Biol Blood Marrow Transplant 22 (2016) 410–417

Biology of Blood and Marrow Transplantation

journal homepage: www.bbmt.org

On Modeling Human Leukocyte Antigen–Identical Sibling Match Probability for Allogeneic Hematopoietic Cell Transplantation: Estimating the Need for an Unrelated Donor Source

Kelsey Besse [1,*], Martin Maiers [1,2], Dennis Confer [1,2], Mark Albrecht [1]

[1] Bioinformatics Research, National Marrow Donor Program, Minneapolis, Minnesota
[2] Center for International Blood and Marrow Transplant Research, Minneapolis, Minnesota

Article history:
Received 18 April 2015
Accepted 10 September 2015

Key Words:
Allogeneic
Hematopoietic cell transplantation
Sibling
Match
Probability
Donor

A B S T R A C T

Prior studies of allogeneic hematopoietic cell transplantation (HCT) therapy for the treatment of malignant or nonmalignant blood disorders assume a 30% likelihood that a patient will find a match among siblings and, therefore, a 70% likelihood of needing an unrelated donor source. This study utilizes birth data and statistical modeling to assess the adequacy of these estimates to describe the probability among US population cohorts segmented by race/ethnicity and age, including ages of greatest HCT utilization. Considerable variation in the likelihood of an HLA-identical sibling was found, ranging from 13% to 51%, depending upon patient age and race/ethnicity. Low sibling match probability, compounded with increased genetic diversity and lower availability among unrelated donors, put the youngest minority patients at the greatest risk for not finding a suitable related or unrelated HCT donor. Furthermore, the present 40-year decline in birth rates is expected to lead to 1.5-fold decrease in access to a matched sibling for today's young adults (18 to 44 years of age) when they reach peak HCT utilization years (near age 61 years) versus their contemporary adult counterparts (44 to 64 years). Understanding the sibling match probability by race/ethnicity and age cohort leads to forecasting the demand for unrelated HCT sources.

INTRODUCTION

An HLA-matched donor is required for allogeneic (allo) hematopoietic cell transplantation (HCT), a potentially curative therapy for the more than 10,000 patients estimated to be eligible for the treatment per year [1]. For a patient seeking alloHCT treatment, each full sibling has a 25% chance of being an HLA-identical match. A greater number of full siblings, therefore, increases the likelihood a patient will have a match among siblings. Based on an average of 2 to 3 children per family, a commonly used estimate is that an individual in the United States has a 30% likelihood of having a full HLA-matched sibling [1-6]. Further, this chance, referred to henceforth as *sibling match probability*, has been used to describe the potential 70% likelihood that a patient seeking an HLA match will need to rely on an unrelated donor or umbilical cord blood rather than a sibling [1,7].

However, the sibling match probability in the context of the current transplantation-eligible population(s) in the United States has not been formally assessed. Consideration of observed changes in female fertility across generations and variations among race/ethnic groups [8,9] would be expected to provide a more reliable description of the expected number of siblings and, therefore, the likelihood of a sibling match. An accurate estimate of the sibling match probability by subpopulations, including those with the highest likelihood of transplantation use, is an important component of forecasting the demand for unrelated alloHCT donation and can be used to improve previous estimates [1]. Efforts that lend to a more detailed picture of the need for unrelated donors by race/ethnic group can also enhance strategic planning and assist in monitoring progress towards the goal of eliminating racial and ethnic barriers to HCT [10]. Additionally, this information can help to better characterize the likelihood of sibling matches for groups of patients diagnosed with diseases with a high incidence rate

Financial disclosure: See Acknowledgments on page 416.

* Correspondence and reprint requests: Kelsey Besse, National Marrow Donor Program 3001 Broadway Street NE, Suite 100 Minneapolis, MN 55416.
E-mail address: kbesse@nmdp.org (K. Besse).

http://dx.doi.org/10.1016/j.bbmt.2015.09.012
1083-8791/

among specific age and/or race/ethnic groups (eg, acute myeloid leukemia: more common in older adults ages 60 years and older; acute lymphocytic leukemia: more common among children under 15 years [11]; sickle cell disease: more common in African Americans [12]). Such information can be useful in informing researchers and the transplantation community of the likely need for an unrelated donor in these disease-specific populations. Overall, such information contributes to better estimating the demand for alloHCT, assessing the usage of the treatment [1], establishing the ideal composition of potential donors to make up the Be the Match Registry, and adding to larger ongoing efforts to successfully serve all patients in need of cellular transplantation.

Using birth data from the National Center for Health Statistics via the National Vital Statistics Reports for the years 1940 to 2012, we estimate the number of siblings and calculate the expected sibling match probability among US subpopulations characterized by age and race/ethnicity and weighted by recent HCT utilization. The distribution of female parity in conjunction with age and race/ethnicity-specific mortality rates was incorporated to improve the accuracy of these estimates by accounting for the number of living siblings, either those that have recently been born (among mothers of fertile ages) or those still living; most relevant to the younger (ages 0 to 19) and older adult (ages 65 to 80) patients, respectively.

METHODS

HLA-Identical Sibling Match Rate Model

Modeling sibling match rate for the US population requires extending the commonly used formula describing number of HLA-identical siblings (matches), Q, for a given patient with known number, n, of related siblings to the population level where n is no longer fixed and must be treated as a random variable, N (capital letters are random variables, lower case letters are fixed). When n is known, $Q \sim \text{bin}(n, p = 1/4)$ where $p = 1/4$ is the rate parameter describing the probability that each sibling is an HLA-identical match. Thus, for a patient having a known number of related siblings, n, the HLA-identical sibling match rate is the expected value of an indicator function $I_{Q\geq 1}$ the distribution of Q

$$E_Q(I_{Q\geq 1}) = I_{Q\geq 1}\cdot[Q = 0] + I_{Q\geq 1}\cdot[Q \geq 1] = 1 - \left(\frac{3}{4}\right)^n \quad (1)$$

where "•" denotes the probability mass function of a discrete random variable.

Unfortunately, the sibling match rate for the US population cannot be described by inserting the expected family size into Equation 1, given that N is a random variable and Jensen's inequality [13] states

$$E_{Q,N}(I_Q \geq 1) \neq 1 - \left(\frac{3}{4}\right)^{E(N)} \quad (2)$$

since match rate is a nonlinear function of $E(N)$. Instead, to derive a valid quantity for $E_{Q,N}(I_{Q\geq 1})$, we must compute the expected value of the nonlinear function

$$E_{Q,N}(I_{Q\geq 1}) = E_N\left(1 - \left(\frac{3}{4}\right)^N\right) = \sum_N \left(1 - \left(\frac{3}{4}\right)^n\right)[N = n] \quad (3)$$

with respect to N. Full details of the methodology are provided in the Supplemental Materials (S1), and a brief description of the key elements follows.

Define the distribution of related siblings, N, available for transplantation

An individual's age determines how many potential siblings are alive and available, based upon the competing factors of birth and mortality; an individual's race/ethnicity provides information regarding family size based upon race/ethnicity-specific differences in birth rates. We, therefore, constructed a model for N that is age and race/ethnicity specific (see Table S1 for data set age and race/ethnicity details) based upon total fertility rate (TFR) data describing the birthing distribution of mothers in the US population (shown in Figure 1) and survival curves reflecting sibling mortality (shown in Figure 2).

Model expected sibling match rate (and expected family size)

Based on the assumptions inherent in N, computing $E_{Q,N}(I_{Q\geq 1})$ in Equation 3 involved taking the integral of a mixture Bernoulli distribution, for which we relied upon Monte Carlo methods for approximation given the lack of closed form expressions.

Donor Demand

Data on patient allogenic HCT utilization from Center for International Blood and Marrow Transplant Research were collected to establish

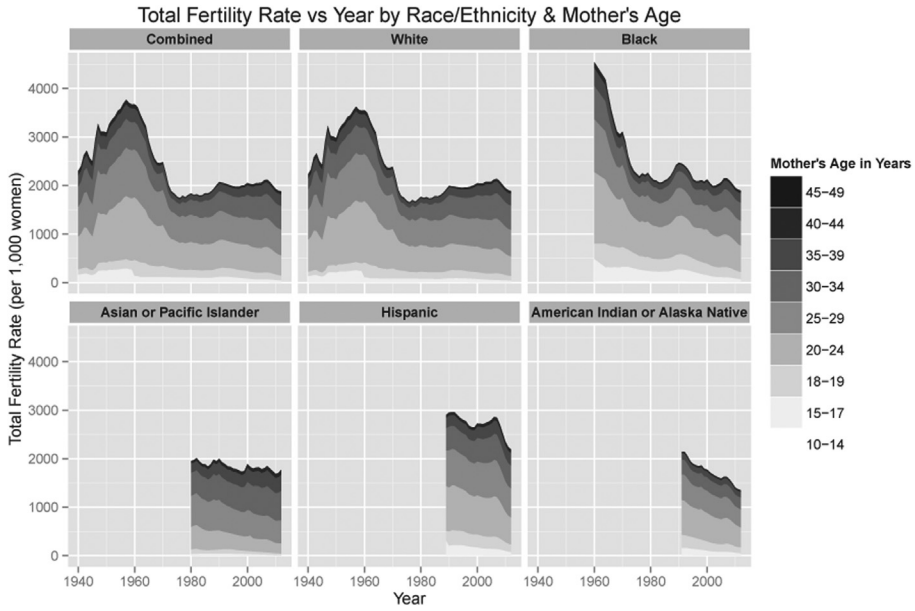

Figure 1. Total fertility rate (TFR) conditional on maternal age, race, and year obtained from National Vital Statistics Reports via the National Center for Health Statistics [8,9]. Partial components of fertility rate are scaled 1000-fold and multiplied by the age interval (ie, $1000 * \pi_{a,r,y} * \Delta a$) to align their units with those of TFR. Gaps in year coverage represent unequal data collection periods for different maternal race and ethnic groups.