Bone Marrow Donor Matching for Patients of Mixed Race

Ted Bergstrom, Rod Garratt, and Damien Sheehan-Connor

September 4, 2009

Background

Many 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 HLA immune systems of the donor and recipient must be a close genetic match. This requires that they have the same set of six alleles located at three genetic loci[1] found on the same chromosome.

In sexual diploid species, two alleles are found in each genetic locus. One of these is inherited from each parent. The string of alleles that is passed from parent to child is known as a haplotype. Because the alleles that determine one's HLA type are located close together on the same chromosome, genetic crossover is rare. Therefore with rare exceptions, HLA haplotypes remain intact from generation to generation, with each parent randomly selecting one of two possible haplotypes to pass to each of its children. Two siblings will be suitable donors for each other only if each received the same haplotype from each parent. This happens with probability one-fourth. About 70% of all patients needing transplants have no matching sibling. Those without a sibling donor must seek a match from the population at large.

Finding a non-sibling match is difficult because the distribution of HLA

[1]These loci are known as HLA-A, HLA-B, and HLA-DRB. Recent research indicates that success probabilities are improved if there is also a match at the loci HLA-C and HLA-DRQ

types is extremely diffuse. There are more than 10 million possible types. Approximately one half of Americans of European ancestry belong to types with frequency less than one in one hundred thousand, while twenty percent belong to types with frequency less than one in a million. The distribution of types is even more diffuse for persons of Asian and African ancestry. Two individuals are far more likely be a match if they are of the same race. Because the world's registries have many more persons of European than of Asian or African ancestry, chances of finding a match are significantly smaller for the latter groups.

In response to this problem, the developed nations of the world have set up national volunteer registries. Registrants pledge that they will contribute stem cells to any needy patient for whom they are the best available match. Each registrant's HLA type is determined by DNA testing and is recorded along with contact information. The U.S. National Marrow Donor Program (NMDP) had roughly 5.5 million persons of European extraction (whites), 520,000 Asian-Americans, 550,000 African-Americans, and 690,000 Hispanics.[2]

The largest available source of data on the distribution of HLA types is the NMDP registry. The NMDP records the self-reported race and determines the HLA type of each registrant. In the early years of the registry, most volunteers were typed only at two loci, HLA-A and HLA-B. As technology improved, new registrants were typed at the three loci, HLA-A, HLA-B, and HLA-DRB1. In a 1997 paper, M. Mori *et al* [9] used a sample of about 400,000 registrants who had been typed at three loci to estimate the distribution of HLA types in each racial subgroup of the U.S. population. More recently, Kollman *et al* [7] performed a similar exercise, with a sample of about two million registrants, including 1.2 million European-Americans, 250,000 Asian-Americans, 280,000 African-Americans, and 320,000 Hispanics.[3] Although the samples used in the Mori and Kollman studies are large, they are not nearly large enough to provide

[2]The registries of the world include about 10 million persons of European ancestry, 1.4 million of Asia ancestry, and 650,000 of African ancestry and 690,000 persons designated as Hispanic-Americans. Japan with 328,000 and Taiwan with 285,000 have the largest registries in Asia. Hong Kong has 67,000. China has 7,000 and India 1,000 registrants. The number of black Africans in African registries is extremely small.

[3]The Kollman study was able to achieve this larger sample partly because the number of the increased number of persons in the registry who have been typed at all three loci and partly because it applied a more general estimation method that is able to use extract information from data about early registrants who had been typed only at two loci.

good direct estimates of the distribution of relatively rare types. Many types HLA-types that are present in the population will simply not appear at all in the sample. But the mechanics of diploid genetics make it possible, with the aid of reasonable assumptions about mating patterns, 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" of the order of 3,000, 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 that report estimates of the frequency distribution of haplotypes in each racial group.

In a recent paper [2], we used Mori's estimates of the haplotype distribution of whites, African-American, Asian-American, and Hispanics in the NMDP registry to estimate the distribution HLA-types in each race. We made these estimates under the assumption that individuals mate within their own race and that within-race mating is random with respect to HLA-type. With these estimates we calculate the probability that individuals of each race would find a matching HLA type in a registry of given size and racial composition. We then conducted a benefit-cost analysis of the value of adding registrants of each race to that registry. We concluded that the expected present value of adding registrants of all races to the NMDP exceeded the costs, with the difference being largest for the minority populations.

In the current paper, we update our estimates of matching probabilities based on the larger and more recent Kollman study [7] and on more recent statistics on the size of the NMDP registry. We also present what we believe to be the first systematic analysis of the probability of finding a match for persons of mixed race.

Matches for Persons of Mixed Race

It has often been observed that finding a match is especially difficult for persons of mixed race. Since a matching donor must have inherited the same genes from both parents, the best prospect for a person of mixed race is someone of the same mixed-race heritage. But persons of specific mixed-race combinations are relatively scarce. Approximately 2 percent of the population in the United States, 1.5 percent of the population of Canada, and 1.4 percent of the

population of the United Kingdom declare themselves to be of mixed race.[4]

A dramatic instance of this problem appeared in an Associated Press story dated May 27, 2009.

> *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, he said, would make it impossible to find a match, because a match 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, Carole Wiegand, recalled...*
>
> *At a time when the number of multiracial Americans is rising, only a tiny fraction of donors on the national bone-marrow registry are of mixed race. The National Marrow Donor Program is trying to change that by seeking more diverse donors for patients suffering from leukemia, lymphoma and other blood diseases. "The truth is, when people of different backgrounds marry and produce offspring, it creates more types that are harder to match," said Michelle Setterholm, the program's director of scientific services. "The probability just gets lower when you have people of mixed ancestral DNA."*

Friends and co-workers of Mr. Glasgow organized a vigorous web-based campaign to find a matching donor for him, with special attention devoted to persons of mixed Japanese-European ancestry. Within a month, not one, but two willing donors of matching HLA type were found. Evidently he had better than a "zero chance" after all.

Awareness of some simple facts of genetics might have improved the quality of this story (and of the Stanford doctor's opinion). Since the relevant HLA-alleles are closely linked, recombination of these alleles is rare, and children of

[4] In the U.S. census of 2000, about 0.28 percent of the population was identified as mixed-race African-American and white and 0.31 percent as mixed-race Asian and white, about 0.8 percent were identified as white and "some other race".

mixed race will almost always carry one haplotype inherited from each parent. If the only information we had about Mr. Glasgow's HLA-type was that he is mixed Japanese-European ancestry and does not have a match in the NMDP registry, we should conclude that his chances of finding a match from further search are better, not worse, than they would have been had he had been entirely of European ancestry. In the latter case, we would know that he had no match in a registry of 5.5 million potential donors. Thus he would almost certainly be of an extremely rare HLA-type and the chances that he would find a match among another thousand or even ten thousand newly recruited donors would be tiny. But given that he has one Japanese grandparent, it is quite possible that he has one haplotype that is not especially rare among Japanese but rare among whites and one that is not especially rare among whites, but rare among Japanese. In this case, his best prospects for a match are persons with one Japanese and one European parent.[5] The number of such persons currently in the registry is probably less than 4,000[6] and so the fact that he did not have a match in the current registry does not imply that his type is extremely rare among persons of Japanese-European ancestry. Thus there seems a reasonable chance that adding a few hundred persons of this background would produce a match.

HLA-Type Distributions By Race and Mixed Race

Even though the Kollman and Mori studies did not have access to large samples of HLA-types of persons of mixed races, their estimated haplotype distributions for the unmixed races can be used to estimate the distribution of HLA-types

[5]It is worth pointing out that the fact that someone is one-fourth Japanese and three-fourths white does not mean that his most likely match is of the same fractional ancestry. Someone with one Japanese and three white grandparents has a probability one half of having one haplotype of Japanese descent and one of European descent and a probability one half of having two haplotypes of European descent. In the former case his best prospects for a match have one parent of each race. In the latter case, his best prospect is someone with two white parents.

[6]We estimate that there are about 42,000 persons of mixed Asian-European ancestry in the registry. Japanese-Americans constitute about 7 percent of the Asian-American population. If they make up the same proportion of persons of Asian-white ancestry in the registry, there would be less than 3,000 such persons in the registry.

for any desired mixed race combination. This is made possible by applying the simple combinatorics of sexual reproduction. A person with one parent of race X and one of race Y could acquire the six alleles a_1, a_2, b_1, b_2, dr_1, and dr_2 in any one of 8 ways. She could inherit the three alleles a_1, b_1, and dr_1 from the parent of race X (in the form of haplotype $a_1b_1dr_1$) and the remaining alleles a_2,b_2, and dr_2 (in the form of haplotype $a_2b_2dr_2$) from the parent of race Y. Alternatively, she could inherit the haplotype $a_2b_1dr_1$ from the parent of race X and the haplotype $a_1b_2dr_2$ from the parent of race Y. There are a total of 8 such combinations. Since we have the estimated haplotype distribution for each race, we can calculate the probability of each of these eight combinations for a person with parents of races X and Y. Adding these 8 probabilities we have the probability that the individual's HLA-type is given by the alleles a_1, a_2, b_1, b_2, dr_1, and dr_2.

Having calculated the distribution of HLA-types for persons of specific mixed races, we use the method described in our earlier paper [2] to calculate the probabilities that persons of each race and mixed race will have an HLA match in the NMDP registry. To perform this calculation we need to estimate the number of persons of each race and mixed race in the registry. The NMDP reports the number of self-designated whites, African-Americans, Asian-Americans and Hispanics in the registry. They also report that

With the estimated distribution of HLA-types for each mixed race group, we can calculate the probability that persons of each single and mixed race group will find a match in a registry, given the number of persons of each race and mixed race in the registry. They also report that 210,000 registrants are of "multiple race" but does not specify the racial combinations. We allocated these persons among the possible combinations of pairs in the proportions that these mixes reported in the 2000 U.S. census. The resulting estimates of numbers of registrants of each type appear in the first column of table 1. In the second column of this table, we report our estimates of the probability that a person in each group does not have a match in the registry.

It is interesting to see that Americans of mixed race do not fare as badly as might be expected. Those with one white parent and one minority parent parent are slightly *more* likely to find a match in the registry than persons with two parents from the same minority group. While these individuals are less likely to have a match in the minority population than those whose parents both belong to that minority, they are more likely to have a match in the white

Table 1: Probability of No Match in Registries

Race	**Number of Registrants**	**Prob No Match in NMDP Registry**
White	5,300,000	0.07
African-American	550,000	0.41
Asian-American	520,000	0.22
Hispanic	690,000	0.17
Mix: African-Amer, White	36,700	0.28
Mix: Asian-Amer, White	42,700	0.20
Mix: Hispanic, White	77,700	0.13
Mix: African-Amer,Asian-Amer	6,700	0.49
Mix: African-Amer,Hispanic	37,400	0.34
Mix: Asian-Amer, Hispanic	8,700	0.27

registry. Table 2 compares the match probabilities of persons of minority races with those of mixed parentage. The first column shows the probability that an individual will find a match of his or her minority parent's race.[7]

[7]These estimated probabilities of not finding a match for the "pure" races differ from the estimates we made in [2] for two reasons. The current estimates are based on haplotype distributions from the Kollman study while the earlier estimates were based on haplotype distributions from the Mori study. The current estimates use the 2008 NMDP registry size and the earlier estimates used the corresponding figures for the 2006 registry. To check for consistency of the Mori-based and Kollman-based results, we compared estimates of probabilities of finding a match with the two estimated haplotype distributions while holding the registry size constant at the 2006 level. The results were reassuringly similar. Using the Mori haplotype distribution, the estimated probabilities of finding a match were .922 for whites, .623 for African-Americans, .793 for Asian-Americans, and .835 for Hispanic-Americans. Using the Kollman haplotype data, these estimates were .922 for whites, .567 for African-Americans, .755 for Asian-Americans, and .807 for Hispanics.

The second column shows the probability that the individual will find a match in the registry of whites and the third column is the probability that the individual will find a match of any race.

Table 2: Probability of Match in NMDP by Source

Race of Patient	**Prob of Minority Match**	**Prob of White Match**	**Prob of Any Match**
Asian	.64	.45	.78
Asian-White Mix	.40	.69	.80
African	.41	.30	.59
African-White Mix	.36	.59	.72
Hispanic	.63	.69	.83
Hispanic-White Mix	.68	.82	.87

Those who have parents of two different minorities do not fare as well. We see from Table 1 that those with one Asian-American parent and one African-American parent are less likely to find a match than those who have two Asian-American or two African-American parents. Those with one Hispanic and one Asian-American parent, have smaller probabilities of a match than those with two Hispanic or two Asian-American parents. Those with one African-American parent and one Hispanic parent have a better chance of a match than those with two African-American parents, but a smaller chance than those with two Hispanic parents.

Benefit Cost Analysis

In our earlier paper [2], we showed how to estimate the expected number of lives saved by adding new registrants of specified race. To do so, we estimate the probability that the new registrant will be the only match for some patient needing a transplant and multiply this probability by the probability that having a match rather than not having a match will save the patient's life. Table

3 reports the results of this calculation for adding adding 1000 registrants of specified race to the NMDP registry.

Table 3: Lives Saved By 1000 New Registrants: By Race of Registrant

Race	**Expected Lives**
White	0.0096
African-Amer	.050
Asian-Amer	.020
Hispanic	0.021
Mix: African-Amer, White	0.030
Mix: Asian-Amer, White	0.018
Mix: Hispanic, White	0.015
Mix: African-Amer, Asian-Amer	0.032
Mix: African-Amer, Hispanic	0.036
Mix: Asian-Amer, Hispanic	0.022

Improving Estimated Haplotype distributions

To estimate the distribution of HLA-types among persons of mixed race, we have used haplotype distributions estimated by Kollman and Mori assumed that individuals always marry within their racial group. Yet we use these estimated haplotype distributions to estimate the HLA-types of persons whose parents did not marry within their race. The Kollman and Mori haplotype distributions for a race could be distorted if a large proportion of those who declare themselves to be of that race actually have parents of two different races.

The U.S. census for 2000 for the first time allowed individuals to declare themselves to be of more than one race. In previous censuses, one had to choose a single race. In the 1990 U.S. census, parents were asked to designate the race of their children, where "mixed race" was not an admissible answer. Zhenchao

Qian [12] examined this census data to estimate the fraction of children of mixed-race marriages who were declared to be of each of the parental races. He found that approximately 60% of children of and African-American and a white American designated their children as African-American and about 40% of children of an Asian-American and a white American designated their children as Asian-American.

In the census of 2000, the number of persons who declared themselves to be white and some other race was less than half of one percent of the white population. The number of persons who declared themselves to be African-American and some other race was about 3.5 percent of the African-American population. The number of persons who declared themselves to be Asian-American and some other race was almost 12 percent of the Asian-American population.

These facts suggest that the assumption of within-race marriage is unlikely to significantly distort estimates of the haplotype distributions for whites or for African-Americans.[8] The estimates for Asian-Americans are more problematic. In addition to the prevalence of marriage to non-Asians, it seems clear that marriage patterns among the parents of the current generation of Asian-Americans were far from homogeneous. According to the U.S. census, the distribution of national origins of the current Asian-American population of the U.S. is as follows:

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

These Asian populations have been geographically separated for many generations with little mating between them. Available data from many studies indicate that the distribution of HLA-types is quite different in these distinct populations [4]. Pritchard, Stephens, and Donnelly [11] study a problem that is similar to this, which they refer to as one of "cryptic" population structure. Assuming that mating is random within sub-populations, the authors propose a

[8] Add some discussion of the genetic evidence in [16].

Bayesian statistical method for extracting from aggregated data on genotypes, the distribution of alleles in sub-populations and the proportions of the total population belonging to each subpopulation.

The problem studied by Pritchard *et al* differs from the one that we pose in some ways. They study distribution of alleles at many separate unlinked loci, while Mori and Kollman work with observations of six alleles-phenotypes with the alleles located in three tightly linked, but extremely polymorphic, loci. We have some additional information that can inform our priors. For example, we know the approximate proportions of the Asian-American population who belong to distinct groups, as stated in the table above. These could be used to set prior probability distributions on the proportions of the registry who belong to each group. There also exist several studies that estimate the distribution of HLA-haplotype distributions in relatively small samples from localized regions. Examples include estimates for Korea [6], China[15] and [14], Taiwan [13]) and India [1] for several regions of Europe [3], for France [10], for Sardinia [8] and for Wales [5]. The Proceedings of the Twelfth International Histocompatibility Workshop has published lists of estimated two and three locus HLA haplotype distributions for quite narrowly-defined national and regional populations throughout the world. [4] These estimates are usually based on samples of only a few hundred individuals and hence can provide reliable information only about the distributions of the most common haplotypes. However, it is possible that they could be used in conjunction with large sample phenotype distributions such as the NMDP or other large national bone marrow registries to estimate the distribution of haplotypes in localized populations. The intuition here is that if one observes a phenotype that is likely to be the union of haplotypes h and h' where h is known to be common in population X and not elsewhere, then this is a hint that haplotype h' is more common in population X than elsewhere.

It is likely that the Mori-Kollman style estimates of haplotype distributions for coarse grained populations could be improved by means of an approach similar to that of Pritchard *et al* so as to produce estimated haplotype distributions for at least major subpopulations of the Asian group. Similar results might make it possible to separate subpopulations from the self-declared "white" and Hispanic groups.

References

[1] Suraksha Agrawal, K. Arundhati, Uddalak Bhardwaj, and Suhasini Bhatnagar. HLA antigen and haplotype frequency in Bhargavas and Chaturvedies of UP (India). *Indian Journal of Human Genetics*, 5(1):25–30, 1999.

[2] Theodore C. Bergstrom, Rodney Garratt, and Damien Sheehan-Connor. One chance in a million: Altruism and the bone marrow registry. to appear in *American Economic Review* currently available at http://repositories.cdlib.org/ucsbecon/dwp/3-07/, September 2007.

[3] 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.

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

[5] C. Darke, M.G. Guttridge, J. Thompson, S. McNamara, and J. Streetand M. Thomas. Hla class i (A, B) and ii (DR, DQ) gene and haplotype frequencies in blood donors from Wales. *Experimental and Clinical Immunogenetics*, 15(2):69–83, 1998.

[6] Se Jong Kim, In Hong Choi, and Jeu Deuk Kim. HLA-DR antigens and hla-b:dr haplotypes in Koreans. *Yonsei Medical Journal*, 24(1):34–37, 1983.

[7] Craig Kollman, Martin Maiers, Loren Gragert, Carlheinz Müller, Michele Setterholm, Machteld Oudshoorn, and Carolyn Katovich Hurley. 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.

[8] Rosanna Lampis, Laura Morelli, Mauro Congia, Maria Doloretta Macisand Annapaola Mulargia, Miriam Loddo, Stefano De Virgiliis, Maria Giovanna Marrosu, John A. Todd, and Francesco Cucca1. The inter-regional distribution of HLA class ii haplotypes indicates the suitability of the Sardinian population for casecontrol association studies in complex diseases. *Human Molecular Genetics*, 9(20):2959–2965, 2000.

[9] 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.

[10] B. Pedrona, K. Yakoubenb, D. Adjaoudb, A. Auvrignonc, J. Landmanc, V. Guerina, G. Levergerc, E. Vilmerc, and G. Sterkers. Listing of common HLA alleles and haplotypes based on the study of 356 families residing in the Paris, France, area: Implications for unrelated hematopoietic stem cell donor selection. *Human Immunology*, 66(6):720–730, June 2005.

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

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

[13] 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.

[14] Chunmei Shen, Bofeng Zhu, Mengli Liu, and Shengbin L. Genetic polymorphisms at HLA-A, -B, and -DRB1 loci in Han population of Xian City in China. *Croatian Medical Journal*, 49(4):476–482, August 2008.

[15] Li Shi, Saeko Ogatab, Jian Kun Yua, Jun Ohashic, Liang Yua, Lei Shia, Hao Suna, Keqin Lina, Xiao Qin Huanga andMasaki Matsushitad, Satoshi Horaie, Masaaki Muramatsub, Jia You Chua, and Katsushi Tokunagac. Distribution of hla alleles and haplotypes in jinuo and wa populations in southwest china. *Human Immunology*, 69(1):58–65, January 2008.

[16] Mark D. Shriver, Estaban J. Parra, Sonia Dos, Carolina Bonilla, Heather Norton, Celina Jovel, Carrie Pfaff, Cecily Jones, Aisha Massac, Neil Cameron, Archie Baron, Tabitha Jackson, George Argyropolulos, Li Jin, Clive J. Hoggart, Paul M. McKeigue, and Rick A. Kittles. Skin pigmentation, biogeographical ancestry and admixture mapping. *Human Genetics*, 112:387–399, 2003.