source of data on the distribution of HLA types is the NMDP registry. The NMDP records the self-reported race and the HLA type of each registrant. In a 1997 paper, M. Mori *et al* [22] used a sample of about 400,000 registrants who had been typed at the HLA-A, HLA-B, and DRB-1 loci in order to estimate the distribution of haplotypes in several racial subgroups of the U.S. population. More recently, Kollman *et al* [21] estimated haplotype using a larger sample of about two million registrants, including 1.2 million European-Americans, 250,000 Asian-Americans, 280,000 African-Americans, and 320,000 Hispanics.

Although the samples used in the Mori and Kollman studies are large, they are not nearly large enough to provide good direct estimates of the distribution of relatively rare types. Since there are tens of millions of possible HLA phenotypes, many of the rare types will not appear even in a sample of 2 million. But the mechanics of diploid genetics make it possible, under reasonable assumptions about mating patterns, to use the observed distribution of phenotypes to construct maximum-likelihood estimates of the distribution of HLA haplotypes for each race. Since the number of possible haplotypes at the three-locus level is "only" in the tens of thousands, the data available from the NMDP registry is sufficient to yield reasonably accurate estimates of haplotype distributions. The Mori and Kollman studies followed this strategy and published tables of estimated frequency distributions of haplotypes in each racial group.

[1] This standard is evolving. Currently, clinicians also attempt to match donors and patients at the loci HLA-C and HLA-DPB1, whenever possible.

2.2 Estimating Phenotype Distributions

We use the Kollman estimates [21] of HLA haplotype frequencies within each race to construct corresponding estimates of the distribution of HLA phenotypes in the population for persons whose parents are of any two specified races. To accomplish this, we assume that matching is random with respect to HLA type. Thus the probability distribution of HLA types for persons of any specified race is independent of the HLA type and the race of their mating partner.

With this assumption, the combinatorics of sexual diploid reproduction allow us to find the distribution of HLA-types for biracial individuals. The phenotype of a person with one parent of race x and one of race y will be determined by six alleles–two in each of the loci, A, B, and $DRB1$. Let us denote these alleles as A_1, A_2, B_1, B_2, D_1, and D_2. Her six alleles could have been obtained in any one of 8 possible ways. For example, she could have inherited the three alleles A_1, B_1, and D_1 from the parent of race x (in the form of haplotype $A_1B_1D_1$) and the remaining alleles A_2, B_2, and D_2 (in the form of haplotype $A_2B_2D_2$) from the parent of race y. Alternatively, she could have inherited the haplotype $A_2B_1D_1$ from the parent of race x and the haplotype $A_1B_2D_2$ from the parent of race y. There are a total of 8 such possibilities. Since we have estimates of the haplotype distribution for each race, we can estimate the probability of each of these eight combinations. With this strategy, and the computational power of MatLab, we are able to calculate the probability of each of the roughly 25 million possible HLA phenotypes for persons of biracial parentage. [2]

2.3 Estimating Match Probabilities

Having estimated the distribution of phenotypes for persons of single race and biracial ancestry, we estimate the probability that a patient of specified ancestry will find a match in a registry of given size and racial composition. To make this estimate, we assume that the decision to join the registry and the need for a bone marrow transplant are both independent of a person's HLA type. Let R_x be the number of persons of race or biracial ancestry x in the registry and let p_i^x be the fraction of the population of group x that is of HLA type i. The distribution of phenotypes in the registry among persons from group x is determined by R_x independent draws, with probability p_i^x that each draw is of type i.

The probability that no type i's are to be found among the R_x registrants from group x is the probability that no type i's are selected in R_x random draws from the population of race x, which is

$$(1 - p_i^x)^{R_x}. \tag{1}$$

A person of type i will find no match in the registry if there are no i's among registrants of any race. Therefore, if R is the vector listing the number of registrants from each group x, the probability that a person of type i has no match of any race in the registry is

$$p_i^0(R) = \prod_x (1 - p_i^x)^{R_x}. \tag{2}$$

[2] Our earlier paper [6] applied this method in the special case where $x = y$ to estimate the distribution of phenotypes within each single race.

It follows that the probability that a person from group x has no match in the registry is

$$\sum_i p_i^x p_i^0(R). \tag{3}$$

2.4 Effective Registry Sizes and Probability of a Match

The NMDP reports the numbers of potential donors who designate themselves as white, African-American, Asian-American, and Hispanic. They also report 250,000 registrants of "multiple race", but do not report the numbers of specific multi-race combinations. To provide such estimates, we make the simplifying assumptions that multi-race registrants are all biracial and that the proportions of each biracial combination in the registry are the same as those found in the U.S. Census.[3] The 2000 U.S. Census, asked respondents to specify their racial or mixed-racial ancestry. About 0.28 percent of the population were identified as mixed-race African-American and white, 0.31 percent as mixed-race Asian and white, and 0.8 percent as white and "some other race" [23].

Not every registrant is able or willing to donate when called upon to do so. The NMDP has provided statistics on the availability rate for donors of each race. We used these estimates of availability to estimate the size of the "effective" registry for each race: the expected number of persons of each race who are registered and will be available if called upon to donate. The number of effective registrants of each race are reported in Table 1.

Using Equations 2 and 3 and our estimates of the effective size and composition of the NMDP registry, we estimate the probability that a randomly selected person of specified racial background will find a match in the NMDP registry. The resulting probabilities are reported in Table 2 of the Section 3.1.

2.5 Lives Saved by an Additional Registrant

In order to estimate the expected number of lives saved by an additional registrant of specified race, we first estimate the probability that this new registrant will provide a match for someone who did not previously have a match in the registry. Let us define $G_{xy}(R)$ as the increase in the probability that a person of race y will have a match in the registry that results from adding a registrant of race x, when the vector of registrants by race is R. Let $p_i^0(R)$ be the probability that there is no potential donor of type i in the registry. The probability that someone of race y is of type i and has no match in the registry is $p_i^y p_i^0(R)$, and the probability that a new registrant of

[3]Because the research on HLA distributions does not treat race and ethnicity in the same way that the U.S. Census does, this required some assumptions. The data on HLA is divided into four categories: white, black, Asian, and Hispanic. The U.S. Census asks respondents to identify with one or more races (white, black, Asian, and Native American or Pacific Islander) and one ethnicity (Hispanic or non-Hispanic). We defined Hispanics as those who identified as "white" and "Hispanic" or "some other race" and "Hispanic." The later group was 30 times larger than "some other race" and "non-Hispanic," so we infer that many Hispanics must have been treating Hispanic as a race in answering the question. Following this logic, we defined mixed race white-Hispanic as those who identified as "white and some other race" and "Hispanic." Determination of the population of other groups was more straightforward.

Table 1: Registrants by Race and Biracial Group

Race or Biracial Group	Number of NMDP Registrants	Probability Registrant is Available	Number of Effective Registrants
Single Race			
White	6,090,000	0.57	3,471,300
Hispanic	800,000	0.34	272,000
Asian-American	561,000	0.35	196,350
African-American	600,000	0.27	162,000
Biracial Group			
Hispanic, White	92,500	0.46	42,100
Asian-American, White	50,900	0.46	23,400
African-American, White	43,700	0.42	18,400
African-American, Hispanic	44,600	0.30	13,600
Asian-American, Hispanic	10,400	0.34	3,600
African-American, Asian-American	8,000	0.31	2,500

Notes: Registry numbers were obtained from the NMDP [24]. Numbers for mixed races are imputed as described in the text. The first four entries of the second column are from the NMDP (personal communication, Martin Maiers). Entries for mixed races are arithmetic means of the rates for constituent single races.

race x is of type i is p_i^x. Therefore the probability that a person of race y is of type i, has no match in the current registry, and will have a match if an additional person of race x is added to the registry is $p_i^x p_i^y p_i^0(R)$. Summing these probabilities over the types, we have

$$G_{xy}(R) = \sum_i p_i^x p_i^y p_i^0(R). \tag{4}$$

Adding one more HLA type to the registry will result in an additional stem cell transplant only if a patient of that HLA type is in need of a transplant. To calculate the probability that adding a person of race x to the registry will result in an additional transplant to a person of race y during a given year, we must multiply $G_{xy}(R)$ by the number of persons of race y who will seek transplants during that year. We display the resulting estimates in Table 5.

2.6 Measuring Benefits and Costs

Table 6 reports the probability that an additional registrant of specified ancestry will be the only available match for some patient in need of a transplant. Not every additional transplant saves a life. Possibly the patient would have survived without a transplant and possibly the patient will die despite receiving a transplant. In [6] we estimated the probability that providing a transplant will actually save a patient's life to be approximately 0.21. To determine the probability that adding an additional registrant will save a life during a single year, we must therefore multiply the probability found in Table 6 by 0.21.

Since no one knows who will need a transplant in the future, and since few know whether their HLA type is common or rare, the stem cell registry is properly viewed as a public good that produces a small increase in survival probability for each person in the community. Economists have developed a tool, known as the "value of a statistical life" for calculating benefits of public projects that enhance public safety. The value of a statistical life (VSL) is an estimate of the rate at which individuals are willing to exchange money for small increments of survival probability. The VSL can be described as the total amount that members of a community would be willing to pay per expected life saved if the increments in survival probability for each individual are small. Discussions of the theoretical and empirical underpinnings of the VSL can be found in [5], [11], [39], [38], [16], [17] and [7].

The U.S. Environmental Protection Agency in "Guidelines for Preparing Economic Analyses" [38] recommends a VSL of 7.9 million 2008 dollars. This is consistent with the results of the most recent survey by Knieser, Viscusi, and coauthors [20] who report that estimates of the value of a statistical life are concentrated in the range \$6 million to \$10 million 2008 dollars. We follow the EPA recommendation, using a VSL of \$7.9 million in our calculations.

Persons who join the bone marrow registry can remain in the registry until they reach age 61. The mean age of new registrants as reported by the NMDP is 35 years. We assume that a new registrant will, on average, remain in the registry for 25 years. Although medical technology is bound to change over the next 25 years and the number of persons annually seeking transplants may change dramatically,[4] we assume that the annual number of persons of each race seeking transplants will remain constant for the next 25 years and that the probability that a transplant saves the patient's life will remain constant as well. Following standard practice in economic benefit cost analysis, we discount future benefit flows by two percent per year. With these assumptions, we can calculate the expected present value of an additional effective registrant of each race. This quantity is found in the second column of Table 7 in Section 3.5.

The NMDP web site reports that the cost of tissue-typing an additional registrant was \$52 in 2007. Personal communication with sources at the NMDP indicates that the total cost of obtaining sample material, tissue-typing, and maintaining a record of a new potential donor's contact information is approximately \$105. Since not all registrants are available when called upon, the registry must on average add more than one registrant to gain an effective registrant. The fractions of registrants who can be located, pass the physical examination, and who consent to make a donation are .57 for white registrants, .27 for African Americans, .35 for Asian Americans, and .34 for Hispanics.

Increasing the number of registrants increases the expected number of transplants and hence the expected total hospital and physician costs of performing these transplants. We estimate that total hospital and physician costs for a transplant are about \$166,000. (This estimate is based on a survey of costs in 2001 by Redeaelli et al [29] and converted to 2007 dollars.) Multiplying this cost by the probability that an additional registration results in an additional transplant, we find that the expected annual

[4]The annual number of stem cell transplants from NMDP-registered donors has increased steadily over the decade from 1998 to 2008 at an average annual rate of 8.5 percent.

hospitalization costs resulting from adding a registrant range from about $7 for whites to about $28 for African American registrants. The third and fourth columns of Table 7 show our estimates of total costs attributable to adding an effective registrant of each racial group and the ratio of the present value of benefits to that of costs.

3 Results

We estimated the distribution of HLA types, the probabilities of finding a matching donor for persons of mixed race, and the cost-effectiveness of adding donors of mixed race. We find results that differ quite sharply from commonly-held beliefs reported in the news media.

3.1 Probability of Finding a Match

The widespread view that multi-racial patients have extremely poor chances of finding a match is expressed dramatically in a recent story in the *Boston Globe* [2].

> "If Nick Glasgow were white, he would have a nearly 90 percent chance of finding a matching bone marrow donor who could cure his leukemia.
>
> But because the 28-year-old bodybuilder is one-quarter Japanese, his doctor warned him the outlook was grim. Glasgow's background would make it almost impossible to find a match, which usually comes from a patient's own ethnic group.
>
> The doctor "didn't say it was slim-to-none. He didn't say it would be hard. He said 'zero chance,' " Glasgow's mother ... recalled."

Our estimates indicate that prospects of finding a match for biracial patients are not so dire as such accounts suggest.[5] Table 2 reports the estimated probability that a randomly selected person of each race or biracial group will find at least one person of matching HLA type enrolled in the NMDP registry who is willing and able to donate. The table shows that although biracial patients have lower matching probabilities than whites, their chances are not very different from those of patients who have two parents of a single minority race. In fact, patients with one white parent and one minority parent are more likely to find a match in the registry than those who have two parents of the minority race.

3.2 Sources of Matching Donors

A recent *Time Magazine* story [35] states the commonly-held view that multiracial patients are largely dependent for matching donors on the registry of their own type.

> "Though there are exceptions, the vast majority of successful matches take place between donors and patients of the same ethnic background."

This view is also reflected in the *Cleveland Plain Dealer* [40].

[5] Eventually, not just one, but two willing donors with matching HLA types were found for Mr. Glasgow.

Table 2: Probability of Finding a Matching Donor

Race or Biracial Group	**Probability of A Match in NMDP Registry**
Single Race	
White	0.93
Hispanic	0.82
Asian-American	0.77
African-American	0.58
Biracial Group	
Hispanic, White	0.87
Asian-American, White	0.80
Asian-American, Hispanic	0.72
African-American, White	0.71
African-American, Hispanic	0.65
African-American, Asian-American	0.50

Notes: Table entries are the probabilities that a person of specified racial background has a 6/6 HLA-A,B,DRB1 match in the NMDP registry with composition described in Table 1.

> "Finding a bone marrow match depends largely on finding someone of the same race to donate, which can be particularly difficult for people of mixed race. ... Mixed-race donors now make up only 3 percent of the 7 million potential adult marrow donors nationally."

If the only source of donors for multiracial patients were others of the same ancestry, their matching prospects would indeed be poor, as we see from the first column of Table 3. But as the second column shows, biracial patients have a significant chance of finding a donor among white registrants. The third column reports the probability of finding at least one donor of *some* racial background.

It is true that a single person of the same mixed race is more likely to be a match for a patient in need than someone of another racial background. But in the search for a match, this fact is overwhelmed by the fact that the registry has about 150 times many whites and as biracial, white and Asian-American registrants and 190 times as many whites as biracial, white and African-American registrants. (See Table 1.) From this larger pool, a majority of biracial individuals will have access to matching donors.

3.3 Matching for Multiple-Race Patients

It might seem that finding a match for someone with a complex multiracial heritage would be more difficult than for someone who is simply biracial. This view is articulated in *Time Magazine* [35].

Table 3: Probability of a Match, by Race of Recipient and Donor

Race of Recipient	Probability of Match with Donor of Own Racial Type	Probability of Match with White Donor	Probability of Match with Some Donor
Single Race			
White	.92	.92	.93
African-American	.16	.30	.58
Asian-American	.25	.43	.77
Hispanic	.27	.70	.82
Biracial Group			
Hispanic, White	.32	.82	.87
Asian-American, White	.16	.70	.80
African-American, White	.25	.59	.71
Asian-American, Hispanic	.04	.56	.72
African-American, Hispanic	.08	.46	.65
African-American, Asian-American	.02	.35	.50

Notes: Table entries are probabilities that a person of specified race or mixed race has a 6/6 HLA-A,B,DRB1 match from a registry of the composition given in Table 1.

> "Athena Mari Asklipiadis, the founder of the California-based Mixed Marrow, one of the only outreach groups devoted to recruiting mixed race donors ...maintains the rates are lower–much lower. 'God forbid I need a match, because I'm a very rare combination,' Asklipiadis says of her mixed Japanese, Italian, Armenian, Egyptian and Greek background."

In the Boston Globe [2] report about Mr. Glasgow, the leukemia patient with one Japanese grandparent and three white grandparents, an NMDP spokesperson is quoted as saying:

> "The truth is, when people of different backgrounds marry and produce offspring, it creates more types that are harder to match. The probability just gets lower when you have people of mixed ancestral DNA."

Fortunately for the matching prospects of multiracial individuals, the genetics of HLA recombination is not like the chromatics of mixing paint. One does not inherit some HLA alleles from each ancestor. Since the alleles controlling HLA type are located near one another on the same chromosome, genetic crossover of these alleles is rare. Therefore the HLA type of an individual, regardless of the racial diversity of her ancestry is almost always determined by just two haplotypes, one of which is passed down from each parent. It follows that the probability of finding a match for someone of complex multi-racial ancestry is not lower than that of biracial individuals, but instead is an average of matching probabilities over the various biracial couples that could be constructed from pairs of her ancestors.

Consider an individual whose four grandparents are of four different single races. The probability distribution of this person's type will be a mixture distribution with

a weight of one-fourth for each of the possible biracial pairs consisting of one paternal and one maternal grandparent.[6] The probability of having a match in the registry will be the arithmetic mean of the probabilities of finding a match for persons of these four possible biracial combinations. Suppose, for example, that a patient has one biracial parent of Asian-American and white descent and one biracial parent of African-American and Hispanic-American descent. Using the estimates of matching probabilities for biracial populations found in Table 2, we calculate that the probability that she will have a match in the NMDP registry is

$$\frac{1}{4}0.50 + \frac{1}{4}0.72 + \frac{1}{4}0.71 + \frac{1}{4}0.87 = 0.70.$$

It is straightforward to extend this method recursively to find matching probabilities for individuals who have one or more multi-racial grandparents.

3.4 Effectiveness of New Mixed-Race Registrants

News stories [35], [9], [37], [13] about multiracial patients seeking donations argue that because multiracial patients have much difficulty in finding matches, there is an urgent need for more multiracial donors in the stem cell registry. While it is true that biracial patients have a lower probability of finding a match in the registry than whites, this, in itself, does not imply that adding a mixed race registrant is more likely to save a life than adding a white registrant. As Table 4 shows, the number of biracial patients seeking stem cell transplants is very small. Consequently the probability that an additional biracial registrant will be a match for some biracial patient in need of a transplant is also small.

Despite the small number of mixed race patients, mixed-race registrants are among the most likely to be the only available match for some patient in need of a transplant. The reason is that patients with the same background are not the only ones for whom an additional mixed race donor might be a unique match. In fact, the probability is much larger that a mixed-race donor is the only available match for a patient of single race than for a mixed-race patient. Each entry in Table 5 is an estimate of the probability that an additional registrant of the racial group indicated for the entry's row will a match for a previously unmatched person of the racial group indicated for the entry's column.

The diagonal elements of the matrix in Table 5 show the probability that a new registrant of specified racial background will be the only available match for some patient of the same background. We see that the diagonal entries for persons of single race are much larger than the diagonal entries for biracial individuals. This means that if the only recipients of donations from biracial donors were of the same biracial background, adding biracial donors to the registry would be less effective than adding single-race registrants of any race.

Reading down the column for white recipients, we see something remarkable. The largest entries in this column are in the rows corresponding to biracial groups. This

[6] This is the distribution generated by first randomly choosing one maternal and one paternal grandparent and then choosing an HLA type from the distribution that applies to individuals with two parents whose races are those of the randomly selected grandparents.

Table 4: Annual Number of Potential Transplants

Race	**Number**
Single Race	
White	3,401
Hispanic	426
African-American	393
Asian-American	205
Biracial Group	
Hispanic, White	22
Asian-American, White	16
African-American, White	10
African-American, Hispanic	10
Asian-American, Hispanic	3
African-American, Asian-American	3

Notes: The number of NMDP-facilitated transplants by single race was obtained from [31]. To impute the number of transplants for mixed race groups, we started with our estimates of the distribution of races and mixed-races from the 2000 US Census as described in footnote 3. We then used estimates by Qian [28] of the ratio in which persons with one white and one minority parent self-identify by single race and assumed that biracial persons with two minority parents are equally likely to identify with each race. This enabled us to impute what proportion of each single race group "actually" belonged to each mixed race group. Finally, the number of potential transplants was inferred by dividing the number of performed transplants by the matching probability for each group reported in Table 2.

means that if one were to select a single registrant to add to the current registry with the objective of improving the matching prospects of white patients, adding a biracial registrant would be more effective than adding a white registrant. As the table shows, the best option for white patients would be to add someone who had one white and one African-American parent, while the second-best would be to add someone with one white and one Asian-American parent. Even a biracial registrant with no white parents is more likely to provide a unique match for a previously unmatched white patient than an additional white registrant.

Because the number of whites in the registry is very large, the types not currently represented in the registry are rare in the white population. As discussed in Section 4.1, these rare types include individuals who are culturally classified as white, but who carry haplotypes passed down from one or more nonwhite ancestors. For such persons, the biracial population is the most promising source of potential matches.

Table 6 summarizes these results. The first column of this table shows the probability that adding a registrant of specified single race or biracial group will result in a match for a patient with *the same* ancestry who has no other match. These entries are smaller for biracial individuals than for persons of single race. The second col-

Table 5: Annual Probability that a Registrant will be the Only Match for a Patient Needing a Transplant; by Race of Registrant and of Recipient (times 10^5)

Race of Registrant	**Race of Recipient**									
	W	Af	As	H	Af-W	As-W	H-W	Af-As	Af-Hi	As-H
W	3.78	0.42	0.19	0.54	0.02	0.02	0.03	0.00	0.01	0.00
Af	3.63	23.42	0.27	2.47	0.22	0.03	0.08	0.05	0.25	0.01
As	3.09	0.51	7.97	0.93	0.02	0.23	0.04	0.03	0.02	0.05
H	4.28	2.28	0.45	5.04	0.06	0.04	0.12	0.02	0.13	0.02
Af-W	5.48	8.25	0.34	2.46	0.31	0.04	0.11	0.05	0.21	0.01
As-W	5.00	0.67	2.85	1.08	0.03	0.49	0.05	0.04	0.03	0.06
H-W	4.48	1.38	0.35	2.21	0.05	0.04	0.10	0.01	0.05	0.01
Af-As	4.89	7.12	2.41	2.81	0.20	0.24	0.09	0.46	0.22	0.10
Af-H	4.84	9.56	0.45	5.47	0.22	0.04	0.12	0.06	0.45	0.02
As-H	4.80	1.48	3.48	2.78	0.04	0.34	0.09	0.09	0.07	0.16

Notes: Equation 4 was used to calculate the effect of adding a registrant of the row race on the probability that a patient of the column race will find a match. The result was multiplied by the number of patients of the column race seeking transplants in 2008, as given in Table 4.

umn records the probability that an additional registrant will produce a match for a patient *of some* race who would otherwise have no match. This column shows that new mixed race registrants are more likely to provide a match for some previously unmatched patient than are new white registrants, though less likely to do so than new African-American registrants.

3.5 Comparing Benefits and Costs

Employing the detailed methods discussed in Section 2.6, we compare the expected present value to the expected cost of adding a new potential donor to the NMDP registry. Table 7 displays the resulting comparisons for new registrants by racial background.

According to our estimates, the expected value of adding a new registrant to the stem cell registry exceeds the cost for registrants of all single-race and biracial backgrounds. The ratio of benefits to costs is highest for African-American registrants, intermediate for other minorities and those of mixed race and lowest for white registrants.

Table 6: Annual Probability that a Registrant Will Be a Unique Match for Some Patient ($\times 10^5$)

Race of Registrant	Probability of Unique Match For Own Race	Probability of Unique Match For Some Race
Single Race		
African-American	23.4	30.4
Asian-American	8.0	12.9
Hispanic	5.0	12.4
White	3.8	5.0
Biracial Group		
African-American, Hispanic	0.4	21.2
African-American, Asian-American	0.5	18.5
African-American, White	0.3	17.3
Asian-American, Hispanic	0.2	13.3
Asian-American, White	0.5	10.3
Hispanic, White	0.1	8.7

Notes: The first column repeats the diagonal entries from Table 5 while the second column is the row sum from that table.

4 Discussion

4.1 Racial Categories and HLA Matching

The four major racial categories into which NMDP registrants are partitioned are coarse and quite arbitrary. Since the recorded race of a registrant is self-declared, it does not necessarily correspond to genetic inheritance. Statistics show, however, that the distribution of HLA types differs markedly across these self-identified categories. For example, our estimates suggest that the probability that a random self-reported white will match another randomly selected white is 34 times that of matching a random Asian-American, 16 times that of matching a random African-American, and 6 times that of matching a random Hispanic.

Our computations are based on the Kollman *et al* [21] estimates of the distribution of haplotypes within each race, which depend on two critical assumptions about marriage patterns. The first is that racial groups are *endogamous* (marriage occurs almost entirely within races). The second is that, conditional on marrying within their group, the probability that two people marry is independent of their HLA types.

Since marriage patterns are more likely to be determined by socially perceived race than by genetic characteristics, the use of self-declared race to determine categories seems appropriate for the model estimated here. Jacobs and Labov [15] collected data on all married heads of households and their spouses from a 1 percent sample of the 1990 U.S. Census. They found that almost 98 percent of marriages of whites and 96 percent of marriages of African-Americans were endogamous. The Jacobs-Labov study shows that approximately 85 percent of Asian-Americans are married to other

Table 7: Present Value, Cost, and Benefit-Cost Ratios of a New Effective Registrant: by Race or Biracial Ancestry of Registrant

Race of Registrant	Present Value	Cost	Benefit-Cost Ratio
Single Race			
African-American	$9,900	$1381	7.2
Asian-American	$4,200	$720	5.8
Hispanic	$4,000	$714	5.6
White	$1,600	$348	4.6
Biracial Group			
African-American, Hispanic	$6,900	$1036	6.7
African-American, White	$5,600	$813	6.9
Asian-American, Hispanic	$4,300	$739	5.8
African-American, Asian-American	$6,000	$943	6.4
Asian-American, White	$3,300	$564	5.9
Hispanic, White	$2,800	$514	5.4

Notes: The annual value of adding an effective registrant was calculated by multiplying the final column of Table 6 by 0.21 statistical lives saved per transplant and multiplying again by $7,900,000 per statistical life. Entries in the table are present values of 25 years of this annual value discounted at 2%, rounded to the nearest $100. Costs were calculated as discussed in the text. The benefit-cost ratio is the ratio of column 2 to column 3.

Asian-Americans and 77 percent of Hispanics are married to other Hispanics.[7] [8]

Although current rates of intermarriage between African-Americans and whites are low, African-Americans carry a significant amount of genetic material obtained from white ancestors. As Kittles *et al* [18] observes, "The vast majority of contemporary African Americans are descendants of enslaved Africans kidnapped and transported to America during the transatlantic slave trade from 1619 to 1850." During the period of slavery, there was substantial mixing of the white and African-American gene pool. Kittles *et al* reports that it is estimated that there were about 4.5 million people of African descent in the U.S. in 1860, of which 600,000 were of mixed ancestry.

Geneticists have developed methods for using genetic markers to estimate *admixture* proportions, which are the proportions of genetic material in a single population that is inherited from members of two or more distinct ancestral populations. (See [36] for a discussion of these methods and further references.) Several studies have estimated admixture proportions from samples of African-Americans. These studies indicate that the percentage of European admixture in the African-American population differs substantially by region, ranging from 3.5 percent in the Gullah sea island community of South Carolina, 10 percent in the rural South, about 20 percent in the industrial North,

[7] Jacobs and Labov report rates of out-marriage for each of several Asian nationalities. We weighted these rates by the number of marriages of each type to find an average rate of out-marriage.

[8] According to Jacobs and Labov, among Hispanics, the marriages of 82 percent of Mexican-Americans, 76 percent of Cuban-Americans and 66 percent of Puerto Ricans were endogamous within these subgroups.

and 22-35 percent on the West Coast, see [18] (Figure 2) and [26]. (The admixture of African-American genetic material in the U.S. white population appears to be much smaller. Shriver *et al* [36] estimates a mean admixture rate of less than one percent.)

The genetic composition of the current population depends on the marriage patterns of their parents' generation. The current population of Asian-Americans and of Hispanics are children of more endogamous populations than is indicated by current marriage patterns. About 2/3 of the existing population of Asian-Americans were born in Asia and their ancestors for many generations would have had little exposure to non-Asians. About one third of the Hispanic population are immigrants from regions where the population is almost entirely Hispanic.

While the assumption that Asian-Americans marry endogamously is not wildly inaccurate, the assumption that marriage among Asian-Americans is random with respect to HLA type is clearly violated. The marriage patterns of the parents of the current generation of Asian-Americans were far from random. Two-thirds of this population are immigrants, coming from several distinct Asian populations that have been geographically separated for many generations. Table 8 reports the distribution of the Asian-American population by national origin. Even after reaching the United States, Asian-Americans have been far more likely to marry within their own nationality than outside of it. Jacobs and Labov [15] find that about 80% of Asian-American marriages are between two people of the same national origin. Since the distribution of HLA types differs significantly between Asian nationalities [8], it follows that the HLA types of Asian American marriage partners are not random draws from the distribution of HLA types in the full Asian American population.

Table 8: National Origin of Asian-American Population

National Origin	Fraction
China & Taiwan	0.24
Indian subcontinent	0.17
Philippines	0.17
Vietnam	0.10
Korea	0.10
Pacific Islander	0.08
Japan	0.07
Other	0.06

Notes: Fractions are calculated from the 2000 U.S. Census publication [3], Table 4.

The Hispanic population of the United States includes distinct sub-populations that differ in ethnic makeup and have had little contact with each other for many generations. About 66 percent of the Hispanic population of the United States is of Mexican extraction, 13 percent come from Central and South America, 9 percent are Puerto Rican, and 4 percent are of Cuban extraction. Genetic admixture studies of Hispanics in the U.S. reveal that Mexican-Americans on average have 30-40 percent Native American ancestry, while immigrants from the Spanish Caribbean have African

genetic contributions that range from 20-40 percent and contributions of about 18 percent from the native American Arawaks and Caribs; see [18] and [19].

Estimates of matching probabilities could be improved with the use of HLA haplotype distributions for more finely distinguished population groups. Many published estimates of frequency distributions of common HLA haplotype distributions for narrowly-defined national and regional populations throughout the world are available. Examples include [33],[8], [30], [34] and [10]). These estimates are usually based on samples of a few hundred individuals and hence provide reliable information only about the distributions of the most common haplotypes. It would be interesting to apply statistical methods along with estimates of the most common haplotypes to disentangle the distribution of haplotypes of highly endogamous subpopulations within the broad racial groups for which we have aggregate data.[9]

4.2 Cord Blood

In recent years, umbilical cord blood collected from newborn infants and cryogenically stored has become an important alternative source for stem cell transplants. In 2010, the NMDP had 145,000 units of cord blood in storage, as compared to 9 million registered potential adult donors, and about 1/5 of all stem cells arranged by the NMDP were from cord blood [25]. As with adult stem cells, the success probability of cord blood transplants depends on the matching of the immune systems of donor and patient, though with cord blood, mismatches of one or two alleles are thought to be less dangerous than they are for adult stem cells. The method used in this paper can be applied to finding the probabilities of full and partial matches in the cord blood registry.

A more thorough benefit-cost analysis would need to explore optimal strategies for cord blood storage and to account for the fact that the adult stem cell registry and the bone marrow registry are partial substitutes. The logistics of the adult stem cell registry are very different from that of the cord blood banks. The adult registry stores only information, while the cord blood banks must store frozen cord blood. Obtaining, testing, and processing a unit of cord blood costs about $2000 and there is an annual cost of about $50 for storage. Most cord blood units are not large enough to serve a full grown adult patient (There is an emerging technology for using cord blood from two infants for patients who cannot find a sufficiently large single unit.) The cost of registering and testing a potential donor for the adult registry is about $100. When a transplant occurs, retrieval of a cord blood unit is cheaper and quicker than obtaining an adult donation, which involves a hospital visit for the donor and related expenses.

In a future paper we intend to explore optimal strategies for maintaining a cord blood registry and the optimal mix of banked cord blood units and registered potential adult donors.

[9] Pritchard, Stephens, and Donnelly [27] address a related problem of "cryptic" population structure, extracting the allele distributions in sub-populations. They study the distribution of alleles at several unlinked loci, while we seek the distribution of haplotypes for tightly linked loci.

5 Summary

We estimated the probability distribution of HLA types for persons of mixed race and calculated the probability that a patient of specified racial background will find a matching donor in the existing stem cell registry. If the only source of matches for multi-racial individuals were donors of the same multi-racial ancestry, their chances of finding matching donors would be small. But biracial individuals stand a good chance of finding a match in the registry of white donors. Just as patients of mixed race are more likely to find a match in the pool of white donors than in the much smaller pool of donors of their own background, those of mixed race who join the stem cell registry are more likely to be the only available match for a white patient than for one of their own race.

Our estimates support the view that recruitment of mixed-race registrants deserves high priority, not so much because of the plight of mixed race patients, as because mixed race individuals are likely to match previously unmatched patients classified as being of single race.

Remarkably, it turns out that adding a biracial registrant (of any biracial combination) is more likely to produce a match for a previously unmatched white patient than adding a white registrant. According to our estimates, the present value of benefits from adding registrants of all races and biracial combinations exceed the costs. The net benefit from additional African-American registrants is highest, followed by the net benefits from biracial individuals and then by the net benefits from other single-race minorities. Benefits from adding a white registrant are lower, but still about four times as large as costs.

References

[1] C. Anasetti, R. Hillgruber, V. Nye, E. Ayala, et al. Patient ethnicity markedly affects the probability of finding an HLA-A, -B,-C amd DRB 1 allele matched unrelated donor for hemopoietic cell transplantation. *Biology of blood and marrow transplantation*, 16(2, Supplement 1):S172, February 2010.

[2] Juliana Barbassa. Mixed-race patients struggle to find match in bone marrow donor. *The Boston Globe*, May 28 2009. `http://www.boston.com/news/nation/articles/2009/05/28/mixed_race_patients_struggle_to_find_marrow_donors/` (accessed Aug 18, 2010).

[3] Jessica Barnes and Claudette Bennett. The Asian population, 2002. `http://www.census.gov/prod/2002pubs/c2kbr01-16.pdf` (accessed June 11, 2010).

[4] Patrick Beatty, Motoro Mori, and Edgar Milford. Impact of racial genetic polymorphism on the probability of finding an HLA-matched donor. *Transplantation*, 60(8):778–783, October 1995.

[5] Theodore Bergstrom. When is a man's life worth more than his human capital? In M. W. Jones-Lee, editor, *The Value of Life and Safety*, pages 3–26. North Holland, Amsterdam, 1982.

[6] Theodore C. Bergstrom, Rodney Garratt, and Damien Sheehan-Connor. One chance in a million: Altruism and the bone marrow registry. *American Economic Review*, 99(4):1309–1334, September 2009.

[7] Glenn C. Blomquist. Self-protection and averting behavior, values of statistical lives, and benefit-cost analysis of environmental policy. *Review of the Economics of the Household*, 2, 89-110 2004.

[8] J G Bodmer, L J Kennedy, J Lindsay, and A M Wasik. Applications of serology and the ethnic distribution of three locus HLA haplotypes. *British Medical Bulletin*, 43(1):94–121, 1987.

[9] Sandra G. Boodman. Multiracial patients struggle to find donors for bone marrow transplants. *Washington Post*, June 1 2010. `http://www.washingtonpost.com/wp-dyn/content/article/2010/05/31/{AR}2010053102481.html` (accessed Aug 18, 2010).

[10] Dominique Charron. *Proceedings of the twelfth international histocompatibility workshop*. EDK, Paris, 1997.

[11] Pierre Dehez and Jacques Drèze. State dependent utility, the demand for insurance, and the value of safety. In M. W. Jones-Lee, editor, *The Value of Life and Safety*, pages 41–65. North Holland, Amsterdam, 1982.

[12] J Dehn, K Buck, S Yang, S Schmidt, et al. 8/8 high-resolution match rate: caucasian and African American patients. *Biology of blood and marrow transplantation*, 17(2, Supplement 1):S170–171, February 2010.

[13] Liane Hansen. Maya's mom, cherishing the 'well' moments. *NPR Weekend Edition*, August 8 2010. radio interview `http://www.npr.org/templates/story/story.php?storyId=129061715` (accessed Aug 24, 2010).

[14] C.K. Hurley, M. Fernandez Vina, and M. Setterholm. Maximizing optimal hematopoietic stem cell donor selection from registries of unrelated adult volunteers. *Tissue Antigens*, 61:415–424, 2003.

[15] Jerry A. Jacobs and Teresa G. Labov. Gender differentials among sixteen race and ethnic groups. *Sociological Forum*, 17(4):621–646, December 2002.

[16] Magnus Johannesson, Per-Olav Johansson, and Karl-Gustav Löfgren. On the value of changes in life expectancy: Blips versus parametric changes. *Journal of Risk and Uncertainty*, 15:221–239, 1997.

[17] Michael Jones-Lee, M. Hammerton, and P.R. Philips. The value of safety: results of a national sample survey. *Economic Journal*, 95:49–72, 1985.

[18] Rick A. Kittles, Eunice R. Santos, S. Oji-Nijkeda Nertiti, and Carolina Bonilla. Race, skin color and genetic ancestry: Implications for biomedical research on health disparities. *California Journal of Health Promotion*, 5:9–23, 2007.

[19] Yann C. Klimentidis, Geoffrey F. Miller, and Mark D. Shriver. Genetic admixture, self-reported ethnicity, self-estimated admixture, and skin pigmentation among hispanics and native americans. *American Journal of Physical Anthropology*, 138(4):375–383, April 2008.

[20] Thomas J. Knieser, W. Kip Viscusi, Christopher Woock, and James P. Ziliac. The value of a statistical life: Evidence from panel data. Vanderbilt University Law School Law and Economics Working Paper Number 11-02, June 2010.

[21] Craig Kollman, Martin Maiers, Loren Gragert, Carlheinz Müller, et al. Estimation of HLA-A -B -DRB1 haplotype frequencies using mixed resolution data from a national registry with selective retyping of volunteers. *Human Immunology*, 68(12):950–958, December 2007.

[22] M Mori, P.G. Beatty, M. Graves, KM Boucher, and F.L. Milford. HLA gene and haplotype frequencies in the North American population: the National Marrow Donor Program Donor Registry. *Transplantation*, 64:1017–1027, 1997.

[23] Social Science Data Analysis Network. Table: Multiple race combinations by frequency, 2010. `http://www.censusscope.org/us/chart_multi.html` (accessed June, 2010).

[24] NMDP. Nmdp facts and figures. http://www.marrow.org/NEWS/MEDIA/Facts_and_Figures/Facts_\%26_Figures\%20\%28PDF\%29.pdf, 2009.

[25] NMDP. Nmdp facts and figures. http://www.marrow.org/NEWS/MEDIA/Facts_and_Figures/Facts_\%26_Figures\%20\%28PDF\%29.pdf, 2010. (accessed April, 2011).

[26] Esteban J. Parra, Amy Marcini, Joshua Akey, Jeremy Martinson, et al. Estimating African American admixture proportions by the use of population specific alleles. *American Journal of Human Genetics*, 63:1839–1851, 1998.

[27] Jonathan Pritchard, Matthew Stephens, and Peter Donnelly. Inference of population structure using multilocus genotype data. *Genetics*, 155:945–959, June 2000.

[28] Zhenchao Qian. Options: Racial/ethnic identification of children of intermarried couples. *Social Science Quarterly*, 85(3):746–66, September 2004.

[29] Alberto Redaelli, Marc F. Botteman, Jennifer M. Stephens, and Chris L. Pashos. Economic burden of acute myeloid leukemia: A literature review. *Cancer Treatment Reviews*, 30(3):237–247, 2004.

[30] S. Rendine, I. Borelli, M. Barbanti, N. Sacchi, S. Roggero, and E. S. Curtoni. HLA polymorphisms in Italian bone marrow donors: a regional analysis. *Tissue Antigens*, 52:135–146, 1998.

[31] Health Resources and Services Administration. Detailed description of registry transplant data, 2009. http://bloodcell.transplant.hrsa.gov/RESEARCH/Transplant_Data/Registry_Tx_Data/LongDesc/index.html (accessed July 28, 2009).

[32] A. H. Schmidt, D Baier, U.V. Solloch, A. Stahr, et al. Estimation of high-resolution HLA-A, -B, -C, -DRB1 allele and haplotype frequencies based on 8862 German stem cell donors and implications for strategic donor registry planning. *Human Immunology*, 70(11):895–902, 2009.

[33] A.H. Schmidt, U.V. Solloch, D. Baier, A. Stahr, R. Wassmuth, G. Ehninger, and C. Rutt. Regional differences in hla antigen and haplotype frequency distributions in Germany and their relevance to the optimization of hematopoietic stem cell donor recruitment. *Tissue Antigens*, 76:362–379, 2010.

[34] C.K. Shaw, L.L. Chen, A. Lee, and T.D. Lee. Distribution of HLA gene and haplotype frequencies in Taiwan: a comparative study among Min-nan, Hakka, Aboriginal, and Mainland Chinese. *Tissue Antigens*, 53(1):51–64, January 1999.

[35] Christopher Shay. Bone marrow transplants: When race is an issue. *Time Magazine*, June 3 2010.

[36] Mark D. Shriver, Estaban J. Parra, Sonia Dos, Carolina Bonilla, et al. Skin pigmentation, biogeographical ancestry and admixture mapping. *Human Genetics*, 112:387–399, 2003.

[37] Jerome Taylor. 125,000/1 against: this boy's chances of finding a bone marrow donor. *The Independent (UK)*, Tuesday, June 8 2010. `http://www.independent.co.uk/life-style/health-and-families/health-news/1250001-against-this-boys-chances-of-finding-a-bone-marrow-donor-1994098.html` (accessed Aug 24, 2010).

[38] U.S. Environmental Protection Agency. Guidelines for preparing economic analyses. Internet publication, December 2010. `http://yosemite.epa.gov/ee/epa/eed.nsf/pages/Guidelines.html/$file/Guidelines.pdf`.

[39] W. Kip Viscusi and Joseph Aldy. The value of a statistical life: A critical review of market estimates throughout the world. *Journal of Risk and Uncertainty*, 27(1):5–76, 2003.

[40] Brie Zeltner. Mixed-race bone marrow donations increasing, but still slow. *The Cleveland Plain Dealer*, September 22 2009. `http://www.cleveland.com/healthfit/index.ssf/2009/09/mixedra_bone_marrow_donation.html` (accessed Mar 30, 2011).