Efficiency of bone marrow donors' registries: some orders of magnitude

F. Feve*, J.P. Florens†

First Draft: February 2005
Preliminary version

*University of Toulouse (IDEI) and INSERM U558
†University of Toulouse (GREMAQ and IDEI) and IUF

1 Main results of this note

This note provides a methodology for the evaluation of a donor's registry design based on the probability for a receiver to find a donor. We compare essentially registries based on random arrival of donors and registries where the selection of the donors is made by an optimal mechanism. Practicable implementations of such a mechanism by filtering processes are not discussed in the paper.

The theoretical results exhibit the main elements of the maximal efficiency of a registry: essentially the efficiency is determinated by the size of the registry, the probability for a donor to be available for a graft, the number of types in the population and the dispersion of the frequencies of types captured by the geometric mean of these frequencies.

The calibration of the model for different scenarii shows essentially the following result: a donor registry system of size corresponding to the actual ones in the main countries is not very efficient (approximatively 10% of receivers find a donor) and this efficiency is difficult to increase. A huge increment of the registry (multiplication by 2 or 3) and an extremely efficient selection of the donors would lead in the best case to increase the efficiency to 20%.

2 Some theoretical results

We consider an abstract model where the number of types is equal to J and the frequency in the population (from which the receivers are drawn) of type $j = 1, ...J$ is p_j. There exists an initial registry of N_0 donors and the number of donors of type j in initial registry is N_{0j} (possibly equal to 0 for some types). The registry increases of N new donors by a sampling process characterized by the frequency q_j. This mechanism is evaluated by the expected probability not to find a donor in the registry for any receiver. This probability is equal to (see Feve and Florens (2005)):

$$\Pi = \sum_{j=1}^{J} p_j (1-a)^{Noj} e^{-aNq_j} \tag{1}$$

where a is the probability for a donor to be available for the transplantation. Then the proportion of receivers who may be transplanted is $1 - \Pi$. A registry design (defined by N and q_j) may then be evaluated by Π. Several registry designs may be considered.

i) Non selective mechanism: the donors are drawn randomly in the same population as the receivers. Then $q_j = p_j$

ii) Optimal selection mechanism. The q_j are selected in order to minimize Π. Under the constraint:

$$\sum_{j=1}^{J} q_j = 1, \tag{2}$$

the optimal value is:

$$q_j^o = \frac{1}{J} + \frac{l}{aN}\left\{ lnp_j - \frac{1}{J}\sum_{\ell=1}^{J} lnp_\ell \right\} + \frac{l}{aN}\left\{ N_{0j} - \frac{N_0}{J} \right\} ln(1-a) \tag{3}$$

This value converges to $\frac{1}{J}$ (equiprobability of the types) where N goes to infinity. This implies, in particular, that the constraint $q_j \geq 0$ is satisfied. However, if J is "large" and N "small", the optimal $q_j^{'}s$ may be negative and not represent a selection mechanism. In that case, the optimal value of Π where q_j is positively constrained is greater than the one obtained with q_j^o. In any case:

$$\Pi^o = \sum_{j=1}^{J} p_j (l-a)^{Noj} e^{-aNq_j^o} \tag{4}$$

is a lower bound of the probability to not find a donor.

The value of registry for the optimal selection of the q_j is surprisingly simple. If we substitute q_j^o given by (2) in (1) we get:

$$\Pi^o = J\bar{p}\,(1-a)^{\frac{N_0}{J}} e^{-\frac{aN}{J}} \tag{5}$$

where $\bar{p}$ is the geometrical means of the p_j.

This result suggests several comments:

1. The result does not depend on the structure of the initial registry (the $N_{oj}'s$ but only on N_o). It does not depend either on the different value of the p_j but only on the geometrical means. This implies that a registry system may be evaluated under the knowledge of a few numbers of elements only ($J, \bar{p}, a, N_o \; and \; N$). Let us underline that this property (independence of the result from the N_{oj}) is only satisfied if the q_j are not constrained to be negative (intuitively speaking, negative values "eliminate" our represented types donors)

2. If a is small, $(1-a)^{\frac{No}{J}}$ is approximately equal to $e^{-\frac{aN_0}{J}}$ (e.g. if $\frac{No}{J} = 2, a = \frac{1}{3}$, the first is equal to .44 and the second to .51 but if $a = \frac{1}{10}$ we get .8 and .818 respectively). In that case:

$$\Pi^o \simeq J\bar{p}e^{-\frac{a(N_o+N)}{J}} \tag{6}$$

   This result has an important implication. If a is small, it is equivalent to start an initial registry of size No and to increase it by N new donors in an optimal way than to draw directly $N + No$ donors by an optimal way.

iii) Semi optimal mechanisms. The optimal q_j^o are not implementable, but several screening processes may be developed (based on SNP, μ sat and observable characteristics). In any case they will lead to an evaluation of the registry between Π^o (optimal registry) and the actual Π^A values (where $q_j = p_j$). In this note we essentially want to compute and compare Π^o and Π^A in different case. This comparison provides an order of magnitude of the potential improvements of the registries.

3 Analysis of the optimal registry based on the MADO sample:

The MADO sample is constituted by 4961 individuals belonging to 4621 types ($J = 4621$). This sample will be used to list the types only and the frequency of each type (given the fact that the type has been observed in the sample) is derived from the France Greffe de Moelle (FGM) file of approximately 100 000 individuals. We are then in the case where the number and the frequencies of the types are known. These frequencies vary between .013 and $9\ 10^{-5}$. As a comparison, the value $\frac{1}{J}$ is equal to .000216. It should be noted

that 2427 types have a frequency equal to the minimum: the shape of the p_j curve is then flat for more than one half of the types. This list of types will be used as a model of the registry problem. All the dimensions of the initial problem should be reduced accordingly to the size of the model. We first create an initial registry of 1000 donors. For simplicity we assume that this registry is composed of the most frequent types proportionally to their frequencies (more precisely the first 823 types are represented by the integer part of 1000 $p_j$plus 1 individuals and the other types have no element in the file). The probability a is chosen equal to 1/3 and different values of N are selected in the computation. $N = 100$ represents an increment of the file of 10% and $N = 1000$ a doubling of the registry.

Remark: if N is selected equal to 1000 the optimal design corresponds to the initial registry 1000 donors. This is not equivalent to increase the file of 100 donors each year and to optimize the q_j at each step. However, in any case, such a sequencialy optimized procedure should give a file better than an optimal file of 2000.

A first set of results is summarized in table 1.

MADO sample

J= 4 621

Table 1 : Expected probability to not find a donor

N0= 1 000

a = 1/3 **823 types**

				elimination based on the initial registry			
N	No selection	Optimal	Optimal renormalized	N0j>=3	N0j>=4	Noj>=5	Noj>=5 et Noj=0
				0.1263	**0.0826**	**0.041**	**0.553**
0	**0.75**	**-**	**-**	**-**	**-**	**-**	**-**
100	**0.74**	**0.641**	**0.73**	**0.737**	**0.736**	**0.735**	**0.731**
500	**0.7**	**0.623**	**0.696**	**0.706**	**0.703**	**0.697**	**0.687**
1 000	**0.666**	**0.601**	**0.655**	**0.67**	**0.666**	**0.657**	**0.648**
1 500	**0.634**	**0.58**	**0.62**	**0.638**	**0.633**	**0.622**	**0.619**
2 000	**0.606**	**0.559**	**0.588**	**0.609**	**0.604**	**0.592**	**0.597**

In this table normalized q_j^o are obtained by transforming negative values of q_j^o into o. The last three columns represent registry evaluation following

fom q_j constructed in the following way : $q_j = 0$ if $N_{oj} \geqslant 3$ (or 4 or 5) and proportional to p_j else. In the last column, $q_j = 0$ if $N_{oj} \geqslant 5$ or if $N_{oj} = 0$.

Remark: we will see later on that the actual FGM file determines a probability of transplantation of less than 10%. In our model, this probability is equal to 25% in the current situation.
At least three phenomena should be underlined:

i) The increment of the file has a low impact on the probability to find a donor. If the actual file is tripled, this probability increases of 14% with the current mechanism of 19% by an optimal one.

ii) The choice of the selection mechanism has a low impact on the evaluation of the registry (less than 10% for any size N). Remember that a screening mechanism will lead to an evaluation between the two given numbers and will be function of the efficiency of screening rule. Some elements will be given later on, on some screening rules.

iii) The difference between actual and optimal rules reduces where the size of the file new registry increases effect of the size.

iv) A strategy both based on elimination of very frequents ($N_{oj} \geqslant 5$ represents two types only) and elimination of a large number of rares (elimination of 3798 types) preforms on an almost similar way as the normalized strategy.

4 Effect of imposing positivity constraints on the optimal sampling probabilities of the donors

In the previous sections we have computed an efficiency bound of the registry designs by computing its optimum under the q_j constrained to $\sum q_j = 1$.We have seen in the empirical example that the optimal q_j obtained from formulae (3) may be negative. The correct value of Π under positivity constraints on the sampling probabilities is then necessarily greater than the computed value. Unfortunately no expression of Π in a closer form of the optimization is done under all the constraints. The computation should be numerical.

We have done this computation using a sub sample in order to keep the size of the computation compatible to standard software on common PC. This sub sample is constructed by retaining only one type each four types

indexed by decreasing frequencies. We get a sub sample of 1156 types. The initial registry is reduced accordingly to the reduction of the number of types and the new $No = 255$. The parameter a is kept equal to $1/3$. We have tried several values of N defined in relation with No ($10\%\ of No$, 50%, $N = No, N = 2No$).

The results are given in table 2 "No selection" means $q_j = p_j$, the "optimal" means q_j optimal without positivity constraint, "optimal realistic" takes into account the constraints and "optimal renormalized" is obtained by optimal q_j without positivity constraint transformed by replacing negative value by 0 and by renormalization.

J= 1 156

Table 2 : Expected probability to not find a donor

N0= 255

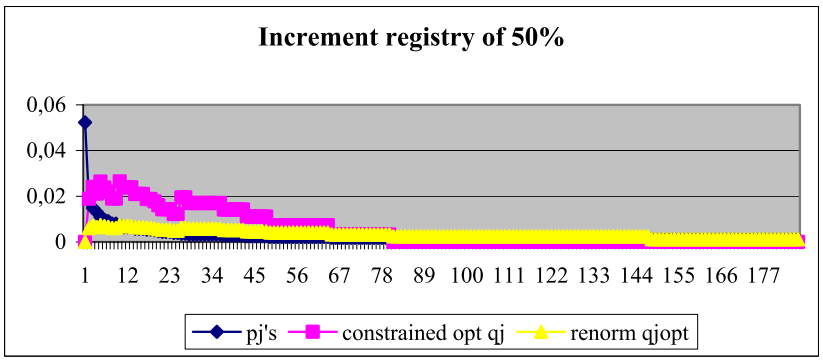
a = 1/3

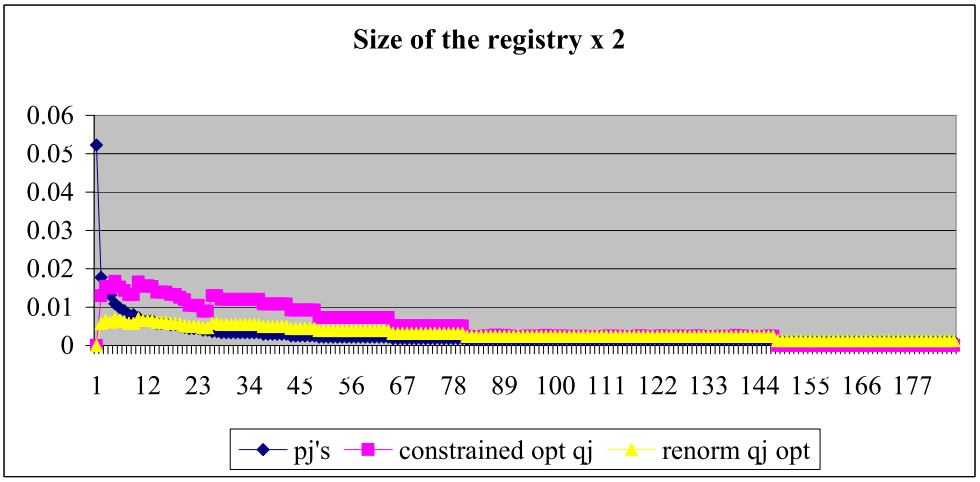
MADO subsample

N	No selection	Optimal	Optimal realistic	Optimal renormalized
10% N0	**0.72**	**0.63**	**0.71**	**0.73**
50% N0	**0.69**	**0.61**	**0.6664**	**0.698**
N0	**0.65**	**0.589**	**0.6294**	**0.6569**
2 N0	**0.59**	**0.547**	**?**	**0.5903**

These simulations result are illuminating: imposing positivity constraints change dramatically the optimum which increases to be closed to the case of no selection. The renormalization procedure is not an efficient way to reach the registry design.

An interesting element is given by the shape of the optimal sampling probabilities. It appears clearly that optimal selection mechanism should over represent the frequent type (at the exception of a very few number of "very frequent" people) and eliminated a large number of rare types (see graphics 1 and 2).





5 Some order of magnitude for the national level (France)

The last step of our computations is to give some quantitative elements about the efficiency of a donor registry at the national level. We have shown that a lower bound for the probability not to find a donor is given by $\Pi_o = J\bar{p}(1-a)^{\frac{No}{J}} e^{\frac{-aN}{J}}$. This value minimizes the probability not to find a donor for any sampling scheme of the donors. As we noticed before, this value depends on a few elements:

- The size of the initial registry No. This number is known and approximatively equal to 100 000.

- The probability a to be available for a transplant. We fix a to be equal to $\frac{1}{3}$. This evaluation corresponds to the experience of FGM.

- The size of the new registry N_0 for which we will try different values.

- The two main unknown elements are J and $\bar{p}$ or, equivalently, J and the product $J\bar{p}$.

In the actual FGM registry, Jo is equal to 62220 and $\bar{p}_o = 0,0000129$ ($J_o\bar{p}_o = 0.8$)

A possible scenario may be based on the hypothesis that $J\bar{p}$ is constant and equal to 0.8 and to a size $N = 100\ 000$ (multiplication by 2 of the actual registry). Several values of J may be used to compute Π.

J	Π optimal
100000	0.57
200000	.68
500000	.75
1000000	.77
2000000	.78

If J increases Π converges to $J\bar{p} = 0.8$. Then the maximal efficiency of a registry of 200 000 individuals would be to find a donor for 20% of receivers. The present registry has an efficiency of less than 10%. Then if the number of types is larger than two millions, a multiplication by two of the size of the registry and an efficient system of screening for the donors will imply an increment of the efficiency from 10% to 20%. It should be underlined that implementable screening mechanism may not be so efficient than the optimal one. Moreover, the optimal does not take into account the constraint $q_j \geq 0$ and the optimal Π, under this constraint, is smaller than 20%.

We have computed another example for the Netherlands case. We have a file of 32 500 donors representing 22983 types and a geometrical means of the frequency equal to 0.0000364. Then, using the same hypothesis as in the French case ($J\bar{p}$ constant, $a = \frac{1}{3}$, $N = No$), if the number of types in the Netherlands is 100 000, the optimal probability not to find a donor is .83. Equivalently the optimal proportion of finding a donor is .17.

REFERENCES

Feve F, Florens J.P. (2005), "Matching Models and Optimal Registry for Voluntary Organ Donation Registries", University of Toulouse