# Maximum number of live births per donor in artificial insemination 

Charlotte Wang ${ }^{1}$, Miao-Yu Tsai ${ }^{4}$, Mei-Hsien Lee ${ }^{1}$, Su-Yun Huang ${ }^{5}$, Chen-Hung Kao ${ }^{5}$, Hong-Nerng Ho ${ }^{2}$ and Chuhsing Kate Hsiao ${ }^{1,3,6}$<br>${ }^{1}$ Division of Biostatistics, Institute of Epidemiology, ${ }^{2}$ Department of Obstetrics and Gynecology, National Taiwan University Hospital and College of Medicine, ${ }^{3}$ Department of Public Health, National Taiwan University, Taipei, Taiwan, ${ }^{4}$ Institute of Statistics and Information Science, National Changhua University of Education, Chang-Hua, Taiwan and ${ }^{5}$ Institute of Statistical Science, Academia Sinica, Taipei, Taiwan<br>${ }^{6}$ To whom correspondence should be addressed at: No. 17, Xu-Zhou Road, Room 523, Taipei 10055, Taiwan. Tel.: +8862 33228032, Fax: +886 2 23418562; E-mail: ckhsiao@ntu.edu.tw


#### Abstract

BACKGROUND: The maximal number of live births ( $k$ ) per donor was usually determined by cultural and social perspective. It was rarely decided on the basis of scientific evidence or discussed from mathematical or probabilistic viewpoint. METHODS AND RESULTS: To recommend a value for $k$, we propose three criteria to evaluate its impact on consanguinity and disease incidence due to artificial insemination by donor (AID). The first approach considers the optimization of $k$ under the criterion of fixed tolerable number of consanguineous mating due to AID. The second approach optimizes $k$ under fixed allowable average coefficient of inbreeding. This approach is particularly helpful when assessing the impact on the public, is of interest. The third criterion considers specific inheritance diseases. This approach is useful when evaluating the individual's risk of genetic diseases. When different diseases are considered, this criterion can be easily adopted. All these derivations are based on the assumption of shortage of gamete donors due to great demand and insufficient supply. CONCLUSION: Our results indicate that strong degree of assortative mating, small population size and insufficient supply in gamete donors will lead to greater risk of consanguinity. Recommendations under other settings are also tabulated for reference. A web site for calculating the limit for live births per donor is available.


Key words: artificial insemination by donor/assortative mating/coefficient of inbreeding/consanguinity/hereditary disease

## Introduction

A legal restriction on the number of live births is usually decided not just for social or cultural considerations, but also for reduction in genetic risks. The regulation on limitation ( $k$ ) of live births per donor varies greatly in different countries. For instance, the limit in France is 5 (Le Lannou et al., 1998), 10 in UK (Deech, 1998), 6 in Spain (Jones and Cohen, 2001) and 1 in Taiwan; whereas the limit in The Netherlands is 25, based on the coefficient of inbreeding $(F)$ (de Boer et al., 1995). The coefficient increases with respect to the number of matings in which both are artificial increminations by donor (AID) children of a single donor or one of the pair is the donor's natural child or relative. Without knowing whether the pair is related, inadvertent mating may happen because the pair is phenotypically similar, called assortative mating. Other factors also play important roles. For instance, when the demand in gametes is keen and when the number of available donors $(D)$ is small, a large value of $k$ may be urged and it may then induce an increase in $F$. The phenomenon of shortage of donors occurs for both semen and oocyte
donors (Kan et al., 1998; Lo et al., 2003; Thum et al., 2003; Magnus and Cho, 2005; Pennings, 2005; Ferraretti et al., 2006). The shortage is increasing possibly due to the recent policy of removal of donor anonymity (Pennings, 2001; Janssens et al., 2006). Furthermore, because not every registered donor produces AID offspring successfully, the number of donors achieving live births can be smaller than $D$. This proportion $S$ may depend on the quality of donated gametes, standards for acceptable gametes in various fertility centres, procedures used during donor treatment, medical condition of recipient and the medical procedures adopted. Some fertility centres may even discard the semen of a donor that does not produce pregnancy after some time of use. In Paul et al. (2006), only 40 men out of 1101 potential sperm donors were not rejected after a suboptimal semen quality examination. These released sperm donors completed 330 clinical pregnancies. Because the centre has adopted a standard higher than the World Health Organization (WHO) criteria for normal semen quality, its release rate is low, but almost every released donor achieved live births, i.e. their proportion $S$ of achieving
live births is $\sim 1$. However, if the clinic has used a different criterion to increase the recruitment rate of donors, it may then associate with a lower number of clinical pregnancy. Carrell et al. (2002) reported significant differences in semen quality between seven semen banks and within a given bank. Consequently, some among the released $D$ donors may fail to achieve a live birth and $S$ will be $<1$. In fact, there may not be a second screening stage for suboptimal semen quality in some clinics. We call $S$ the effective ratio to stand for the proportion of donors who achieve live birth. It differs from the release rate of donors (Paul et al., 2006) and is not the pregnancy rate (ESHRE, 2006) per transfer for various assisted reproductive techniques (ART) either. Therefore, the product of $D$ and $S$ can be considered the number of donors who successfully achieve pregnancy. Some fertility centres report the number $D$ and some report the value of $D S$.

A more convincing and evidence-based recommendation for the number of live births should be provided on the basis of transmission probability of hereditary diseases, degree of assortative mating, probability of consanguinity, coefficients of inbreeding by kinship and the availability of donors. The formulae proposed by Curie-Cohen (1980) and de Boer et al. (1995) have included many variables. They calculated $F$ from individual viewpoint and set a threshold for the tolerable coefficient, without assuming genetic risks in the model. The number 25 regulated in The Netherlands was derived on the basis of their recommendations. However, further examination of this number has been urged, especially after the event that a sperm donor with 18 children developed a brain disease many years after he stopped donation (Sheldon, 2002; Janssens, 2003), as well as the introduction of a new law on abolition of donor anonymity in The Netherlands (Janssens et al., 2006). Furthermore, the prevalence of hereditary diseases of interest should be considered in the maximal number as well.

In this article, under the assumption of insufficient supply, we first set up a model for $Y$, the potential number of unwittingly consanguineous mating due to AID, and then explore its relation to the number of live births $(k)$, number of donors $(D)$, degree of assortative mating $(C)$ and effective ratio $(S)$. Therefore, an optimal $k$ can be derived under a constrained $Y$ and fixed $D, C$ and $S$. Next, we follow the same idea from previous researchers but adopt the viewpoint from population perspective to derive the average $F$ (Hedrick, 2005) due to donor insemination. In other words, we focus on the coefficient induced by AID ( $F_{\text {AID }}$ ) beyond the already existing coefficient of inbreeding $\left(F_{0}\right)$ in a given society with no AID children. We then examine the relation between the total coefficient of inbreeding ( $F=F_{0}+F_{\text {AID }}$ ) and the number of live births per donor. Our results imply that the influence of $k$ on $Y$ and $F$ will be large if the population size is small, if the number of donors is limited and/or if the degree of assortative mating is strong.

Alternatively, we incorporate the risk of a certain hereditary disease in the computation to evaluate the possible elevation in incidence and prevalence due to AID. Some infertility centres require the donors to take screening tests for genetic disorders such as sickle cell anaemia or chromosome abnormalities (Lewis et al., 1999). This will effectively exclude certain
genetic diseases, but it cannot reduce completely the risks of other hereditary diseases that are not screened. From probabilistic perspective, we construct the corresponding incidence due to donor insemination for any given prevalence of a disease and its mode of inheritance. We next demonstrate how to determine the maximal limit on the basis of the information of disease characteristics, population data and donor statistics under tolerable increased incidence.

## Materials and methods

## Number of consanguineous mating

Curie-Cohen (1980) proposed the formula $Y=D \bar{m} P$ for the number of consanguineous mating, where $\bar{m}$ depends on the number of live births per donor $(k)$ and is the expected number of potential unwitting marriages between an AID child and his/her unknown relative. $P$ stands for the probability of mating between any random pair. We extend this equality to accommodate the effective ratio $S$ and obtain $Y=D S \bar{m} P$ to reflect that $D S$ is the number of effective donors that produce offspring successfully. We will assume the demand of gametes is keen and there is thus no limit on the number of recipients.

The number of donors per year, $D$, can be estimated through registration systems in centres performing ART. The ratio $S$ can be estimated using reported proportion of donors with AID children. Depending on their definitions, however, some clinics report $D$ and some report $D S$. For instance, $S$ is estimated around $25 \%$ in Taiwan (Bureau of Health Promotion, 2006) and $D S$ is $40 \%$ in Paul et al. (2006). In the following illustrations, we use 80 and $100 \%$ for $S$ in calculations. The computation of expected number $\bar{m}$ of various consanguineous marriages, including half-sibling, biological father or mother, uncle-niece, aunt-nephew, half uncle-niece, half auntnephew, first cousin, half first cousin, first cousin once removed and half first cousin once removed, can be found in Hajnal (1960), with a modification for possible multiple delivery, as listed in the Appendix. The probability of mating $P$ depends on age, geographic location and phenotypes involved in positive assortative mating. Its value can be derived via demographic data, and the magnitude is usually small. For instance, the maximum number of $P$ was around $1.78 \times 10^{-5}$ in Vermont and the minimum was around $7.31 \times 10^{-7}$ in California, USA (Curie-Cohen, 1980). For an area with population size of 23 millions and 1.5 total fertility rate, $P$ is around $1 \times 10^{-5}$. More detailed explanations and calculations can be found in Hajnal (1960), Cavalli-Sforza (1966) and CurieCohen (1980).

## Average level of $F$ in population with AID

The average level of $F$ in a population with AID should include three parts: (i) the original $F$ existing in a population $\left(F_{0}\right)$; (ii) coefficient due to intentional mating involving an AID child and a known relative; (iii) coefficient due to inadvertent consanguineous mating between an AID child and one unknown relative. The second part is simply half of $F_{0}$ once $k$ is positive because the known relative is the recipient's (not donor's) relative. Therefore, the expected coefficient is half of that when relatives of both parents are considered.

The third part concerns the mating between AID child and unknown relative. It is a weighted average of coefficients with respect to different degrees of consanguinity, where the weights depend on the corresponding percentages of mating. The sum of the last two parts is the coefficient of inbreeding due to AID ( $F_{\text {AID }}$ ). The resulting $F$, average coefficient including that due to AID, becomes

$$
F=F_{0}+F_{\mathrm{AID}}=F_{0}+\left[\frac{1}{2} F_{0} \cdot I(k>0)+\sum_{i} \frac{Y_{i}}{M} \cdot F_{i}\right]
$$

Explanations of further notations are in the Appendix. The relation between $F$ and $k$ can be examined through $Y_{\mathrm{i}}$, the number of matings of the $i$ th kinships.

## Incidence of hereditary disease due to AID

The incidence of a given hereditary disease due to AID depends on whether the donor is a carrier, the characteristics of the disease and/or the inheritance mode of the disease. Several countries or clinics enforce some diseases to be screened at the stage of examination before donation. However, there are many hereditary diseases that are not screened or tested, such as the autosomal dominant cerebellar ataxia (Janssens, 2003). In this case, it is of interest to assess the increase in incidence of the disease with respect to the choice for $k$.

The increased incidence can be interpreted as the 'probability of positive number of cases due to AID' $\left(P_{\mathrm{I}}\right)$. It can be shown that this probability is dominated by three factors: (i) the donor carries the affected gene; (ii) the offspring mates with others; (iii) number of offspring of the same affected donor. This probability $P_{\mathrm{I}}$ is approximately linear in $k$ because other non-linear terms are small and negligible. More technical details are explained in the Appendix.

## Results

We demonstrate first in Figure 1A-C the number of consanguineous matings due to AID per year $(Y)$ and in Figure 1D-F the average coefficient of inbreeding $(F)$ with respect to the number of live births ( $k$ ) per donor under different degrees of assortative mating $(C)$, population sizes and numbers of effective donors (DS). Curie-Cohen (1980) suggested 1.44 when considering IQ, stature and ear length as three independent factors for assortative mating. Redden and Allison (2006) selected several traits that have been observed to associate with assortative mating, including body mass index, total energy intake (per kg), depressive symptoms, introversionextraversion personality, uric acid levels, urinary sodium excretion and systolic blood pressure. Here, we consider the numbers of factors to be $3,12,18$ or 24 , and thus obtain correspondingly $1.5(\approx 1.44), 4.3\left(\approx 1.44^{12 / 3}\right), 9\left(\approx 1.44^{18 / 3}\right)$ and $18.5\left(\approx 1.44^{24 / 3}\right.$ ) for $C$, respectively. In Figure 1 A and $D$, the number of consanguineous matings and average coefficient correlate positively with $k$ for an area of 23 millions population, $D S=300$, and 1.5 fertility rate. The increase enhances if $C$ is larger. For instance, when $k=20$, there are $\sim 5.8$ pairs of consanguineous mating due to AID per year and $F$ becomes 151.4
if $C$ is 9 , whereas the number of pairs increases to 11.9 and $F$ reaches 152.8 when $C$ is 18.5 . Figure 1 B and 1 E demonstrates the effect of population size under $D S=300$ and $C=9$. In an area of smaller population, $Y$ and $F$ escalate faster than an area of larger population. Similarly, Figure 1C and F illustrates that both $Y$ and $F$ correlate positively with respect to $k$. Furthermore, if the number of donors is larger, more will be contributed to the overall number of consanguineous mating and average $F$, assuming the demand for gametes is much higher than supply. The fertility rate does not have much influence (values not shown here) and hence has been set to 1.5 throughout all computations.

It should be noted that in Figure 1D-F, the scales for average $F$ including AID (left $y$-axis) has been multiplied by $1 \times 10^{6} \quad(=1000000)$ for better readability. We have adopted $1.501 \times 10^{-4}$ and $1.505 \times 10^{-4}$ as two choices for the tolerance. It is because the tolerance in de Boer et al. (1995) was around $9 \times 10^{-5}$ for $F_{0}$, which is almost $1.0 \times 10^{-4}$. We consider the tolerance $(1+0.5+0.001) \times$ $1 \times 10^{-4}$, where 0.001 represents a ratio of increase compared with $F_{0}$. The coefficients in other countries are of similar magnitude (Bittles and Sullivan, 2005). Alternatively, we have also considered the ratio $F / F_{0}$ as another presentation (right y-axis in Figure 1D-F).

Table I lists the maximal numbers under two different tolerance values of $F$. The multiplication of $D$ and $S$ indicates the number of effective donors considered. Under the threshold $1.501 \times 10^{-4}$ and $C=9$, the maximal number of live births per donor is $\sim 5$ if $D S=240$ for an area of 23 millions and 1.5 fertility rate, whereas the limit is 4 if $D S=300$ (the second row in Table I). On the other hand, when the threshold increases to $1.505 \times 10^{-4}$, the corresponding maximal numbers become 13 and 11 , respectively (the second row in the lower part of Table I). Other combinations of population size, $D S$ and $C$ will produce different values of $k$, as listed in Table I. Table II provides the same information as Table I, but focuses on only half-sibling mating. As expected, the maximal number is larger than that in Table I.

Figure 2A-C illustrates the increased incidence of two autosomal recessive inheritance diseases, the hereditary haemochromatosis with $1 / 250$ prevalence (Ellerbik et al., 2001; Scotet et al., 2003) and cystic fibrosis with $1 / 2500$ prevalence (Ratjeu and Döring, 2003) and an autosomal dominant disease, spinocerebellar ataxia, with $1 / 25000$ prevalence (Manto, 2005). The number of new cases due to AID increases with respect to $k$ and the increase escalates, when the prevalence is higher and when $C$ is large. Figure 2D and E further displays the increased incidence due to AID as a function of $k$ for polygenetic disorders of quantitative trait, such as schizophrenia with $81 \%$ heritability in liability (Sullivan et al., 2003) and around $1 \%$ life time prevalence rate (Tsuang et al., 2001) and depressive disorder with low genetic etiological influence and around $5.9 \%$ prevalence rate (Bland, 1997). When $C=9$ and $k=10$, there will be around 50 and 500 new cases of schizophrenia and depression, respectively. Figure $2 \mathrm{~A}-\mathrm{E}$ is based on the population size of 23 million, $D S=240$ and 1.5 fertility rate.


Figure 1. $(\mathbf{A}-\mathbf{C})$ the numbers of consanguineous matings due to AID per year $(Y)$; $(\mathbf{D}-\mathbf{F})$ the average population coefficients of inbreeding $(F)$ with respect to the maximal number $k$ under different degrees of assortative matings $(C)$, population size and number of effective donors ( $D S$ ). In $(\mathbf{A})$ and $(\mathbf{D}), D S=300$ and the population size is 23 millions. $\operatorname{In}(\mathbf{B})$ and $(\mathbf{E}), D S=300$ and $C=9$. $\operatorname{In}(\mathbf{C})$ and $(\mathbf{F}), C=9$ and the population size is 23 millions. The fertility rate is fixed at 1.5 .

Some caution should be employed while reading the figures. The increased numbers of cases in Figure 2A-C are calculated under 10 kinship relations and 3 relations in Figure 2D-E. It may take a long time to cumulate the risks and observe the cases if all conditions, such as the chance of recruiting affected donors, deficiency of screening policy or technique and
medical advancement, remain constant. Therefore, the displayed values are not immediately comparable to the current incidence rates. Because the value is extremely small when compared with the incidence per year, some may prefer to consider it irrelevant, especially for autosomal dominant diseases (Janssens, 2003).

Table I. The maximal numbers of live births $k$ per donor under tolerance $F=1.501 \times 10^{-4}$ and $1.505 \times 10^{-4}$ (where $1 \times 10^{6}=1000000$ ), respectively

| Tolerance $F$ | Population size (million) | Fertility rate | Number of donors | Effective ratio $S=100 \%$ |  |  |  | Effective ratio $S=80 \%$ |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  | $C=18.5$ | $C=9$ | $C=4.3$ | $C=1.5$ | $C=18.5$ | $C=9$ | $C=4.3$ | $C=1.5$ |
| $F=1.501 \times 10^{-4}$ | 16 | 1.7 | 400 | 0 | 1 | 2 | 5 | 1 | 2 | 3 | 6 |
|  | 23 | 1.5 | 300 | 3 | 4 | 7 | 13 | 3 | 5 | 8 | 14 |
|  | 43 | 1.3 | 500 | 4 | 6 | 10 | 17 | 5 | 7 | 11 | 19 |
|  | 60 | 1.9 | 400 | 5 | 8 | 12 | 21 | 6 | 9 | 14 | 24 |
| $F=1.505 \times 10^{-4}$ | 16 | 1.7 | 400 | 3 | 4 | 7 | 13 | 3 | 5 | 8 | 15 |
|  | 23 | 1.5 | 300 | 7 | 11 | 17 | 30 | 8 | 13 | 19 | 34 |
|  | 43 | 1.3 | 500 | 10 | 15 | 23 | 40 | 12 | 17 | 26 | 45 |
|  | 60 | 1.9 | 400 | 13 | 19 | 29 | 49 | 24 | 15 | 22 | 32 |

$F$, coefficient of inbreeding; $C$, degree of assortative matings.

## Discussion

In this article, we propose three methods to assess the influence of the maximum number of live births per donor under the assumption that the supply of gamete donors is far less than demand. The first one focuses on the number of consanguineous matings. The second approach computes the average $F$. It can provide legislators with an estimate of allowable maximal number at a tolerable average population $F$. Major factors include the degree of assortative mating, population size and number of donors per year. A more careful estimation of the number $k$ based on demographic variables of the area under study can be conducted when setting the limit for legal purposes. Under various conditions, such as population size (16, 23, 43 and 60 millions), fertility rate ( $1.3,1.5,1.7$ and 1.9 ), number of donors ( $300-600$ ) and degree of consanguineous mating, we derive the maximal $k$, as listed in Table I. When population coefficient of inbreeding is set below $1.501 \times 10^{-4}, k$ needs to be smaller. These values of $k$ are $>1$ and $<10$ in UK and 25 in The Netherlands. However, when the constraint for $F$ increases to $1.505 \times 10^{-4}$, most $k$ 's are $>10$. The setting with a population of 16 millions is similar to the case in The Netherlands, although the number of donors and degree of consanguineous mating are approximates. The maximal values, however, are $<25$, suggested in de Boer et al. (1995). The main reason is, in contrast to their $\bar{m}_{i}$ (expected number of consanguineous mating of the $i$ th kinship for an AID child), we consider here the average $F$ from the population perspective with $Y_{i}$ (expected number of inadvertent mating of the $i$ th kinship due to AID) as part of the weight. By doing so, every donor in $D$ (or $D S$ ) will be
included and thus the corresponding $F$ is larger. A more detailed presentation of $F$ versus $k$ can be found in Figure 3. The settings in Figure 3A-D are approximate numbers of those in The Netherlands, Taiwan, Spain and France, respectively (WHO, 2005). It seems that the legal limits 1,6 and 5 in Taiwan, Spain and France, respectively, are conservative, whereas the legal limit 25 in The Netherlands is larger. Calculations of other settings can be done through the author's webpage http://homepage.ntu.edu.tw/~ckhsiao/art.htm.

Our third approach evaluates $k$ on the basis of the incidence of a genetic disease. This is helpful in assessing the probability of disease status of an AID child of a specific donor who was not screened for a hereditary disease, but developed it years after successful transfer. It can also be used when someone needs to predict the possible risk of the child before receiving ART. The only prerequisite information before calculation is the mode of inheritance and prevalence of the disease. The results indicate that there will be more cases if the prevalence of the recessive disease is higher. In addition, there are more cases with recessive than with dominant disease. This is because, for dominant diseases, we assume the parent receiving donated gamete is not affected (recipient is normal homozygote) and the AID child will not mate with those affected (those who carry mutant homozygote or heterozygote). For complex diseases, the probability of offspring's disease status conditional on parents' disease status is required. Figure 2D and E shows the increased incidence due to AID as a function of $k$ for schizophrenia and depressive disorder. The calculations in Figure 2D and E cover only the relations of halfsibling, biological father or mother and grand-parents because

Table II. The maximal numbers of live births $k$ per donor under tolerance $F=1.501 \times 10^{-4}$ and $1.505 \times 10^{-4}$, respectively, for half-sibling matings only

| Tolerance $F$ | Population size (million) | Fertility rate | Number of donors | Effective ratio $S=100 \%$ |  |  |  | Effective ratio $S=80 \%$ |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  | $C=18.5$ | $C=9$ | $C=4.3$ | $C=1.5$ | $C=18.5$ | $C=9$ | $C=4.3$ | $C=1.5$ |
| $F=1.501 \times 10^{-4}$ | 16 | 1.7 | 400 | 3 | 5 | 8 | 14 | 4 | 6 | 9 | 15 |
|  | 23 | 1.5 | 300 | 7 | 11 | 16 | 28 | 8 | 12 | 18 | 31 |
|  | 43 | 1.3 | 500 | 9 | 13 | 19 | 33 | 10 | 15 | 22 | 37 |
|  | 60 | 1.9 | 400 | 14 | 21 | 31 | 53 | 16 | 24 | 34 | 59 |
| $F=1.505 \times 10^{-4}$ | 16 | 1.7 | 400 | 9 | 13 | 18 | 31 | 10 | 14 | 21 | 35 |
|  | 23 | 1.5 | 300 | 17 | 25 | 36 | 62 | 19 | 28 | 41 | 69 |
|  | 43 | 1.3 | 500 | 21 | 30 | 44 | 74 | 23 | 34 | 49 | 83 |
|  | 60 | 1.9 | 400 | 33 | 48 | 70 | 118 | 37 | 54 | 78 | 132 |



Figure 2. Increased number of cases of $(\mathbf{A})$ hereditary haemochromatosis with prevalence $=1 / 250$; $(\mathbf{B})$ cystic fibrosis with prevalence $=$ $1 / 2500$; $(\mathbf{C})$ spinocerebellar ataxia with prevalence $=4 / 100,000 ;(\mathbf{D})$ schizophrenia with prevalence $=1 \%$; $(\mathbf{E})$ depressive disorder with prevalence $=5.9 \%$ versus the maximal number $k$.
the conditional probability of disease status under other kinship is not available. All the increased numbers of cases in Figure 2 are derived on the basis of different kinship relations. Therefore, it may take a long period of time to observe the cases and these values may not be considered comparable to current incidence per year.

Our first approach considers the number of consanguineous mating, and this number is then used to compute the average $F$ in the second approach or to evaluate the increased incidence of a genetic disease in the third approach. It may not be fair to compare directly among the three approaches because they are not based on the same selection criteria. Nevertheless,


Figure 3. Average population coefficients of inbreeding versus $k$ where (A) is for the population size of 16 millions, $D S=320,1.7$ fertility rate; (B) is for the population size of 23 millions, $D S=240,1.5$ fertility rate; $(\mathbf{C})$ is for the population size of 43 millions, $D S=400,1.3$ fertility rate; $(\mathbf{D})$ is for the population size of 60 millions, $D S=320,1.9$ fertility rate.
they do not result in maximal numbers that differ greatly. For instance, in Figure 1A and D, under a somewhat stringent condition for $Y<1$ or $F<1.501 \times 10^{-4}, k$ is around 5 for the population of 23 millions. If $Y$ is set $<2$ and $F$ $<1.502 \times 10^{-4}$, then $k$ is between 5 and 10 when $C=9$. For the third approach, the prevalence and inheritance mode need to be specified a priori. Disease with a higher prevalence always results in more new cases due to AID.

The factors influencing the maximal number are certainly beyond what we have considered here. Further investigations are worth pursuing. For instance, due to self-reporting bias and various screening criteria, the estimates of $D$ and $S$ may be biased. In addition, when gamete donations are possible from other nearby countries or globally (such as the Cryos International Sperm Bank Ltd, 2006), the curves shown here will become steeper and lead to a smaller $k$, if most recipients considered donors from the same company. On the other hand, the act will reduce the risk if it increases the donor pool and lessens the stress of demand. However, as the anonymity policy is no longer followed (such as in North Europe, England and several states in USA), the number of donors may drop dramatically (Janssens et al., 2006), which can result in a shortage of supply and lead to a request of more
live births per donor. A balance between the increase in demand and the request for suppress in $k$ is always a task.

Furthermore, we have assumed the phenotypes to be independent when computing assortative mating coefficient $C$. It is an over-simplified assumption. For instance, correlation between IQ scores may relate to that of education received, and correlation in facial features may depend on correlation between the couple's stature or fitness. In other words, phenotypes are correlated with each other. Therefore, $C$ may be smaller under dependence assumption. On the other hand, it is argued that mating behaviour is much more complex and more factors should be considered (Luo and Klohnen, 2005). Therefore, the determination of its range prior to the computation of $P$ and the maximal number $k$ for the area under study is essential. In addition, due to numerous illegitimate children, there are more inadvertent matings than considered here. In other words, the population $F_{0}$ may be larger and then leads to smaller ratio $F / F_{0}$. A more careful and refined conclusion for $k$ is worth pursuing after these matings are incorporated. As an alternative proposal, Le Lannou et al. (1998) suggested, instead of the number of pregnancies or children, to constrain on the number of families where the children are born because a brother and a sister of the same family are
unlikely to marry. This approach can be incorporated in the current model as multiple deliveries ( $n_{i}$ in the Appendix), and the computation of $k$ can be carried out in the same way. However, our current approach does not include the number of families as a variable. Further modifications are necessary. Finally, to assess the precision of $Y$, we assume $k$ is fixed for every donor and approximate the variance of number of consanguineous matings by $(\bar{m} P)^{2} \operatorname{Var}(D S)$, where $P$ and the expected value $\bar{m}$ depend on demographic data. The variance, $\operatorname{Var}(D S)$, can be estimated, assuming a Poisson distribution. The approximation, however, is much smaller than the magnitude of $Y$ itself (data not shown here). Similarly, the variability in $F$ and in the increased incidence due to AID is both extremely small and negligible. When $k$ varies among donors or when more private donors are involved, its uncertainty needs to be incorporated in the model, and its variability will influence the variance of $Y$ and $F$. As a consequence, the risk of consanguinity may decrease.

## Acknowledgements

The authors are grateful to Sung-Chiang Lin for his help in preparing the calculations on web page. The authors appreciate the comments from reviewers and associate editor, which have greatly improved this manuscript.

## References

Bland RC (1997) Epidemiology of affective disorders: a review. Can J Psychiat 42(2),367-377.
Bittles A and Sullivan S (2005) Consang.net (http://www.consang.net).
de Boer A, Oosterwijk JC and Rigters-Aris CAE (1995) Determination of a maximum number of artificial inseminations by donor children per sperm donor. Fertil Steril 63(2),419-421.
Bureau of Health Promotion (2006) Report of artificial insemination technology in Taiwan, 2004. Bureau of Health Promotion, Department of Health, Taiwan, R.O.C. http://www.bhp.doh.gov.tw/BHP/index.jsp.
Carrell DT, Cartmill D, Jones KP, Hatasaka HH and Peterson CM (2002) Prospective, randomized, blinded evaluation of donor semen quality provided by seven commercial sperm banks. Fertil Steril 78,16-21
Cavalli-Sforza LL, Kimura M and Barrai I (1966) The probability of consanguineous marriages. Genetics 54,37-60.
Cryos International Sperm Bank Ltd. Retrieved March 14, 2006 (http:// www.cryos.dk/).
Curie-Cohen $M$ (1980) The frequency of consanguineous matings due to multiple use of donors in artificial insemination. Am J Hum Genet 32,589-600
Deech R (1998) Legal and ethical responsibilities of gamete banks. Hum Reprod 13(Suppl 2),80-89
Ellervik C, Mandrup-Poulsen T, Nordestgaard B, Larsen L, Appleyard M, Frandsen M, Petersen P, Schlichting P, Saermark T and Tybjaerg-Hansen A (2001) Prevalence of hereditary haemochromatosis in late-onset type 1 diabetes mellitus: a retrospective study. Lancet 358(9291),1405-1409.
ESHRE (2006) Assisted reproductive technology in Europe, 2002 Results generated from European registers by ESHRE. Hum Reprod 21(7), 1680-1697.
Ferraretti AP, Pennings G, Gianaroli L and Magli MC (2006) Semen donor recruitment in an oocyte donation programme. Hum Reprod 21,2482-2485.
Hajnal J (1960) Artificial insemination and the frequency of incestuous marriages J R Stat Soc Series A 123,182-194.
Hedrick PW (2005) Genetics of populations, 3rd edn. Jones and Bartlett Publishers, Sudbury.
Janssens PMW (2003) No reason for a reduction in the number of offspring per sperm donor because of possible transmission of autosomal dominant disease. Hum Reprod 18(4),669-671.
Janssens PMW, Simons AHM, van Kooij RJ, Blokzijl E and Dunselman GAJ (2006) A new Dutch law regulating provision of identifying information
of donors to offspring: background, content and impact. Hum Reprod 21(4),852-856.
Jones HW, Jr and Cohen J (2001) IFFS surveillance 01. Fertil Steril 76(5, Suppl 2), s5-s36.

Kan AKS, Abdalla HI, Ogunyemi BOO, Korea L and Latarche E (1998) A survey of anonymous oocyte donors: demographics. Hum Reprod 13, 1762-1766.
Le Lannou D, Thépot F and Jouannet P (1998) Multicentre approaches to donor insemination in the French cecos federation: nationwide evaluation, donor matching, screening for genetic diseases and consanguinity. Hum Reprod 13(Suppl 2),35-54.
Lewis V, Saller DN, Jr, Kouides RW and Garza J (1999) Survey of genetic screening for oocyte donors. Fertil Steril 71(2),278-281.
Lo B, Chou V, Cedars MI, Gates E, Taylor RN, Wagner RM, Wolf L and Yamamoto KR (2003) Consent from donors for embryo and stem cell research. Science 301,921.
Luo S and Klohnen EC (2005) Assortative mating and marital quality in newlyweds: a couple-centered approach. J Pers Soc Psychol 88,304-326.
Magnus D and Cho MK (2005) Issues in oocyte donation for stem cell research. Science 308, 1747-1748.
Manto MU (2005) The wide spectrum of spinocerebellar ataxias (SCAs). Cerebellum 4(1),2-6.
Paul S, Harbottle S and Stewart JA (2006) Recruitment of sperm donors: the Newcastle-upon-Tyne experience 1994-2003. Hum Reprod 21,150-158.
Pennings G (2001) The reduction of sperm donor candidates due to the abolition of the anonymity rule: analysis of an argument. J Assist Reprod Genet 18,617-622.
Pennings $G$ (2005) Gamete donation in a system of need-adjusted reciprocity. Hum Reprod 20,2990-2993.
Ratjen F and Döring G (2003) Cystic fibrosis. Lancet 361(9358),681-689.
Redden DT and Allison DB (2006) The effect of assortative mating upon genetic association studies: spurious association and population substructure in the absence of admixture. Behav Genet 36,678-686.
Reynolds MA, Schieve LA, Martin JA, Jeng G and Macaluso M (2003) Trends in multiple births conceived using assisted reproductive technology, United States, 1997-2000. Pediatrics 111(5),1159-1162.
Scotet V, Mérour M, Mercier A, Chanu B, Le Faou T, Raguénes O, Le Gac G, Mura C, Nousbaum J and Férec C (2003) Hereditary hemochromatosis: effect of excessive alcohol consumption on disease expression in patients homozygous for the $c 282 y$ mutation. Am J Epidemiol 158(2),129-134.
Sheldon T (2002) Children at risk after sperm donor develops late onset genetic disease. BMJ 324(7338),631.
Sullivan PF, Kendler KS and Neale MC (2003) Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. Arch Gen Psychiat 60(12),1187-1192.
Thum MY, Gafar A, Wren M, Faris R, Ogunyemi B, Korea L, Scott L and Abdalla HI (2003) Does egg-sharing compromise the chance of donors or recipients achieving a live birth? Hum Reprod 18,2363-2367.
Tsuang MT, Stone WS and Faraone SV (2001) Genes, environment and schizophrenia. Br J Psychiat 178(Suppl 40),s18-s24.
World Health Organization (2005) retrived November 10, 2005 (http:// www.who.int/).

Submitted on August 4, 2006; resubmitted on October 1, 2006; accepted on December 2, 2006

## APPENDIX

## The probability of mating ( $P$ )

A male and female, chosen randomly from a population, are more likely to mate if they are similar with respect to age, residence and phenotype as a result of positive assortative mating. Curie-Cohen (1980) assumed that mates were assortatively chosen by the above three factors. We consider the same factors and hence the probability of mating was estimated by multiplying the coefficients of these three. According to Hajnal (1960), Cavalli-Sforza (1966) and Curie-Cohen (1980), de Boer et al. (1995), the probability of mating was

$$
P=2 \bar{d} C \frac{Q}{A}
$$

where $l$ is the chance that a newborn will ever reproduce, $\bar{d}$ the factor of assortative mating for age, $C$ the factor of assortative mating for phenotype, $Q$ the factor that people mate in the area where they were born and $A$ the average number of newborns in a year.

## The number of potential marriages between offspring and relatives ( $\bar{m}$ )

Hajnal (1960) presented a method to calculate the number of potential marriages between a donor's offspring and a relative. He estimated the expected number of unintentional consanguinities under one AID child per pregnancy. However, rates of multiple births using ART rose substantially in the 1980s and 1990s (Reynolds et al., 2003), and hence this model needs modification.

Take half-siblings for example. Under the assumption of multiple use of a single donor's gametes, each donor has $k$ live births and $n_{i}$ $(i=1, \ldots, k)$ AID children per pregnancy, where $n_{1}, \ldots, n_{k}$ are independently and identically distributed with mean $\bar{n}$ and variance $\sigma_{n}^{2}$. Furthermore, assume that every donor has $f$ natural children, where $f$ follows a normal distribution with mean $\bar{f}$ and variance $\sigma_{f}^{2}$. Among the AID children of a given donor, let $v_{i}$ be the number of girls per live birth, $q$ be the probability of giving birth to a girl and $p$ for boy, $p+q=1$, and then $v_{i}$ is a binomial distribution, $\operatorname{Bin}\left(n_{i}, q\right), i=1, \ldots, k$. Among natural children, let $w$ be the number of girls with probability $s$ and $(f-w)$ for boys with probability $r$, where $s+r=1$, then $w$ is binomially distributed.

There are two kinds of unwitting incestuous marriages between half-siblings, those between AID children from different live births and those between AID and natural child. As for the first case, the potential number of consanguinities is

$$
E_{v}\left[\sum_{i \neq j} v_{i}\left(n_{j}-v_{j}\right) \mid n_{i}, n_{j}\right]=\sum_{i \neq j} n_{i} n_{j} p q .
$$

If $p=q=1 / 2$, then the above equation becomes $\sum_{i \neq j} n_{i} n_{j} / 4$. Similarly, for the second case, the potential number of consanguinities is

$$
E_{w, v}\left\{\sum_{i=1}^{k}\left[w\left(n_{i}-v_{i}\right)+v_{i}(f-w)\right] \mid n_{i}, f\right\}=\sum_{i=1}^{k} f n_{i}(p s+q r) .
$$

If $r=s=1 / 2$, then it becomes $\sum_{i=1}^{k} f n_{i} / 2$. Adding both numbers, we obtain the estimated number $m$ of potential matings among the children of a donor. Hence, the expected value of unintentional consanguineous mating between one single donor's children is

$$
\bar{m}=E_{n, f}\left(\frac{1}{4} \sum_{i \neq j} n_{i} n_{j}+\frac{1}{2} \sum_{i=1}^{k} f n_{i}\right)=k(k-1) \frac{\bar{n}^{2}}{4}+k \frac{\bar{n} \bar{f}}{2} .
$$

For other types of consanguinities, $\bar{m}=(1 / 2) k \bar{n}$ for biological father or mother, $m=(1 / 2) k \bar{n}(\bar{f}-1)$ for uncle-niece or auntnephew, $\bar{m}=k \bar{n} \bar{f}^{2}+(1 / 2) k(k-1) \bar{n}^{2} \bar{f}$ for half uncle-niece or half aunt-nephew, $\bar{m}=(1 / 2) k \bar{n} \bar{f}(\bar{f}-1)$ for first cousin, $\bar{m}=(1 /$ 2) $k \bar{n} \bar{f}^{3}+(1 / 4) k(k-1) \bar{n}^{2} \bar{f}^{2}$ for half first cousin, $\bar{m}=k \bar{n} \bar{f}(\bar{f}-1)$ $(\bar{f}+1 / 2)$ for first cousin once removed and $\bar{m}=k \bar{n} \bar{f}^{4}+(1 /$ 2) $k(k-1) \bar{n}^{2} \bar{f}^{3}$ for half first cousin once removed. Detailed derivations can be found on the author's webpage.

## Average level of coefficient of inbreeding in population with AID (F)

The average coefficient $F$, including that due to AID, is

$$
F=F_{0}+F_{\mathrm{AID}}=F_{0}+\left[\frac{1}{2} F_{0} \cdot I(k>0)+\sum_{i} \frac{Y_{i}}{M} \cdot F_{i}\right],
$$

where $I(\cdot)$ is an indicator function, $i$ indexes various degrees of consanguinity, $Y_{i} / M$ is the corresponding proportion of mating and $M$ is the average number of marriages per year. The coefficient of inbreeding due to AID ( $F_{\text {AID }}$ ) is the sum of the last two items. This formula extends what was previously proposed in de Boer et al. (1995) and Curie-Cohen (1980) in two places. First, the effective ratio $S$ is considered in $Y_{i}$. Second, we use $Y_{i} / M$ as the weight when summing $F_{i}$. The effects of $D, S, \bar{m}$ and $P$ on $F$ can be assessed through $Y_{i}$ in the above equation and may vary among countries or clinics.

## The incidence rate of autosomal recessive inherited disease

Taking an autosomal recessive inherited disease for an example. Assume the frequency of normal allele is $p$ and that for mutant allele is $q$. Under Hardy-Weinberg equilibrium, the frequency of diseased is $f$ (diseased) $=q^{2}$ (also the prevalence), the frequency of non-diseased is $f($ normal $)=p^{2}$ and the frequency of carrier is $f($ carrier $)=2 p q \approx 2 q=2 \sqrt{ }$ prevalence. Then, the incidence rate due to AID is

$$
\begin{aligned}
P_{\mathrm{I}}= & \operatorname{Pr}(\text { case } \geq 1 \text { due to AID }) \\
= & \operatorname{Pr}(\text { case } \geq 1 \text { due to AID } \mid \text { donor is normal }) \\
& \times \operatorname{Pr}(\text { donor is normal }) \\
& +\operatorname{Pr}(\text { case } \geq 1 \text { due to AID } \mid \text { donor is carrier }) \\
& \times \operatorname{Pr}(\text { donor is carrier }) \\
& +\operatorname{Pr}(\text { case } \geq 1 \text { due to AID } \mid \text { donor is diseased }) \\
& \times \operatorname{Pr}(\text { donor is diseased })
\end{aligned}
$$

Because the mutant rate of genes is very low, the chance of a child of normal parents being affected is extremely slim and negligible. Therefore, the following probability is close to zero, $\operatorname{Pr}($ case $\geq 1$ due to AID $\mid$ donor is normal $) \approx 0$. It is non-zero only when mutation occurs. Then, the probability and its approximation are

$$
\begin{aligned}
P_{I} \approx & \operatorname{Pr}(\text { case } \geq 1 \text { due to } \operatorname{AID} \mid \text { donor is a carrier }) \\
& \times \operatorname{Pr}(\text { donor is a carrier }) \\
& +\operatorname{Pr}(\text { case } \geq 1 \text { due to AID } \mid \text { donor is diseased }) \\
& \times \operatorname{Pr}(\text { donor is diseased })
\end{aligned}
$$

where each conditional probability depends on $Y$, probability of having an affected AID child, average number of children per couple and average number of new borns per year. The two marginal probabilities of donor being carrier or diseased depend on the prevalence of the disease. All these probabilities can be obtained
or derived from demographic data. Next,

$$
\begin{aligned}
P_{\mathrm{I}} & =\operatorname{Pr}(\text { case } \geq 1 \text { due to AID }) \\
& \approx \operatorname{Pr}(\text { case } \geq 1 \text { due to AID }
\end{aligned}
$$

donor is carrier and recipient is normal)
$\times p^{2} \times 2 p q+\operatorname{Pr}$ (case $\geq 1$ due to AID
donor is carrier and recipient is carrier)
$\times 2 p q \times 2 p q+\operatorname{Pr}$ (case $\geq 1$ due to AID
donor is carrier and recipient is diseased)
$\times q^{2} \times 2 p q+\operatorname{Pr}$ (case $\geq 1$ due to AID
donor is diseased and recipient is normal)
$\times p^{2} \times q^{2}+\operatorname{Pr}$ (case $\geq 1$ due to AID
donor is diseased and recipient is carrier)
$\times 2 p q \times q^{2}+\operatorname{Pr}$ (case $\geq 1$ due to AID
donor is diseased and recipient is diseased) $\times q^{2} \times q^{2}$
where $\operatorname{Pr}$ (case $\geq 1$ due to $\operatorname{AID} \mid$ donor is carrier and recipient is normal) $=\sum_{i} Y_{i} \times \bar{f} \times P_{i}^{\prime} / A, A$ represents the average number of newborns per year, $Y_{i}$ the number of the $i$ th type of consanguineous
matings, $\bar{f}$ the average number of natural children per couple and $P_{i}^{\prime}$ the probability of the child being affected under the $i$ th type of consanguineous mating. The approximate probability $P_{\mathrm{I}}$ now is taken linear in $k$ because other non-linear terms are small and negligible. By fixing a desirable magnitude for the increased incidence, the maximal allowable number $k$ per single donor can then be determined.

For autosomal dominant disease, the calculation differs.
$P_{I} \approx \operatorname{Pr}$ (case $\geq 1$ due to AID|donor is heterozygous)
$\times \operatorname{Pr}$ (donor is heterozygous)
$\approx \operatorname{Pr}$ (case $\geq 1$ due to AID
donor is heterozygous and recipient is normal homozygous)
$\times \operatorname{Pr}($ recipient is normal homozygous)
$\times \operatorname{Pr}$ (donor is heterozygous)
$=\operatorname{Pr}($ case $\geq 1$ due to AID $\mid$
donor is heterozygous and recipient is normal homozygous)
$\times p 2 \times 2 p q$

