Working Paper Series n° 623

May, 2010

"A Mathematical Model for Bone Marrow Donors' Registries and Cord Blood Banks"

Frédérique FÈVE and
Jean-Pierre FLORENS

Institut d'Economie Industrielle (IDEI) – Manufacture des Tabacs
Aile Jean-Jacques Laffont – 21, allée de Brienne – 31000 TOULOUSE – FRANCE
Tél. + 33(0)5 61 12 85 89 – Fax + 33(0)5 61 12 86 37 – **www.idei.fr** – contact@cict.fr

INSTITUT D'ECONOMIE INDUSTRIELLE

A Mathematical Model for Bone Marrow Donors' Registries and Cord Blood Banks*

F. Fève and J.P. Florens

Toulouse School of Economics

May 2010

*This work was supported by the European contract MADO No: QLG7-2001-00065) and EFG (Etablissement Français des Greffes). The authors thank Anne Cambon-Thomsen (INSERM U558, Toulouse) for initially raising the question, FGM (France Greffe de Moelle) and the French histocompatibility laboratories for providing HLA data, Gregory Katz–Benichou (ESSEC, France) for helpful comments about cord blood banks. They are grateful of Fabrice Collard, Pierre Dubois, Jean–François Eliaou, Patrick Fève and Thomas Mariotti for helpful discussions and comments.

Address of the authors: IDEI, Université Toulouse 1 Sciences Sociales, Manufacture des Tabacs - Bât F, 31000 Toulouse, France, Tel: +33(0)5.61.12.85.90, Fax: +33(0)5.61.12.86.37, email: feve@cict.fr, florens@cict.fr

Abstract

This paper constructs an economic analysis of some organ donation organizations. The two main examples are voluntary marrow donor registries and cord blood banks. The main characteristic of this system is to facilitate the graft of bone marrow or cord blood to patients. These grafts require a high degree of compatibility between donors and receivers and the efficiency of this system is not always satisfactory despite sizes of the registries. This paper gives a framework to understand the key parameters of this problem and to proceed to simulations. We consider the case without screening or the case of optimal selection. These models may be used to infer an economic evaluation of the registries and of cord blood banks.

Keywords: Decision theory, evaluation of information system, marrow donors registry.

JEL Classification: I18, C70, C61.

1 Introduction and Motivations

The treatment of certain blood diseases (leukaemia in particular) sometimes requires the appeal to a transplant of HSC (Hematopoietic Stem Cell, known as " bone marrow"). This requires the use of a compatible donor and a donor family member of the recipient is sought on a priority basis. Such a compatible donor family exists only in 30% of cases. To treat other patients, a record of voluntary donors was organized in some countries. The first registry was established in United Kingdom (Anthony Nolan [1]). A registry consists in a list of volunteers individuals for whom a number of genetic characteristics determining compatibility are recorded. For a given patient, we search through the registry if one or more donors exist. If so, a more precise analysis of the compatibility will be conducted and whether the customer is always voluntary graft will be possible. It will be achieved only in the light of the state of health of the recipient but this is beyond the scope of this paper. The compatibility "principal recipient" can be summarized in the identity of the HLA (Human Leucocytes Antigens) system characterized by a set of genes on the $6th$ chromosome. Each individual has a pair of these chromosomes. On each of them a set of genetic positions (loci) could take several modalities (alleles) and two people are compatible if their alleles coincide at each locus. This system can be encoded at multiple levels of detail (choice of the number of loci-3 to 6 - and precision tracking alleles - among 20 to 40-per locus). But even at a level of precision enough roughage, then the number of possible types is immense. This great heterogeneity of HLA types explains the low efficiency of the system registry. More than 130,000 donors are recorded in the French registry and less than 10% of the patients involved (approximately 1,000 per year) finds a donor perfectly compatible and available. It is therefore natural to ask how to improve the effectiveness of the system. Two approaches are possible: either swell the size of the registry by recruiting more donors (a doubling in 10 years is scheduled in France) or by developing the international connection of registries or select the best donors. Indeed the size of the registry is constrained both by the cost of typing and the limited budget of the association in charge of the registry. The selection of donors is difficult: it could be based on observable variables (birthplace), or on genetic testing, but it is necessary before considering such an approach to study potential improvements.

[1] www.anthonynolan.org.uk

To address these two issues, we develop a mathematical model of registry system that simulates the effect of the increase and the maximum effect of selection under reasonable assumptions based on the distribution of types in the population. Our results show that, unfortunately, only a very significant increase in the registry would be useful and selection procedures would be marginal effectiveness on the system's efficiency.
The process of selecting a donor in a register to satisfy a particular recipient comprises two stages. The register contains a set of individuals and their phenotypes. It tries in the register all donors who have the same type A, B, DR low resolution (Speiser et al., 1994) that the receiver (Sonnenberg et al., 1994). If no compatible donor is the process stops (Oudshoorn et al. 1997). If further examinations are carried out on donors (including the search for effective availability of donor), so as to refine the compatibility test. If a fully compatible donor is found, the transplant can take place, provided that the patient's condition allows. It was thus two levels of compatibility criteria: a first level that is revealed by the register and a second level which is unknown until the donor is considered for a particular recipient.

The division between the two types of compatibility poses a relevant economic issue which will be simply referred to here, so as to clarify the interpretation of parameters. It is clear that all the criteria of compatibility or availability varying over time (incompatibility resulting from the fact that the donor himself may be sick or unavailability of a donor due to pregnancy, for example) can only part of the second level (unregistered and revealed during the search.

However fixed criteria (sex, HLA typing at high resolution, Speiser et al.,1994) can be completed in the registry (at a cost of typing high, therefore a smaller size of the register given budget) or not all be sought at the time of registration.

The number of new registered donors each year in the registry is determined by a budget. The price of typing is expressed by a rate which has been evaluated and that is the price paid to laboratories responsible for this benefit. The number of volunteers could exceed the number of individuals that may be typed or may not correspond to the territorial distribution sought. The level of dissemination of recruitment campaigns trying to bring closer the number of volunteers searched over a given period (depending on the available budget) and the number of volunteers who present themselves effectively. The 34 french centers approved nationally and internationally (accreditation by the WMDA and the European Federation for Immunogenetics (EFI))are responsible for the

recruitment of donors and are responsible for their HLA typing and their inclusion on the national registry. For example in France, they receive a lump sum of 183€ by typing in return. The techniques vary typing laboratories. All laboratories are subject to quality control organized by the AFSSAPS (French Agency for Health Safety of Health Products). In the case of a hospital structure, which carries out typages register for the french for example, increasing returns to scale mean that the average cost in the long term is decreasing: there is economies of scale and marginal cost is still below the average cost. The optimum pricing which would equalize the prices and the marginal cost then inevitably lead to a deficit of the hospital structure. The financing of typing therefore funding fixed costs of hospitals in charge of typing, by the powers of public funds. Subsidies aimed at eliminating the deficit of a public monopoly (which is the case of the hospital here) can not always be put in place (especially long-term), even if the deficit is justified by the criterion of 'optimality collective that represents the marginal cost pricing. It is often more reasonable to assume that the public monopoly is required to comply with a balanced budget constraint: finance production costs by income at least equivalent becomes a constraint that should be taken into account to set the tariff policy. It is then led to define pricing that maximizes the collective surplus, or welfare, under this additional constraint is a balanced budget. The collective surplus is a measure of the net benefit provided by the constitution and management of a register of bone marrow donors (CSH).

This article focuses on this model and only considers the economic aspects (see Fève et al 2007) in the conclusion and it does not analyse the statistical inference on distributions types (see Fève 2006). We believe the effectiveness in terms of probability of finding a donor for a recipient without any objective equity between social groups. (with same probabilities considered for the various ethnic minorities). Moreover, it will focus on a model for a single country, although in reality the files of donors are interconnected. Our model could nevertheless be applied throughout the world with an appropriate adjustment parameters.

Under these hypotheses, our model shows that, for increasing the register has an impact within the meaning of our efficiency criterion, it is necessary that it is massive. Also optimize the register has little impact (optimization would be optimal in eliminating donors rare). The equality between ethnic groups

also had little impact. It should ensure that the mismatch is possible, whatever the type of transplant, grafting haematopoietic stem cells (CSH) or from cord blood. This article focuses on CSH but it can be easily applied to other bank management: the cord blood banks pose problems of a similar nature. In cord blood banks, cord bloods are preserved during long time and are recorded with their types. The main differences between voluntary donors registries and cord blood banks are the following. The number of cord bloods in the bank is small (5 000 in France) but the types are recorded at a high precision and if a cord blood is compatible at the level of the recording in the bank, the probability to realize the graft is very high. The main interest of cord blood transplantation is that the perfect matching between donor and receiver is not required and some mismatches are accepted. To cover this case we extend our model in section 6 to the case of graft with mismatches. In that case the model may cover voluntary donors registries and cord blood banks simultaneously by suitable choice of parameters.

Our paper is organized in the following way: section 2 defines the model of registry without mismatches and optimal registry is analysed in that case in section 3. In section 4 we perform some simulations experiments derived from the French data. In section 5 we introduce some equality constraints between ethnic groups.Section 6 extends our approach to mismatch with empirical analysis of the cord blood banks. In a last section, we briefly discuss the economical implications of our study.

Different previous analysis have considered the matching probabilities for bone marrow donors registries. Most of these papers have an empirical approach based on a set of observations (see Ottinger et al. 1994, Oudshoorn et al. 1997, Rendine et al. 1999, Schipper et al. 1996, Speiser et al. 1994, Tiercy et al. 2000) or propose algorithms to determine these probabilities (Hoffman-Smith 1993, Muller 2002, Schuler et al. 2000, Takahashi 1989, Kollman et al. 2004, Sonnenberg et al. 1989). The case of minorities has also been treated in few papers (see Anderson et al. 1996).

Numerous references exist on the number and the distribution of HLA in the population (see Hurley et al. 1997, Lonjou et al. 1995, Mori et al. 1997 for example).

2 Donors, receivers and registry

Each individual of the population possesses a set of characteristics and a graft is possible between individuals having the same characteristics. We call type the part of these characteristics which is recorded in the registry of voluntary donors. For example, the type may be the HLA phenotype restricted to a set of loci on the two chromosomes. The more frequent case for the type is to consider the A, B and DR loci and to record the alleles at a low precision level ("two digits" coding). A registry is defined by the list of recorded types $1, ..., J$. The receivers are assumed to be drawn in the population where the frequencies of types are described by a probability vector p_j ($p_j > 0, \sum_{j=1}^{J} p_j = 1$). By construction of the list of the types all the p_j are strictly positive. Two conditions are necessary in order to do a transplant to a given receiver :

i, The existence in the registry of individual of the same type. We assume first that only perfect matching transplants are realized. Such a hypothesis will be relaxed in section 5.

ii, The compatibility at the level of coding in the registry is not sufficient: more compatibility tests should be done. Moreover, some donors may also be not available (pregnancy, professional requirements, illness...). We summarize these complex phenomena by assuming that a donor compatible with a receiver of type j has a probability a_j to be fully compatible and to accept a transplant. Full compatibility and acceptation for different donors are independent events. In general we consider cases where a depends on j. Indeed if the typing is done in the registry at a low precision, a graft will be done only under high precision compatibility. The number of possible high precision types corresponding to a low precision type depends on the type and usually determines a specific to j.

Remark 1: The level of recorded types determines jointly the list (and the probabilities) of types and the a_j's values. A low level of typing will generate a "small" number J of types and small a_j's values. Reciprocally a very precise typing in the registry ("Four digits" and more alleles) increases J and determines a_j's probabilities closer to 1. The decision between levels of typing in the registry is an economic decision : the cost of typing is higher for a precise typing but the search of a donor is more efficient and increases the value of the registry.

A registry design is defined by two components :

1. an initial registry : this initial registry is characterized by its size N_0 and by the number N_{0j} of donors of type j. This number may be equal to 0 for many types.

2. an increment process defined by the number N of new donors introduced in the registry and by a sampling mechanism of the types described by a vector q_j ($q_j \geq 0, \sum_{j=1}^{J} q_j = 1$) of frequencies. For example if donors and receivers are drawn randomly in the same population $p_j = q_j$ and are equal to the frequency of type j in the population.

We should underline that individuals in the population and registry management ignore the types. Typing is a complex operation only realized where a new donor is introduced in the registry. Then, in practice, the registry management cannot choose q_j. However we first imagine situation where $(q_j)_j$ can be selected arbitrarily by the registry management and we may imagine optimal choice for the q_j. This optimal q_j mechanisms will give an upper bound for the efficiency of the registry. In practice the policy of the registry management may consist in the implementation of screening procedure for the selection of the donors. These screening mechanism are not considered here (see Fève 2006). In the paper we will analyse different values of the level N of the increment of the registry but this value may be constrained by the arrival process of voluntary or by economic considerations (see Fève et alii, 2007).

We consider in this paper only a single period model : starting for an initial registry we consider its improvement by a unique increment and not by several increments through multiple periods. This multiple period problem is a dynamic programming problem and more mathematical tools are needed and are beyond the scope of this paper.

As an illustration consider the French registry of bone marrow donors. The types are defined by HLA haplotypes A,B,DR recorded at low precision. The current registry contains approximatively 130 000 donors and an increment of 10000 by year is scheduled. If we want to analyze the one year mechanism N is 10 000 but we may also consider a long term variation $N = 100\,000$ or more. The number J of possible types is a crucial element of the analysis and will be discussed later on. More than 66 000 different types are present in the registry. At the world level the interconnection between the national registries gives a

total registry of more than 11 millions which also increases by several hundred of thousands people each year. More than 400 000 types have been observed.[2]

Définition 1: A registry system is defined by the list types $\{1, ..., J\}$, their probability in the population p_j, the values $a'_j s$, the initial registry, the size N of the increment and the sampling probabilities $q'_j s$.

We propose to evaluate a registry by the expected probability to find a donor for a receiver. To illustrate this concept consider just the simple case where $N_0 = 0$ (no stock) and $a = 1$ (all compatible donors are fully compatible). For any receiver the non-realization of a transplant (all the donors have a type different of j) is a random event which has a probability of $(1 - q_j)^N$ if the type of the receiver is j. As the types of the future receivers are not given when the registry is designed we consider the expectation of these probabilities through the different types :

$$\pi = 1 - \sum_{j=1}^{J} p_j (1 - q_j)^N.$$

This quantity may be viewed as the value of the registry system.

Remark 2: An alternative criterium for the evaluation of the registry design would be based on the expected waiting time of a patient. Let us assume that N new donors are drawn with probabilities $(q_j)_{j=1,...,J}$ at each period. For a patient where type is j, the waiting time is 0 with probability $1 - (1 - q_j)^N, 1$ with a probability of $(1 - q_j)^N (1 - (1 - q_j)^N)$. In general the waiting time is t with a probability of $(1 - q_j)^{Nt} (1 - (1 - q_j)^N)$. The expected waiting time of this Pascal distribution is equal to $\frac{1}{1-(1-q_j)^N}$. We average this expected time with respect to the patient's type and we get the following evaluation criterion:

$$L = \sum_{j=1}^{J} p_j \frac{1}{1 - (1 - q_j)^N}.$$

Remark 3: We may also introduce a value A for finding a donor and of cost B if a donor is not found. In that case the evaluation of the registry in our simple

[2]The list of possible alleles on each locus A,B, DR is probably known and then defines a huge list of potential types. However most of associations don't exist and the number of sequences A,B, DR is smaller than the product of the number of alleles on each locus.

case ($N_0 = 0 \quad a_j = 1 \quad \forall_j$) becomes :

$$\sum_j p_j \left\{ A[1-(1-q_j)^N] - B(1-q_j)^N \right\}.$$

We concentrate our analysis to the first criterium (the probability to find a donor) and this probability will be evaluate in the general case.

Proposition 1: For N large and q_j small, the value of the registry may be approximated by :

$$\pi = 1 - \sum_{j=1}^{J} p_j (1-a_j)^{N_{0j}} e^{-a_j N q_j}.$$

Proof : let fix j the type of a receiver. The number of donors of this type is $N_{0j} + m_j$ where m_j is drawn by a Binomial distribution :

$$Prob(m_j) = \left(\begin{array}{c} m_j \\ N \end{array} \right) q_j^{m_j} (1-q_j)^{N-m_j}.$$

Given j and m_j the probability of finding no donor is $(1-a_j)^{N_{0j}+m_j}$. Then given j only, the probability to not find a donor is

$$\sum_{m_j=0}^{N} (1-a_j)^{N_0 j + m_j} \left(\begin{array}{c} m_j \\ N \end{array} \right) q_j^{m_j} (1-q_j)^{N-m_j}.$$

Then

$$\pi = 1 - \sum_{j=1}^{J} p_j \sum_{m_j=0}^{N} (1-a_j)^{N_{0j}+m_j} \left(\begin{array}{c} m_j \\ N \end{array} \right) q_j^{m_j} (1-q_j)^{N-m_j}$$

$$\pi = 1 - \sum_{j=1}^{J} p_j (1-a_j)^{N_{0j}} E[(1-a_j)^{m_j}],$$

where m_j is drawn by the binomial distribution. For large N and small q_j it is wellknown that this Binomial distribution is approximatively a Poisson distribution parametrized by $\lambda_j = N q_j$. If X is generated by a Poisson distribution of parameter λ, then $E(b^X) = e^{\lambda(b-1)}$. Then :

$$\pi = 1 - \sum_{j=1}^{J} p_j (1-a_j)^{N_{0j}} e^{-a_j N q_j}.$$

□

We will use in section 4 this formula to evaluate this probability in some particular examples.

3 Optimal registry

The problem then reduces to maximize the evaluation criterion with respect to the drawing design of the donors $(q_j)_{j=1,...,J}$ for any given value of N. This analysis has only a theoretical objective because the result will require to be implemented the knowledge of the types. The result has however an interest as a reference theoretical optimal registry.

Equivalently the problem is to minimize

$$1 - \pi = \sum_{j=1}^{J} p_j (1-a)^{N_{0j}} e^{-aNq_j}$$

with respect to the q_j's under the constraints:

$$\sum_{j=1}^{J} q_j = 1 \text{ and } q_j \geq 0 \; \forall j = 1, ..., J.$$

Proposition 2 : Let $r_j = a_j p_j (1-a_j)^{N_{oj}}$. We assume that the types $j \in \{1, \ldots, J\}$ are ranked such that the r_j are non increasing. Then there exists $J_o \leq J$ such that :

i) The optimal selection mechanism verifies :

$$q_j^0 = 0 \quad \text{if} \quad j > J_0$$
$$q_j^0 = \frac{\frac{1}{a_j}}{\sum_{j=1}^{J_0} \frac{1}{a_j}} + \frac{1}{a_j N} \{\ln a_j - \ln \bar{a} + \ln r_j - \ln \bar{r} + \ln s_j - \ln \bar{s}\}$$

where

$$\ln \bar{a} = \frac{\sum_{j=1}^{J_0} \frac{1}{a_j} \ln a_j}{\sum_{j=1}^{J_0} \frac{1}{a_j}} \qquad \ln \bar{r} = \frac{\sum_{j=1}^{J_0} \frac{1}{a_j} \ln r_j}{\sum_{j=1}^{J_0} \frac{1}{a_j}}$$

$$s_j = (1 - a_j) \qquad \ln \bar{s} = \frac{\sum_{j=1}^{J_0} \frac{1}{a_j} \ln s_j}{\sum_{j=1}^{J_0} \frac{1}{a_j}}$$

($\bar{a}, \bar{r}$ and $\bar{s}$ are the weighted geometric means of the a_j, r_j and s_j)

ii) The optimal probability to find a donor is

$$\pi^0 = 1 - \left\{ \bar{a}\ \bar{p}\ \bar{s} \sum_{j=1}^{J_0} \frac{1}{a_j} e^{-\frac{N}{\sum_{j=1}^{J_0} \frac{1}{a_j}}} + \sum_{j=J_0+1}^{J} p_j (1-a_j)^{N_{oj}} \right\}$$

Proof: We consider the optimization problem:

$$\min_{(q_j)_j} \sum_{j=1}^{J} p_j (1-a_j)^{N_{oj}} e^{-a_j N q_j}$$

$$\text{under } \sum_{j=1}^{J} q_j = 1 \text{ and } q_j \geq 0.$$

Consider the Lagrangian :

$$\varphi(q) = \sum_{j=1}^{J} p_j (1-a_j)^{N_{oj}} e^{-a_j N q_j} + \rho \left(\sum_{j=1}^{J} q_j - 1 \right)$$

The usual Kuhn and Tucker method leads to the conditions :

$$\frac{\partial \varphi}{\partial q_j} = -a_j N_{p_j} (1-a_j)^{N_{oj}} e^{-a_j N q_j} + \rho = 0 \quad \text{if} \quad q_j \geq 0$$

$$\geq 0 \quad \text{else}$$

The first equality is equivalent to :

$$q_j = \frac{1}{a_j N} ln \frac{a_j p_j (1-a_j)^{N_{oj}}}{\rho}$$

and $q_j \geq 0$ if and only if $r_j = a_j p_j (1-a_j)^{N_{oj}} \geq \frac{\lambda}{N}$

As r_j is non increasing this is equivalent to say that there exists a unique $J_0 \in I\!N$ such that $j \leq J_0 \leq J$ implies $q_j \geq 0$.

Moreover $j > J_0$ implies $q_j = 0$.

The condition $\sum_{j=1}^{J} q_j = 1$ is used to determine λ and the optimal solution for the q_j is :

$$\begin{cases} q_j^0 & = \frac{\frac{1}{a_j}}{\sum_{j=1}^{J_0} \frac{1}{a_j}} + \frac{1}{a_j N} \left\{ lnr_j - \frac{\sum_{j=1}^{J_0} \frac{1}{a_j} lnr_j}{\sum_{j=1}^{J_0} \frac{1}{a_j}} \right\} j \leq J_0 \\ q_j^0 & = 0 \quad j > J_0 \end{cases}$$

which leads to the formula given in the theorem.
The value of π^0 is obtained by replacing q_j by q_j^0.

□

The search of J_0 is simplified if we assume that the probabilities a_j do not depend on j ($a_j = a \quad \forall_j$). In that case the q_j^0 are non increasing (with the same ranking as above) because $q_j^0 = \frac{1}{aN} \ln r_j$ + constant (or 0 after J_0). So before the breakpoint J_0 the optimal selection probability is an affine transformation of the logarithm of $p_j(1-a)^{N_{oj}}$.

Select an arbitrary J_1 and construct $q_j(J_1)$ exactly as q_j^0 by replacing J_0 by J_1. The minimum value of the $q_j(J_1)$ is $q_{J_1}(J_1)$. Then we may increase J^1 until the two inequalities : $q_{J_1}(J_1) \geq 0$ and $q_{J_1+1}(J_1+1) < 0$. In that case J_0 is equal to J_1.

In order to illustrate this general result, we consider a version of the model where no initial registry is available and where $a_j = a \quad \forall j$. In that case the optimal sampling frequencies are:

$$\begin{cases} q_j = 0 & j \geq J_0 \\ q_j = \frac{1}{J_0} + \frac{1}{aN}\{\ln p_j - \frac{1}{J_0}\sum_{j=1}^{J_0} \ln p_j\} & j \leq J_0 \end{cases}$$

This optimal mechanism eliminates the rare types (remember that the j are ranked by decreasing frequencies). For the sampled types the q_j are equal to the uniform probability ($\frac{1}{J_0}$) plus a term depending on the difference between the logarithm of the p_j and the mean of the logarithms of the p_j. Frequent types (with a frequency larger than the geometrical mean) should be sampled with a superior probability than the uniform.
The value of the registry for the optimal sampling is equal to:

$$\pi^0 = 1 - \{J_0 \bar{p}\, e^{-\frac{aN}{J_0}} + \sum_{j>J_0} p_j\}$$

where $\bar{p}$ is the geometrical mean of the p_j truncated at J_0. Note that

$$\pi^0 \leq 1 - J\bar{\bar{p}}\, e^{-\frac{aN}{J}}$$

where $J\bar{\bar{p}}$ is the geometrical mean of all the p_j. This bound is obtained by maximizing π relaxing the positivity constraint on the q_j and is then larger than the maximum under constraints. This upper bound of the efficiency is a decreasing function of the number of types and increases with the size of the registry and with a. This dependance between N and π shows that π_0 increases slowly with N. Remember that this value dominates the efficiency for all the sampling mechanism for the donors. The value of $J\bar{\bar{p}}$ is an element of $[0, 1]$ (because the geometrical mean is smaller than the arithmetical mean of the p_j equal to $\frac{1}{J}$) and may be viewed as a measurement of the heterogeneity of the probabilities of the types. For populations with almost uniform distribution the types $J\bar{\bar{p}}$ is almost equal to 1 but for heterogenous populations $J\bar{\bar{p}}$ declines. Contrarily to some intuition the efficiency of the optimal registry is higher for heterogenous populations, because the optimal mechanism will sample in the frequent types.

4 Simulations

We illustrate our theoretical results by simulations experiments comparable to real situations. We essentially based our examples on the French registry system (France Greffe de Moelle) now managed by a public organization, l'Agence de Biomédecine. This registry records individuals' typings for three loci (A,B,DR) at a low precision (two digits). A type is then constituted by three pairs of numbers representing the alleles of the genes A,B and DR on the two chromosomes. At the end of 2003, 107 925 individuals were recorded, 66 164 different types are observed. Due to the large number of types the individuals values a_j are ignored and we will assume that $a_j = a$ for all j. A realistic value for a is $\frac{1}{3}$ ($\frac{1}{3}$ of donors compatible at the registry level are fully compatible and available for a graft. This choice is based on a qualitative opinion provided by France Greffe de Moelle Registry). The calibration of a registry model using real data is difficult. Indeed even if, for example, the French sample is large, its size remains small with respect to the number of types. Different models have been developed to estimate the number of HLA genotypes in the French population and an estimation of 500 000 seems valid (see Fève and Florens, 2007). The estimation of the characteristics of the distribution is also difficult and we prefer to calibrate the model with different scenarii and not to use a statistical estimation.
In order to compute the probability to find a donor for different sizes of the reg-

istry [3], we will make a mathematical tractable approximation of the problem. Even if the number of types is discrete, we assume a continuous distribution on the real positive numbers. Let us underline that we do not develop a registry model with a continuous number of types but we approximate the result obtained in discrete case by a continuous model. All the previous results remain valid where the sum are transformed into integrals. To simplify our example we consider cases where no initial registry is available. More precisely we assume that the types are randomly generated from an exponential distribution of density $\lambda\, e^{-\lambda j}$ ($j \in [0, +\infty[$). Then, π is given by:

$$\pi = 1 - \int_0^\infty \lambda e^{-\lambda j} e^{-aNq(j)} dj.$$

In case of no selection $q(j) = p(j)$ and this formulae becomes:

$$\pi = 1 - \int_0^\infty \lambda e^{-\lambda j} e^{-aN\lambda e^{-\lambda j}} dj$$

$$\pi = 1 - \frac{1}{aN\lambda}(1 - e^{-aN\lambda})$$

We need to evaluate λ. This will be done by fixing the median of the distribution. For example, if we assume that the X more frequent types represent 50% of the population we have $\lambda = -\frac{1}{X} \ln \frac{1}{2}$.

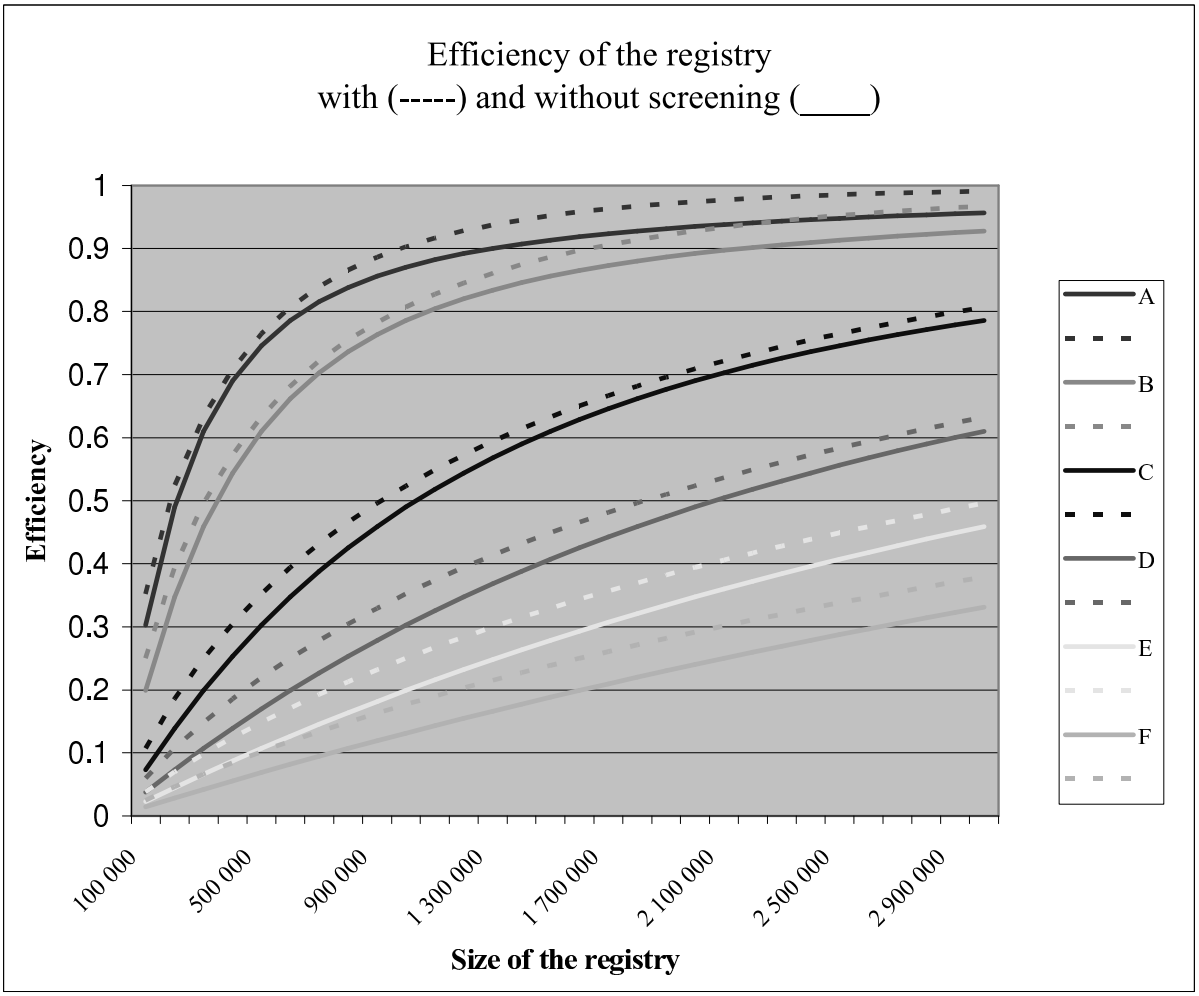
We consider six scenarii from $X = 30\,000$ to $X = 800\,000$

Scenario A	$X = 30\,000$
Scenario B	$X = 50\,000$
Scenario C	$X = 150\,000$
Scenario D	$X = 300\,000$
Scenario E	$X = 500\,000$
Scenario F	$X = 800\,000$

We then evaluate the efficiency of the registry with and without screening (see Figure 1).

[3] what was one of the objectives of the MADO project (2001-2005): Optimisation of typing policies for European hematopoïetic stem cell donor Registries: socioeconomic evaluation of molecular techniques and recruitment strategies.

Figure 1:


Efficiency of the registry
with (-----) and without screening (____)
Efficiency
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
100 000
500 000
900 000
1 300 000
1 700 000
2 100 000
2 500 000
2 900 000
Size of the registry
A
B
C
D
E
F

Under this exponential approximation the optimal sampling mechanism for the donors has the following structure. The first order condition of the maximization of π gives:

$$q_j = \frac{1}{aN} \ln \frac{aNp_j}{\rho}$$

if $q_j \geq 0$ where ρ is the Lagrange multiplier of the constraint $\int q(j)dj = 1$. The condition $p_j \geq 0$ is equivalent to $\frac{aNp_j}{\rho} \geq 1$ or to p_j greater to some value which is equivalent to $j \leq J_0$. We compute J_0 from the inequality:

$$q_j = \frac{1}{J_0} + \frac{1}{aN}\{\ln p_j - \int_0^{J_0} \ln p_j \, dj\} \geq 0$$

which gives J_0 solution of

$$J_0 \ \ln p_j - \int_0^{J_0} \ln \ p_j \, dj - aN = 0$$

or in the exponential case

$$J_0 = \sqrt{\frac{2aN}{\lambda}}.$$

In that case, we easily get:

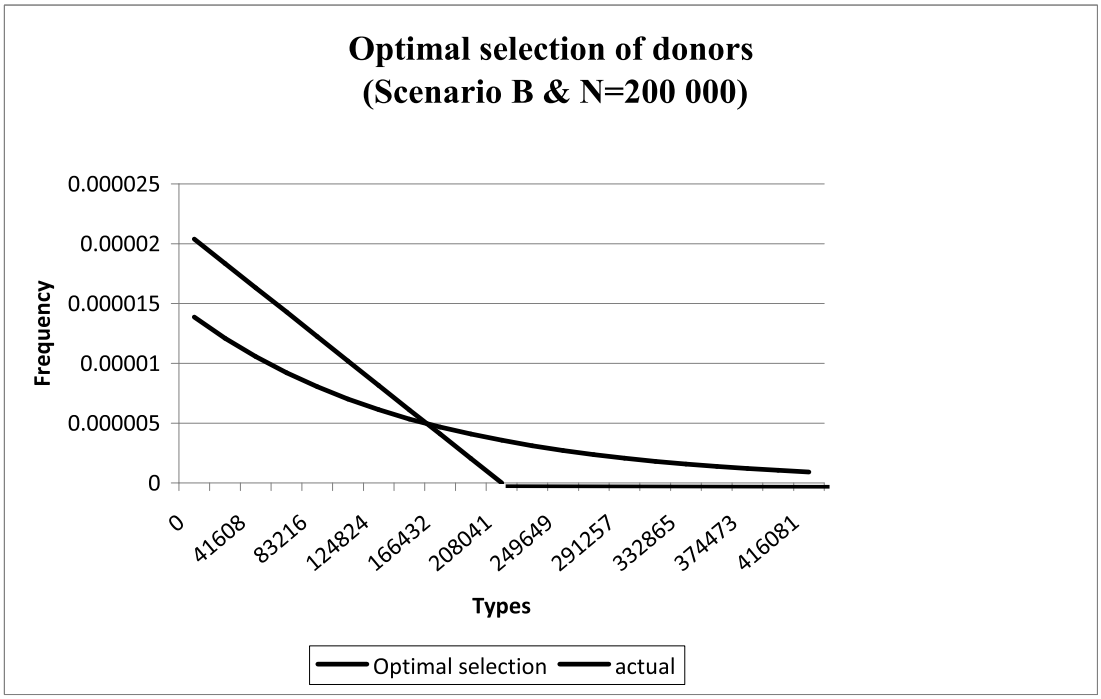
$$\pi_0 = 1 - \{J_0 \ \lambda e^{-\frac{\lambda J_0}{2} - \frac{aN}{J_0}} + e^{-\lambda \ J_0}\}$$

and the optimal sampling mechanism is:

$$\begin{cases} q_j = \frac{1}{J_0} + \frac{\lambda}{aN}\{\frac{J_0}{2} - j\} & j \leq J_0 \\ q_j = 0 & j > J_0 \end{cases}$$

We illustrate this mechanism by the following graph (see Figure 2):

Figure 2:



Finally the efficiency of the optimal mechanism is represented in figure 1 for different values of X.

The two main comments following from these results are the requirement of a large registry and the low efficiency of a screening procedure. Even for scenario B (which is "optimistic" in terms of heterogeneity of HLA genotypes) a multiplication by 10 of the present French registry is required to reach more than 80% of efficiency. Moreover the difference between dashed and continuous curves shows that the impact of any selection of donors would be very low. Contrarily to the intuition of numerous researchers in this field the optimal mechanism is not oriented to over sample the rare types but to the reverse. In our case (Figure 2), $e^{-J_0\lambda} = 0.25$ of the population is excluded of the possibility of a graft and frequent types should be over sampled. If a population has for example ethnic minorities (and then very rare types) the strict argument of efficiency leads to an elimination of this group. In other terms the objective to give the same chance to receive a graft for each member of an ethnic group requires a size of the registry (and then a cost) larger than the size required to get this chance for an individual randomly drawn in the population.

5 Equality constraints between ethnic groups

If a registry of voluntary donors is formed without donor's selection the probability of finding a donor may be very different between ethnic groups, for example in the case where the groups are a large majority of people and a small ethnic minority (see Bergstrom et al., 2008). As we have shown in the previous sections, an optimal selection of donors which maximizes the ex-ante matching probability strenghtens this phenomena by the elimination of the "rare" types which belongs to the minority's types. Then the objective of the registry regulator may be to design the registry in order to equalize the probability of finding a donor in each group. We assume that the belonging to a group is unambiguous and known both by the registry management and the individuals.
We simplify our presentation by considering two groups in the population. The extension to some groups follows obviously.
Let us assume that each group is characterized by the following characteristics:

i) A set of types and a probability distribution $(p_j^r)_{j=1,...,J^r}$ for group $r(r = 1, 2)$

ii) a value of a^r for each group. For simplicity we don't consider various values for a^r for different types but a may depends on the group.

At the level of the whole population we still have a single list of types, a distribution $(p_j)_{j=1,...,J}$ and a value a equal to the mean between the a^r weighted by the proportions of the two populations; We specify also our presentation by considering the case where no initial registry exists but here also the extension is obvious.
Let us assume that the total size N of the registry if fixed and the objective of the registry is to equalize the probability of finding a donor by fixing N^1 and $N_2 = N - N^1$, the size of the registry in each community. Two cases may be considered:

- no optimal selection in each group. In that case it follows from proposition 1 that N_1 is obtained by solving in N_1:

$$\sum_{j=1}^{J^1} p_j^1 \, e^{-a^1 N^1 p_j^1} = \sum_{j=2}^{J^2} p_j^2 \, e^{-a^2 (N-N^1) p_j^2}.$$

The cost in term of ex-ante probability may be obtained by comparing 1 minus the previous value to

$$1 - \sum_{j=1}^{J} p_j \, e^{-aNp_j}$$

- optimal selection in each group. We may apply the results of proposition $2ii$). From each group we derive an optimal probability $\pi^{01}_{(N_1)}$ and $\pi^{02}_{(N-N_1)}$ and N_1 is obtained by solving:

$$\pi^{01}(N^1) = \pi^{02}(N - N^1)$$

This value may be compared to $\pi^0(N)$, the optimal value without equality constraint. If more than two groups are considered we get $r-1$ unknowns values $(N^1, ..., N^{r-1}$ if r groups exist and we have $r-1$ equations. In the case of optimal selection, we have:

$$\pi^{01}(N_1) = \pi^{02}(N_2)..... = \pi^{0r-1}(N_r - 1)$$
$$= \pi^{0r}(N - (N_1 + ... + N_{r-1})).$$

This computation may be illustrated by considering the case of two groups with no selection where the frequencies of the types are approximated by exponential distributions in the two groups. Following the notations of section 2 where the characteristics of each group are indexed by 1 and 2, the sizes of the two registries N^1 and $N - N^1$ which equalize the probabilities of finding a donor, satisfies:

$$\frac{1}{a^1 N^1 \lambda^1}(1 - e^{-a^1 N^1 \lambda^1}) = \frac{1}{a^2 N^2 \lambda^2}(1 - e^{-a^2 N^2 \lambda^2})$$

which is equivalent to:

$$a^1 N^1 \lambda^1 = a^2 N^2 \lambda^2$$

or, using the relation between λ^1, λ^2 and X^1, X^2:

$$\frac{a^1 N^1}{X^1} = \frac{a^2 (N - N^1)}{X^2}$$

$$\frac{a^1 N^1}{X^1} = \frac{a^2 (N - N^1)}{X^2} \Rightarrow \begin{cases} N1 = \dfrac{\frac{a^2}{X^2}}{\frac{a^1}{X^1} + \frac{a^2}{X^2}} N \\ \\ N2 = \dfrac{\frac{a^1}{X^1}}{\frac{a^1}{X^1} + \frac{a^2}{X^2}} N \end{cases}$$

Let us simplify by assuming $a_1 = a_2$ and define α_1 and α_2 by $X_1 = \alpha_1 P_1$ and $X_2 = \alpha_2 P_2$ where P_1 and P_2 are the populations of the groups. The values α_1 and α_2 are in $[0, 1]$ and measure the heterogeneity in each group. Hence,

$$N^1 = \frac{\frac{\alpha_1}{\alpha_2}\frac{P_1}{P_2}}{1 + \frac{\alpha_1}{\alpha_2}\frac{P_1}{P_2}} N$$

$$N^2 = \frac{\frac{\alpha_2}{\alpha_1}\frac{P_2}{P_1}}{1 + \frac{\alpha_2}{\alpha_1}\frac{P_2}{P_1}} N$$

The size of each group depends on the two ratios $\frac{P_1}{P_2}$ and $\frac{\alpha_1}{\alpha_2}$ the relative size and the relative heterogeneity of each group. The extension of this results to many groups and gives:

$$N^0 = \frac{\alpha^l P^l}{\sum_{l=1}^{r} \alpha^l P^l} N$$

where r is the number of groups.

In order to evaluate the impact of the equality constraint under the probability of finding a donor, we have performed some simulations in the spirit of section 4. We assume that the shape of frequencies probabilities may be approximated by exponential densities both in each group and in the whole population. If X, X_1, X_2 represent the number of types of the whole population and for the two sub groups needed to describe one half of the population, we assume firstly that the families of genotypes are totally distinct ($X = X_1 + X_2$). Then it is possible to consider the case where some types are common ($X \leq X_1 + X_2$). We assume a difference between the a^1 and a^2 parameters and several values of the size of the registry N are selected.

From previous results, we obtain (see table 1):

- the value of π (no selection, no equality constraint)
- the values of $\pi^1 = \pi^2$ (no selection, under equality)
- the value of π^0 (optimal selection, no constraint)
- the values of $\pi^{01} = \pi^{02}$ (optimal selection in each group under equality constraint)

The table shows also the size of N^1, N^2 the two registries for the two groups under equality constraint.

The final empirical simulation consists in the computation of the size $\tilde{N}$ of the

whole registry which gives the same probability for each group (as the probability without constraint with a registry of size N). Intuitively, $\tilde{N} - N$ is the price to pay (in terms of the increment of the registry) of the equality constraint keeping the probability of finding a compatible donor.

Table 1: Impact of equality constraint between ethnic groups under efficiency

N	200 000
$pop1$	50 000 000
$pop2$	13 000 000
a^1	0.40
a^2	0.30

$X1$	200 000		$X1$	200 000		$X1$	200 000		$X1$	200 000	
$X2$	500 000		$X2$	500 000		$X2$	100 000		$X2$	100 000	
N^1	46 154		N^1	46 154		N^1	120 000		N^1	120 000	
N^2	153 846		N^2	153 846		N^2	80 000		N^2	80 000	
X	700 000		X	500 000		X	300 000		X	200 000	
$\tilde{N}$	234 845		$\tilde{N}$	328 783		$\tilde{N}$	210 758		$\tilde{N}$	316 138	
π	$\pi^1 = \pi^2$	$\Delta\ \pi$	π	$\pi^1 = \pi^2$	$\Delta\ \pi$	$\pi^1 = \pi^2$	$\Delta\ \pi$	π	π	$\pi^1 = \pi^2$	$\Delta\ \pi$
0.037	0.031	0.005	0.051	0.031	0.019	0.083	0.079	0.004	0.121	0.079	0.042
π^0	$\pi 01 = \pi^{02}$	$\Delta\ \pi^0$	π^0	$\pi^{01} = \pi^{02}$	$\Delta\ \pi^0$	π^0	$\pi^{01} = \pi^{02}$	$\Delta\ \pi^0$	π^0	$\pi^{01} = \pi^{02}$	$\Delta\ \pi^0$
0.058	0.051	0.008	0.078	0.051	0.027	0.119	0.114	0.005	0.165	0.114	0.050

N	500 000
$pop1$	58 000 000
$pop2$	3 000 000
a^1	0.40
a^2	0.30

$X1$	200 000		$X1$	200 000		$X1$	200 000		$X1$	200 000	
$X2$	300 000		$X2$	300 000		$X2$	50 000		$X2$	50 000	
N^1	166 667		N^1	166 667		N^1	375 000		N^1	375 000	
N^2	333 333		N^2	333 333		N^2	125 000		N^2	125 000	
X	500 000		X	250 000		X	250 000		X	220 000	
$\tilde{N}$	592 623		$\tilde{N}$	1 185 246		$\tilde{N}$	526 776		$\tilde{N}$	598 609	
π	$\pi^1 = \pi^2$	$\Delta\ \pi$	π	$\pi^1 = \pi^2$	$\Delta\ \pi$	$\pi^1 = \pi^2$	$\Delta\ \pi$	π	π	$\pi^1 = \pi^2$	$\Delta\ \pi$
0.125	0.107	0.018	0.023	0.107	0.123	0.230	0.220	0.010	0.256	0.220	0.035
$\pi 0$	$\pi^{01} = \pi^{02}$	$\Delta\ \pi 0$	π^0	$\pi^{01} = \pi^{02}$	$\Delta\ \pi^0$	π^0	$\pi^{01} = \pi^{02}$	$\Delta\ \pi^0$	π^0	$\pi^{01} = \pi^{02}$	$\Delta\ \pi^0$
0.170	0.149	0.021	0.281	0.149	0.133	0.281	0.271	0.010	0.307	0.271	0.035

N	1 000 000
pop1	62 000 000
pop2	1 000 000
a^1	0.4
a^2	0.3
$X1$	250 000
$X2$	20 000
N^1	903 614
N^2	96 386
X	270 000
$\tilde{N}$	1 020 625

π	$\pi^1 = \pi^2$	$\Delta\ \pi$
0.374	0.368	0.005
π^0	$\pi^{01} = \pi^{02}$	$\Delta\ \pi^0$
0.419	0.414	0.005

These simulations are just examples and we see that in most of cases the difference between cases without or with constraint ($\Delta\pi$ or $\Delta\pi^0$) is not very important or equivalently that the value of $\tilde{N} - N$ is relatively small. An important difference is computed in the case of a small minority (3 000 000 compared to 58 000 000) very heterogenous ($X^1 = 200000, X^2 = 300000$ and $\frac{\alpha_1}{\alpha_2} = 0.034$). This situation may be realistic in some circumstances. In that case the cost of equality is a division by 2 of the efficiency of the registry and the registry should be multiply by more than 2 in order to keep the efficiency constraint.

6 Extension to the possibility of mismatch: Application to cord blood bank efficiency

In the previous analysis only perfect matching between donor and receiver is considered. We now analyze a model where mismatch is accepted under some conditions. This generalized hypothesis is suitable to consider the application of our model to cord blood banks. Cord blood are preserved in several banks in different countries (see Katz–Benichou, 2007). The size of these banks are smaller than the size of the registries (5 000 in France for example) but non fully compatible transplants are realized. For example it is commonly accepted that four identical alleles under the six considered are sufficient to decide the transplant.

The model is identical to the previous one but we introduce a supplementary mechanism. We call $Z_j \subset \{1, ..., J\}$ the set of types compatible to a specific receiver of type j. In case of perfect matching Z_j reduces to $\{j\}$ but in case of imperfect matching Z_j may have many components (including j).

We consider the extension of our previous analysis under two simplifications: we assume that no initial registry or bank exists (we evaluate the efficiency of the realization of a bank and not of the increment of an existing bank) and we consider the case where $a_j = a \ \forall j$. In the case of cord blood the level of typing is usually high and there is no individual decision to accept to be a donor or not. The main element determining the non use of the cord blood is its possible bad quality which may be assumed to have a probability independent of the type. It is natural to consider case where a is high (0.9 for example) but with a great number of types because the typing is extremely precise. Note that the possibility of mismatch concerns only the recorded type and not the hidden phenomena modelized by the $a'_j s$.

We first extend proposition 1.

Proposition 3: The efficiency of a registry or a bank in case of imperfect matching is approximated by:

$$\pi = 1 - \sum_{j=1}^{J} p_j \ e^{-aN \sum_{j=1}^{J} \mathbb{1}(l \in Z_j) q_j}$$

where $\mathbb{1}(l \in Z_j) = 1$ if $l \in Z_j$ and 0 else.

Proof: The proof is immediate. Fix j and consider that all the types in Z_j are identical for the purpose of the graft. This group of types has a probability of $\sum_{l=1}^{J} \mathbb{1}(l \in Z_j) q_j = Q_j$. The proof of proposition 1 may be reproduced replacing q_j by Q_j and we get the result.

□

Let us remark that the family $(Q_j)_{j=1,...,J}$ does not constitute a probability because $\sum_{j=1}^{J} Q_j = K \neq 1$. We call K the weighted average number of compatible types and we may normalize the Q_j into $\tilde{q}_j = \frac{1}{K} Q_j$ which satisfy the condition of a probability measure on $\{1, ..., J\}$. Then the efficiency may be rewritten:

$$\pi = 1 - \sum_{j=1}^{J} p_j \, e^{-aNK\tilde{q}_j}.$$

The effect of mismatch is then to transform the initial sampling mechanism from q_j to $\tilde{q}_j$ and to multiply the size N of the registry by K.
We have used in the section 4 an exponential approximation to calibrate the efficiency of the registry. An identical computation may be realized here.
Assume that the number of types is infinite and exponentially distributed with a parameter λ and that $\tilde{q}_j$ is also exponentially distributed with a parameter μ then the efficiency of the registry is equal to:

$$\pi = 1 - \frac{\lambda}{\mu (aNK\mu)^{\frac{\lambda}{\mu}}} \, \gamma(\frac{\lambda}{\mu}, aNK\mu)$$

where $\gamma(a, x)$ is the lower incomplete gamma function ($\gamma(a, x) = \int_0^x t^{a-1} e^{-t} dt$).
This result follows immediately from the computation:

$$\int p(j) \, e^{-aN\tilde{q}(j)} dj = \int \lambda \, e^{-\lambda j} e^{-aNK\mu \, e^{-\mu j}} dj$$

$$= \frac{\lambda}{\mu (aNK\mu)^{\frac{\lambda}{\mu}}} \int_0^{anK\mu} y^{\frac{\lambda}{\mu} - 1} \, e^{-y} \, dy$$

by change of variable ($y = aNK\mu \, e^{-\mu j}$).
This result may be used to perform some calibrations reproducing the French situation. We know that the French cord blood bank has around 5 000 cord bloods and that 30% of the potential receivers find a compatible cord blood in

the French bank. The exercise is then to find a reasonable set of parameters compatible to this result.

Our simulations are realized under the assumption of absence of screening ($p_j = q_j$) and under continuous approximation. We will consider two cases: first $\lambda = \mu$ (the frequencies of the aggregated types are similar to the original one) and second that $\lambda = 1.5\ \mu$. This second case is motivated by an empirical argument: using our French data set and using the definition of mismatch founded of at least four alleles in common over six, we estimate this ratio to 1.42[4]. In the cord blood case types are more precise than in the marrow donor registries (four digits typing plus more information on the compatibility criteria). As in section 4 we calibrate λ by choosing a value X such that the 50% more frequent individuals has a type smaller than X (if the types are ranked by decreasing frequencies). We select $X = 500\ 000$, $1\ 000\ 000$ and $1\ 500\ 000$. Remember that $\lambda = -\frac{1}{X} \ln \frac{1}{2}$ and the X is higher as in section 4 because we have in mind a high precision typing.

- The value of a is high in the cord blood bank and we choose $a = 0.9$.

- The value of K is a key parameter. To evaluate this number we use the French registry of voluntary donors which gives 66 164 types and we compute $K = 297$ with the previous definition of mismatch. This number may be interpreted: it means that at a low typing precision, a cord blood is equivalent to 297 donors. This number should be manipulated with caution: this result is derived from a sample and computed to a low precision typing. The actual value is probably lower than this value and we try K = 100, 200 and 300.

The results are summarized in Table 1:

Under the previous assumption, if we consider for example the case where K=200, X=500 000, $\frac{\lambda}{\mu} = 1.5$ and a=0.9 the evolution of the efficiency as a function of the size of the cord blood bank is summarized in the following table:

[4]The estimations of $\frac{\lambda}{\mu}$ is based on the property that for exponential distribution, $\frac{\lambda}{\mu}$ is proportional to the ratio of $e^{E\ ln(p)}/e^{E\ ln(q))}$ and this ratio is estimated by the ratio of the geometrical means of p_j and q_j.

Table 2: Efficiency of a cord blood bank of size 5 000

	$\lambda = \mu$		
$X \setminus K$	100	200	300
500 000	0.26	0.43	0.55
1 000 000	0.14	0.26	0.35
1 500 000	0.10	0.18	0.26

	$\frac{\lambda}{\mu} = 1.5$		
$X \setminus K$	100	200	300
500 000	0.22	0.38	0.50
1 000 000	0.12	0.22	0.30
1 500 000	0.08	0.15	0.22

Table 3: Efficiency of a cord blood bank as function of its size

X	π
5 000	0.38
10 000	0.59
15 000	0.72
20 000	0.80
25 000	0.85
30 000	0.88

We see that the observed performance may be reproduced by these values but more statistical analysis are required to confirm these results.
The computation of the optimal registry under mismatch assumption is difficult and up to our knowledge no solution on closer form may be derived; However if we optimize the efficiency without imposing the constraint $q_j \geq 0$ we may derive an upper bound of the efficiency.
Let us define the following matrices:

$$B = (b_{jl})_{j,l} \qquad b_{jl} = \mathbb{1}(l \in Z_j)$$

where B is assumed to be invertible

$$P = \begin{pmatrix} p_1 & & 0 \\ & \cdot & \\ & & \cdot \\ 0 & & p_j \end{pmatrix} \quad p = \begin{pmatrix} p_1 \\ \cdot \\ \cdot \\ p_J \end{pmatrix}$$

$A = BP$

$$\text{where } \mathbb{1} \in \mathbb{R}^{\mathbb{J}} = \begin{pmatrix} 1 \\ . \\ . \\ . \\ 1 \end{pmatrix} \text{ if } x = \begin{pmatrix} x_1 \\ . \\ . \\ x_J \end{pmatrix} \ln x = \begin{pmatrix} \ln x_1 \\ . \\ . \\ \ln x_J \end{pmatrix} e^x = \begin{pmatrix} e^{x^1} \\ . \\ . \\ e^{x^J} \end{pmatrix}$$

Proposition 4: The optimal efficiency π^0 in presence of mismatch satisfies the inequality:

$$\pi^0 \leq 1 - (\mathbb{1}'\ B^{-1}\ \mathbb{1})\ e^{\frac{\mathbb{1}'\ B^{-1}\ \ln(B^{-1}\mathbb{1})}{\mathbb{1}'\ B^{-1}\ \mathbb{1}}}\ e^{\frac{\mathbb{1}'\ B^{-1}\ \ln(p)}{\mathbb{1}'\ B^{-1}\ \mathbb{1}}}\ e^{\frac{-aN}{\mathbb{1}'B^{-1}\mathbb{1}}}$$

Proof:

We maximise $\pi = 1 - \sum_j p_j\ e^{-aN \sum_l q_l \mathbb{1}(l \in Z_j)}$ under the constraint $\sum_j q_j = 1$ and not under the positivity constraints. As some constraints are relaxed the maximum reached is greater than the maximum under constraints.

Technically using the previous notations the first order conditions are summarized by:

$$A\ e^{-aNBq} = \lambda \mathbb{1}$$

Or equivalently there exists a constant c such that:

$$q = -\frac{1}{aN}\ B^{-1}\ \ln(A^{-1}\mathbb{1}) + c\ B^{-1}\mathbb{1}$$

using $\mathbb{1}'q = 1$ we get

$$c = \frac{1}{\mathbb{1}'B^{-1}\mathbb{1}}(1 + \frac{1}{aN}\ \mathbb{1}'B^{-1}\ln(A^{-1}\mathbb{1}))$$

and

$$q = -\frac{1}{aN}\ B^{-1}\ \ln(A^{-1}\mathbb{1}) + \frac{1}{\mathbb{1}'B^{-1}\mathbb{1}}B^{-1}\mathbb{1} + \frac{1}{aN\mathbb{1}'B^{-1}\mathbb{1}}\mathbb{1}'B^{-1}\ln(A^{-1}\mathbb{1})B^{-1}\mathbb{1}$$

and plugging this value, we get:

$$\pi^0 \leq 1 - (p'A^{-1}\mathbb{1})\ e^{\frac{-\mathbb{1}B^{-1}\ln(A^{-1}\mathbb{1})}{-\mathbb{1}B^{-1}\mathbb{1}}}\ e^{\frac{-aN}{-\mathbb{1}B^{-1}\mathbb{1}}}$$

objective function. Moreover, elementary computation shows that $p'A^{-1}\mathbb{1} = \mathbb{1}'B^{-1}P^{-1}p = \mathbb{1}'B^{-1}\mathbb{1}$ and $\mathbb{1}'B^{-1}\ln(A^{-1}\mathbb{1}) = \mathbb{1}'B^{-1}\ln(p) + \mathbb{1}'B^{-1}\ln(B^{-1}\mathbb{1})$ and the result follows.

□

This result is complex but it may be simplified in a particular case. Let us assume that the cardinality of each Z_j is identical for any j and equal to L. In that case $B^{-1}\mathbb{1} = \frac{1}{L}\mathbb{1}$ and $\mathbb{1}'B^{-1}\mathbb{1} = \frac{J}{L}$, the total number of types divided by the number of types compatible to any type. In that case $\mathbb{1}'B^{-1}\ ln(B^{-1}\mathbb{1}) = -\mathbb{1}'B^{-1}\mathbb{1}\ln L$ and the inequality becomes:

$$\pi^0 \leq 1 - J\ e^{\frac{L}{J}\mathbb{1}'B^{-1}\ln p}\ e^{\frac{-aNL}{J}}$$

In the case of cord blood banks the matrix B is usually symmetric: for example, the mismatch is possible if donors and receivers have four alleles in common among the six. In that case using the previous argument $\mathbb{1}'B^{-1} = \frac{1}{L}\mathbb{1}'$ and the formulae may be simplified by :

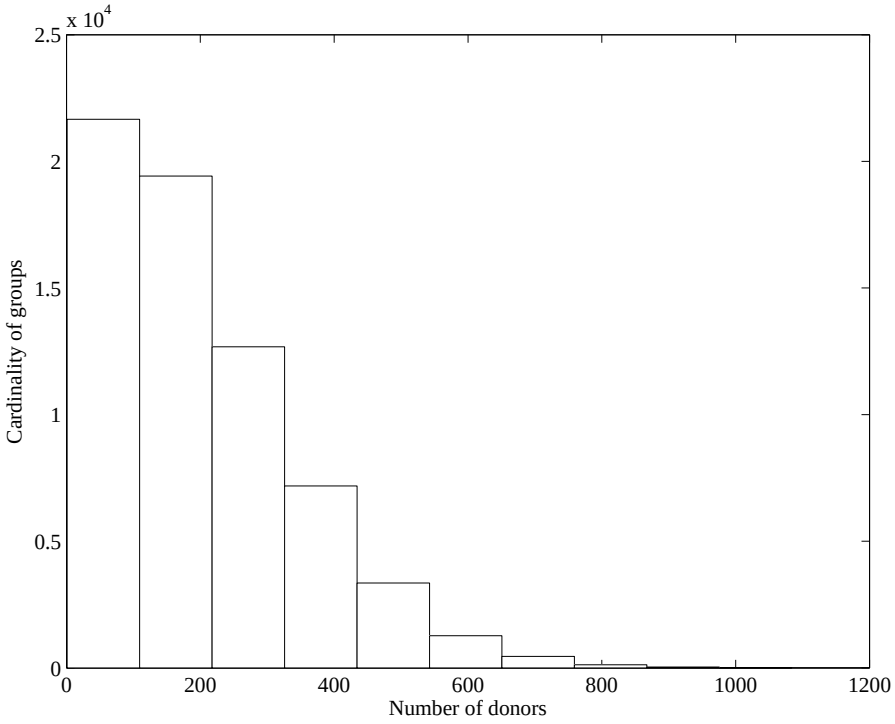
$$\pi^0 \leq 1 - J\ \bar{\bar{p}}\ e^{\frac{-aNL}{J}}$$

where $\bar{\bar{p}}$ has been defined as the geometrical mean of the frequencies. In that case the upper bound is exactly identical to one obtained in section 4 except that J is replaced by $\frac{J}{L}$.
We may see if our assumption (cardZ_j constant) is validated empirically. If we take the French sample of marrow donor registry and the previous definition of possible mismatch, the cardinality of Z_j is not constant as shown in Figure 3.
Let us remind that our sample is small in comparison to the number of types. A statistical theory of this estimation is needed to reject correctly our assumption. This assumption may be used to give an order of magnitude of the efficiency of a cord blood bank without very complex computations.

7 Conclusion

In this paper we have constructed a model which formalizes the matching between donors and receivers of Haematopoietic Stem cells. This model has been extended to some possibility of mismatch between the type of the donor and the type of the receiver and then it may be used to compare the efficiency of the two systems. A registry of size N determines a probability $\pi(N)$ of finding a fully compatible donor to any receiver. If the number of receivers during a given period is equal to M and if the value of finding a compatible donor is V, the value of the registry is equal to $\pi(N)MV$ and its cost is evaluated to $(1+o)C(N)$ where $C(N)$ is the cost function of a registry of size N and o is the opportunity cost of public fund in a case where the registry is financed by the

Figure 3: Histogram of the cardinality of the group of possible donors (Z_j)



state. [5]Then the first order condition which determines the optimal size of the registry is equal to:

$$\frac{\partial}{\partial N}\pi(N) = \frac{(1+o)}{VM}\frac{\partial}{\partial N}C(N).$$

This equality is obtained by equalizing the derivative of the value of the system $\Pi(N)VM$ and of the cost $(1+o)C(N)$ (see Fève et al. 2007). We illustrate this computation for two cases roughly representing the value of a voluntary donor registry and a cord blood bank for a country like France. The cost is assumed to be linear with no fixed cost and a marginal cost equal to c.

We consider the case where the model is approximated by a model of continuous types exponentially distributed. In case of mismatches, we assume that there is no selection of donors and that the normalized frequencies of the groups of possible donors are also exponential. The scenarii are then characterized by λ (or the median X), the ratio $\frac{\lambda}{\mu}$, a, M, K, c and o. The previous equation is used to compute numerically the optimal size of the registry and we derive the value $\pi(N)$ and $\pi(N) - (1+o)\,C(N)$.

[5]The costs considered in the model are only the costs of the registration and the costs of maintenance of the registry and not the search costs which are negligible

Table 4: Economic evaluation of registries and cord blood banks

Parameters	Voluntary donors registry		Cord blood bank	
	I	II	III	IV
X	100 000	200 000	500 000	1 000 000
λ/μ^*	-	-	1.5	1.5
K	1	1	200	200
a	1/3	1/3	0.9	0.9
V	60 000 €	60 000 €	60 000 €	60 000 €
M (for 10 years)	10 000	10 000	10 000	10 000
δ	180 €	180 €	1 850 €	1 850 €
o	0.5	0.5	0.5	0.5
Optimal dimension	655 966	329 494	32 456	46 737
Probability to find a donor	0.48	0.17	0.89	0.83
Social value **	291 $\bar{M}$ €	100 $\bar{M}$ €	537 $\bar{M}$ €	501 $\bar{M}$ €

$^*\lambda = \frac{-1}{X}\, ln\frac{1}{2}$ $\quad$ $^{**}\pi(N)VM - cN$

The registry of voluntary donors is assumed to be recorded at a low precision but the cord bloods are typed at a high precision and we have considered a low hypothesis about the number of types (I and III) and a high hypothesis (II and IV). This result should be taken with caution because the number and the distribution of HLA genotypes at a high precision coding is largely unknown; These results are calibrated for a country like France. The costs are the values usually accepted and the value of finding a donor is derived from Fève (2006) and the value V is obtained by Fève et al. (2007),(see also Viscusi,Aldy, 2003). We assume that the value of a marrow transplant and of a cord blood transplant are the same. We also assume that this value does not depend on the quality of matching between the cord blood and the receiver. The cost of cord blood bank comes from a private communication given by Gregory Katz–Benichou. We have discussed in the present paper the value of K.

Under these hypotheses the results are non ambiguous in favor of the cord blood bank which seems to give an excellent efficiency at a low cost. This result shows that the size of the actual French bank should be multiplied by at least 6 to get

the optimal efficiency. We want to insist on the fact that these results require more empirical studies to fix the parameters:the comparison between the two systems seems to be very robust.

At least two aspects of registries and cord blood banks are not considered in this paper: the international interconnection and the minorities. These two aspects are related because the improvement of the probability of a graft in a minority may be obtained in a better way with international collaboration than by a distortion of the national registry. We are working as extension on our model where we quantify this type of question.

References

Anderson J., Anasetti C., Appelbaum F., Schoch G., Gooley T., Hansen J., Buckner C., Sanders J., Sullivan K., Storb R. (1996), Unrelated donor marrow transplantation for myelodysplasia (mds) and mds–related acute myeloid leukemia. British Journal of Hematology, 93(1):59-67.

Bergstrom T.C., Garrat R., Sheehan-Connor D., september 2007, One Chance in a Million: Altruism and the Bone Marrow Registry, working paper, department of Economics, UCSB

Fève F., " Economie et Statistique du Don de Cellules Souches Hématopoiétiques : Contributions à la Gestion de Registres de Donneurs Volontaires", Thèse de Doctorat de Sciences Economiques, Université Toulouse III, 2006. http://www.idei.fr/doc/by/ffeve/these.pdf

Fève F., Cambon–Thomsen A., Eliaou J.-F., Gourraud P.A., Raffoux C., Florens J.-P. (2007), Economic evaluation of the organization of a registry of haematopoietic stem cell donors. Revue d'Epidémiologie et de Santé Publique 55; 275-284.

Fève F.,Florens J.-P. (2007), How many different HLA genotypes exist in a population ?, *discussion paper Toulouse School of Economics*.

Hoffman-Smith C. (1993), Matching marrow donors & recipients: downsized system helps save more lives. Health Inform. Sep;10(9):18, 20.

Hurley CK, Schreuder GM, Marsh SG, Lau M, Middleton D, Noreen H. (1997), The search for HLA-matched donors: a summary of HLA-A,-B,-DRB1/3/4/5 alleles and their association with serologically defined HLA-A,-B,-DR antigens. Tissue Antigens. Oct;50(4):401-18.

Katz–Benichou G., Umbilical cord blood banking: economic and therapeutic challenges, Int.Healthcare Technology and Management, Vol.8, Nř5, 2007.

Kollman C., Abella A., Baitty R.L., Beatty P.G., Chakraborty R., Christiansen C.L., hartzman R.J., Hurly C.K., Milford E., Nyman J.A., Smith T.J., Switzer

G.E., Wada R.K., Setterholm M (2004), Assessment of optimal size and composition of the U.S. national registry of haematopoietic stem cell donors. transplantation, 78(1):89-95.

Lonjou C, Clayton J, Cambon-Thomsen A, Raffoux C. (1995),HLA-A,-B,-DR haplotype frequencies in France implications for recruitment of potential bone marrow donors. Transplantation. Aug 27;60(4):375-83.

Mori M., Beatty P.G., Graves M., Boucher KM, Milford F.L. (1997). HLA gene and haplotype frequencies in the North American population: the National Marrow Donor Program registry. Transplantation, 64:1017-1027.

Muller CR. (2002) Computer applications in the search for unrelated stem cell donors.TransplImmunol.Aug;10(2-3):227-40.

Ottinger H, Grosse-Wilde M, Schmitz A, Grosse-Wilde H. (1994),Immunogenetic marrow donor search for 1012 patients: a retrospective analysis of strategies, outcome and costs. Bone Marrow Transplant. 1994-14 Suppl 4:S34-8.

Oudshoorn M, Cornelissen JJ, Fibbe WE, de Graeff-Meeder ER, Lie JL, Schreuder GM, Sintnicolaas K, Willemze R, Vossen JM, van Rood JJ. (1997), Problems and possible solutions in finding an unrelated bone marrow donor. Results of consecutive searches for 240 Dutch patients. Bone Marrow Transplant. Dec;20(12):1011-7.

Rendine S, Barbanti M, Borelli I, Dall'Omo AM, Roggero S, Sacchi N, Curtoni ES. (1999), The Italian Registry of Bone Marrow Donors: genetic structure and recruitment strategy - Ann Ist Super Sanita.35(1):21-34. Italian.

Schipper RF, Oudshoorn M, D'Amaro J, van der Zanden HG, de Lange P, Bakker JT, Bakker J, van Rood JJ. (1996), Validation of large data sets, an essential prerequisite for data analysis: an analytical survey of the Bone Marrow Donors Worldwide. Tissue Antigens. Mar;47(3):169-78.

Schuler U, Rutt C, Baier D, Keller JV, Stahr A, Grathwohl A, Ehninger G.(2000), Approaches to managing volunteer marrow donor registry HLA data. Algo-

rithms for directing donor center-initiated HLA-DR typing of selected donors. Rev Immunogenet. 2000-2(4):541-6.

Speiser DE, Tiercy JM, Rufer N, Chapuis B, Morell A, Kern M, Gmur J, Gratwohl A, Roosnek E, Jeannet M.(1994),Relation between the resolution of HLA-typing and the chance of finding an unrelated bone marrow donor. Bone Marrow Transplant. 1994 Jun;13(6):805-9.

Sonnenberg FA, Eckman MH, Pauker SG. (1989),Bone marrow donor registries: the relation between registry size and probability of finding complete and partial matches. Blood. 1989 Nov 15;74(7):2569-78.

Takahashi K, Juji T, Miyazaki H. (1989), Determination of an appropriate size of unrelated donor pool to be registered for HLA-matched bone marrow transplantation. Transfusion. 1989 May;29(4):311-6.

Tiercy JM, Bujan-Lose M, Chapuis B, Gratwohl A, Gmur J, Seger R, Kern M, Morell A, Roosnek E. (2000), Bone marrow transplantation with unrelated donors: what is the probability of identifying an HLA-A/B/Cw/DRB1/B3/B5/DQB1-matched donor? Bone Marrow Transplant. 2000 Aug;26(4):437-41.

Viscusi K., Aldy J. (2003). The value of a statistical life: A critical review of market estimates throughout the world. Journal of risk and Uncertainty, 27(1):5-76.