A simple method for calculating risks before DNA analysis

MARC JEANPIERRE
From Service de Biochimie Génétique et Unité INSERM U129, CHU Cochin, 24 rue du Faubourg Saint-Jacques, 75014 Paris, France.

SUMMARY Calculation of carrier risk of an X linked disease may be performed on a small computer after DNA analysis, but a method for rapid hand estimation of the risk is still useful for a quick check of the results and weighing the relative importance of each element of information, such as the determination of a haplotype. Each risk estimation is a function of a prior risk and the product of likelihood ratios and these terms are derived themselves from parameters such as fitness or the relative mutation rate in male and female gametes. Even if it is often difficult to have strong experimental estimation of these variables, the existence of a normal father or grandfather must be considered whenever male fitness is not null. The likelihood ratio for a woman for not being a carrier, when her father is not affected and her mother has herself a likelihood R for not having the mutated gene, may be expressed as the ratio \(2R/(C_m R + 1)\), with \(C_m\) being a function of male fitness and relative mutation rate. \(C_m\) represents the odds ratio for the mother of a carrier not to be a carrier, given that the father of the known carrier is not affected. This formula can be used recurrently and reduces to \(2R/(R + 1)\) in lethal X linked disease. When likelihood ratios are expressed as an algebraic function, maximum values are easily determined, hence fixing the limits of DNA analysis.

In recent years at least four computer programs have been introduced which can provide a genetic risk figure combining pedigree information, carrier detection tests, and DNA analysis.\(^1\)\(^-\)\(^4\) These computations are especially complex when the phase of the marker with the disease gene is unknown or flanking probes are used. Geneticists face one or other situation in most cases of X linked diseases and computer analysis of the data is the only sensible solution to the problem. However, there is still a requirement for hand calculation of pedigree data before DNA analysis to avoid molecular studies in families where potential carriers have a very low risk, or to determine in others those persons who need further study. Most of the parameters needed for computation are derived from experimental data, and relative mutation rates in male and female gametes, male fitness, and quantitative carrier detection tests are only estimates, so it may be wise to use only simple calculations. There may also be a special role for hand computing as a rough estimation of the risk to check the computed result rapidly: the computer program is assumed to be right but it may be fed with incorrect data. This work will be limited to X linked diseases.

What is the shortest way to the risk figure?

The risk is often expressed using the odds ratio for the disease, which is more easily comprehended by relatives of the patient. The most simple expression of a Bayesian calculation is:

\[
\text{Posterior odds} = (\text{prior odds}) \times (\text{likelihood ratios}).
\]

If we have two or more independent elements of information, the computed likelihood ratio is the product of likelihood ratios for each element. Working with odds ratios has the double advantage of a concise and widely understood figure of risk estimation.\(^6\) There is, however, a serious drawback in the genetic context: each time we jump from one generation to the next or to the previous one, the probability is simply multiplied by a factor such as 0.5, while this is not possible for odds ratios. The relation between odds ratios and probability is usually written as:

\[
\text{odds} = \frac{\text{probability}}{1 - \text{probability}}.
\]

With probability below 0.5 a literal use of this
The odds of a geneticist arriving between the expressing of a mother and the inverse of the two carriers of a mutated gene can be calculated using the formula:

\[ d = \frac{1}{d_{M}} = \text{odds for not being a carrier} + 1. \]

Thus, if, for example, \( p = 0.027 \) then \( d = \frac{1}{p} = 37 \) and the odds are 36:1. In the clinical context, any geneticist arriving at a probability figure of 0-027 for being a carrier would be likely to convert it to a ratio of two integers, such as 1/37. If we are dealing with the inverse of the probability we may take advantage of a simple formula expressing the relation between the risk of a daughter (1/d) and the risk of her mother (1/dM):

\[ d_D = (R_D + 1) \times d_M. \]

\( R_D \) is the likelihood ratio for the daughter for not being a carrier; the latter relation is just a way of expressing the risk of the daughter as Pr(daughter is a carrier if mother is a carrier) \( \times \) Pr(mother is a carrier), assuming that only one mutation has occurred in the pedigree. The former factor is derived from the relation between odds and probability:

\[ \frac{1}{\text{Pr(D carrier if M carrier)}} = \frac{\text{prior odds}}{(\text{likelihood ratio}) + 1}. \]

The existence of a healthy son is an element of information which may be represented by the ratio of two probabilities. Such a likelihood ratio may be defined as being higher than the unit for each situation where the probability of having the mutated gene is lower. The likelihood ratio for a mother of a healthy son is then:

\[ R = \frac{\text{Probability of having a healthy son if non-carrier}}{\text{Probability of having a healthy son if carrier}} = \frac{1/0.5 = 2}{1 \times 0.5 = 2}. \]

One reason for the choice of the likelihood of not being a heterozygote as a computation factor is easier hand calculation (that is, \( 2^3 = 0.5^3 \)) and another is a convenient scaling of the factors since calculations are always aimed to lower the risk: the odds ratios for being healthy or a non-carrier parallels the likelihood ratios. This choice is arbitrary and not grounded in mathematics as the alternative formula would give the same result with an appropriate formula, since \( R(\text{carrier}) = 1/R(\text{non-carrier}) \). If necessary, we may compute the probability of being a healthy homozygote from the expression of \( d \):

\[ \text{Pr (non-carrier)} = 1 - \frac{1}{d} = \frac{d - 1}{d}. \]

As a rule, starting from the most proximal known carrier of the mutated gene, either a hemizygote or heterozygote, it is possible to incorporate in the computation all the information available as the product of a prior odds ratio and likelihood ratios, and to derive the risk for the rest of the family with the above formulae.

**Example**

Fig 1 is an example of a lethal recessive X linked disease, Duchenne muscular dystrophy, where the prior risk in a mother of an isolated case is 2/3, and thus prior odds are 0.5:1, if we assume in this section that the mutation rate is equal in male and female gametes. In this example, the probability that \( A \) is a heterozygote is \( 1/d \) with \( d = 1+0.5 \times R \), with \( R \) being the product of the following likelihood ratios:

\[ R = 2 \text{ for one son.} \]

\[ R = 3 \text{ for a biological test (CK) in the normal range.} \]

This figure stems from a two class division of the population: about 33% of known carriers have a CK value below a limit defined by the normal 95th centile. Emery had suggested further dividing the normal range of values into more than two discrete classes and assigning to each class the ratio of two probabilities, the four likelihood ratios being 5-20, 4-48, 2-35, and 0-24. Indeed there is no reason to consider a limited number of classes, since Emery's original data showed that the log of CK fits a normal distribution nicely. The likelihood ratio for having a CK of any value may be described as a ratio of two probability density functions, such as two Gaussian curves. This avoids a 10 times jump
when CK rises from 65 IU to 75 IU and such a ratio is already taken into account in some computer programs. The four parameters are the mean and standard deviation of the two populations (carrier and non-carrier). The existence of a value in the normal range must still be expressed by a ratio of 3 if not obtained with the protocol of the population study, since this method may lead to odd results.

\[ R = \frac{1}{0.5 + 0.5 \times R_D} = \frac{2 \times R_D}{R_D + 1} \]

Since \( R \) is itself a likelihood ratio, recurrent use of this formula is possible in multigeneration families.

\[ R = \frac{1}{0.5 + 0.5 \times R_D} = \frac{2 \times R_D}{R_D + 1} \]

If \( R \) is 1.33 for each daughter with an unknown CK but who has one normal son: \( R = 4/(2 + 1) \).

\[ R = 1.82 \]

For the last daughter of this example, since \( R = 2 \times (3 \times 2 \times 12/7)/(3 \times 2 \times 12/7 + 1) = 144/79 \) (each daughter with a CK in the normal range and one normal son is represented by a ratio of \( 2 \times 7/(6 + 1) = 12/7 \)).

The final result is thus:

\[ d_A = 1 + 0.5 \times 2 \times 3 \times 1.5 \times 1.33 \times 1.82 = 11.9 \]

The probability that \( A \) is a carrier is close to 1/12 and the odds for \( A \) not being a carrier are 10.9:1. The risk of daughter \( D \) is calculated as follows:

\[ d_D = 2 \times 3 \times 1.72 + 1 \times 11.9 = 135 \]

Daughter \( D \) has a risk of 1/135 of being a carrier.

Do we have to consider the father?

The notation of the section is derived from the modernisation of the classical usage introduced by Edwards. If the X linked disease has a ratio of mutation in male to female gametes of \( k \), and the fitness of affected males is \( f \), the odds for a mother of an affected child not being a carrier are:

\[ \text{Prior odds} = \frac{1 - f}{k + 1 + f} \]

Table 1 shows the values of this factor as a function of fitness of male hemizygotmes and the relative mutation rate. The phenotype of the father of the person seeking advice is usually known and, with this knowledge, the expression of the prior probability is more complex, when the disease is not lethal. We may define one more likelihood ratio \( (A_f) \) which may be used together with the general relation.

\[ A_f = \frac{1}{\text{Probability that the father of a known carrier is unaffected}} \]

If \( q \) stands for the frequency of affected males, \( h \) being the proportion of heterozygotes and \( u \) the mutation rate at oogenesis, we have:

\[ A_f = \frac{h}{h - qf} \]

This ratio is 1.00 where fitness is null and 1.64 where fitness = 0.70 and \( k = 1.00 \) (in Becker type muscular dystrophy, for example), as shown in table 2.

If one considers now what may be learned from the maternal side, such as healthy brothers or maternal biological tests in the normal range, it is necessary to calculate the probability that the mother of a known carrier is herself a carrier. This figure has been known for more than 50 years to be 0.5 in one population, but in a family where the father of an obligate carrier is unaffected and \( f > 0 \), a correction must be made.

It is convenient to consider the odds ratio for a mother of a known carrier not being a carrier, given a normal phenotype of the father of the carrier, and this ratio may be designated \( C_m \), standing for a maternal carriage constant. If we consider the
TABLE 2  Likelihood ratio for not being a carrier when the phenotype of the father of the person for whom risk is calculated is known to be normal. This ratio is notated $A_f$ here.

<table>
<thead>
<tr>
<th></th>
<th>$k=1/8$</th>
<th>$k=1/4$</th>
<th>$k=1/2$</th>
<th>$k=1$</th>
<th>$k=2$</th>
<th>$k=4$</th>
<th>$k=8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f$</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.10</td>
<td>0.09</td>
<td>0.09</td>
<td>0.08</td>
<td>0.08</td>
<td>0.07</td>
<td>0.06</td>
<td>0.06</td>
</tr>
<tr>
<td>0.20</td>
<td>0.19</td>
<td>0.18</td>
<td>0.17</td>
<td>0.16</td>
<td>0.14</td>
<td>0.12</td>
<td>0.12</td>
</tr>
<tr>
<td>0.30</td>
<td>0.29</td>
<td>0.28</td>
<td>0.26</td>
<td>0.24</td>
<td>0.22</td>
<td>0.19</td>
<td>0.19</td>
</tr>
<tr>
<td>0.40</td>
<td>0.39</td>
<td>0.38</td>
<td>0.36</td>
<td>0.34</td>
<td>0.31</td>
<td>0.28</td>
<td>0.28</td>
</tr>
<tr>
<td>0.50</td>
<td>0.49</td>
<td>0.48</td>
<td>0.46</td>
<td>0.44</td>
<td>0.42</td>
<td>0.39</td>
<td>0.39</td>
</tr>
<tr>
<td>0.60</td>
<td>0.59</td>
<td>0.57</td>
<td>0.55</td>
<td>0.53</td>
<td>0.51</td>
<td>0.48</td>
<td>0.48</td>
</tr>
<tr>
<td>0.70</td>
<td>0.69</td>
<td>0.67</td>
<td>0.66</td>
<td>0.64</td>
<td>0.62</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>0.80</td>
<td>0.89</td>
<td>0.87</td>
<td>0.86</td>
<td>0.84</td>
<td>0.82</td>
<td>0.80</td>
<td>0.80</td>
</tr>
<tr>
<td>0.90</td>
<td>0.99</td>
<td>0.98</td>
<td>0.97</td>
<td>0.95</td>
<td>0.93</td>
<td>0.91</td>
<td>0.91</td>
</tr>
</tbody>
</table>

pedigree in fig 2, the risk for B being heterozygous may be computed with the help of the same relation as in the former section:

$$d_B = (1 + \text{odds} \times R_B) \times d_A.$$  

The odds for a mother of a carrier not being a carrier, given that the father of the carrier is healthy, are:

$$C_m = \frac{u(1 + k)}{h/2} = \frac{(1 + k)(1 - f)}{k + 1 + f}.$$  

The likelihood ratio including the information for ascendants on the maternal side not being carriers is then:

$$A_m = \frac{1}{1/(C_m + 1) \times R_m + (1 - 1/(C_m + 1))}.$$  

$$A_m = \frac{(C_m + 1) R_m}{C_m R_m + 1}.$$  

When $f = 0$, $C_m$ reduces to 1 and computation of risk from data from the mother or daughters shows obvious similarities, which are exemplified in fig 3. It may be noted that $A_f \times (C_m + 1) = 2$, since $A_f \times (C_m + 1) = \frac{2(k + 1 + f)(1 + k)(1 - f)}{k(2 - f) + 2 + (k + 1 + f)} = 2$.  

Fig 2  Pedigree of family with an X linked lethal disease. The risk for A of being a carrier may be computed in one step and the risk for B expressed as a factor of the former result: $d_A = 1 + 0.5 \times 4 \times 32 / 17 = 81 / 17$  

$d_B = 81 / 17 \times (16 + 1) = 81$.

Fig 3  These two hypothetical families illustrate the homology of the expression of the risk for subjects A and B. The expression of $d_A$ is for both families: 

$$d_A = (R_a + 1) R_a + 1.$$  

This is true only if the male hemizygotes do not reproduce, otherwise the existence of a normal father of A in the family on the right should be taken into account in the calculation.

TABLE 3  Expression of the odds ratios for the mother of a carrier not being a carrier herself when the father of the carrier is known to have a normal phenotype (otherwise, if the phenotype of the father is unknown, this factor would be 1.00 throughout). This carriership constant is notated $C_m$.

<table>
<thead>
<tr>
<th></th>
<th>$k=1/8$</th>
<th>$k=1/4$</th>
<th>$k=1/2$</th>
<th>$k=1$</th>
<th>$k=2$</th>
<th>$k=4$</th>
<th>$k=8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f$</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.10</td>
<td>0.83</td>
<td>0.83</td>
<td>0.84</td>
<td>0.86</td>
<td>0.87</td>
<td>0.88</td>
<td>0.88</td>
</tr>
<tr>
<td>0.20</td>
<td>0.68</td>
<td>0.69</td>
<td>0.71</td>
<td>0.73</td>
<td>0.75</td>
<td>0.77</td>
<td>0.78</td>
</tr>
<tr>
<td>0.30</td>
<td>0.55</td>
<td>0.56</td>
<td>0.58</td>
<td>0.61</td>
<td>0.64</td>
<td>0.66</td>
<td>0.68</td>
</tr>
<tr>
<td>0.40</td>
<td>0.44</td>
<td>0.45</td>
<td>0.47</td>
<td>0.50</td>
<td>0.53</td>
<td>0.56</td>
<td>0.57</td>
</tr>
<tr>
<td>0.50</td>
<td>0.35</td>
<td>0.36</td>
<td>0.38</td>
<td>0.40</td>
<td>0.43</td>
<td>0.45</td>
<td>0.47</td>
</tr>
<tr>
<td>0.60</td>
<td>0.26</td>
<td>0.27</td>
<td>0.29</td>
<td>0.31</td>
<td>0.33</td>
<td>0.36</td>
<td>0.38</td>
</tr>
<tr>
<td>0.70</td>
<td>0.18</td>
<td>0.19</td>
<td>0.20</td>
<td>0.22</td>
<td>0.24</td>
<td>0.26</td>
<td>0.28</td>
</tr>
<tr>
<td>0.80</td>
<td>0.12</td>
<td>0.12</td>
<td>0.13</td>
<td>0.14</td>
<td>0.16</td>
<td>0.17</td>
<td>0.18</td>
</tr>
<tr>
<td>0.90</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
<td>0.07</td>
<td>0.08</td>
<td>0.08</td>
<td>0.09</td>
</tr>
</tbody>
</table>
A simple method for calculating risks before DNA analysis

This is not fortuitous since the product can be shown to be the inverse of the probability for a mother of a carrier being a carrier in the population, and this figure is 0.5. The factor $C_m$ may be obtained from table 3 or computed from $f$ and $k$. The expression of the two likelihood ratios $A_f$ and $A_m$ can be condensed to give:

$$A = \frac{2R_m}{C_mR_m + 1}.$$

When no information can be obtained from maternal ancestors, $R_m$ is 1, and $A$ represents the knowledge of the existence of a normal father:

$$A = 2/(C_m+1) = A_f.$$

Fig 4 is an example of a complex pedigree. What are the risks for each woman:

- if $f=0$ and $k=1$ (that is, DMD),
- if $f=0.70$ and $k=4$ (that is, haemophilia A)?

For $B$, the calculations are directly computed in one step:

- case 1: $d_B = 1+1.78 \times 1.60 \times 1.78 = 11.1$
- case 2: $d_B = 1+0.26 \times 5.20 \times 1.78 = 8.7$

In this case, it may be noted that the existence of three normal brothers and a normal father represents a likelihood ratio of $2 \times 2^3/(0.26 \times 2^3 + 1) = 5.20$.

For $A$, we have:

- case 1: $d_A = 11.1 \times (2^3 + 1) = 100$
- case 2: $d_A = 8.7 \times (0.26 \times 2^3 + 1) = 40$

For $D$, we have:

- case 1: $d_D = 11.1 \times (2^3 + 1) = 55.5$
- case 2: $d_D = 8.7 \times (2^3 + 1) = 43.5$

In the latter case, it may be noted that the expression of the factor to jump from the risk of a mother to her daughter’s risk is independent of the values of $k$ and $f$.

**How useful are ancestors?**

Information from parents may be expressed as a function of the likelihood ratio for the mother of the person for whom risk is calculated. This relation reaches a maximum for $R = C_m \gg 1$ and if we designate this limit $\max(A)$, we have the expression:

$$\max(A) = 2 \frac{R_u}{C_m(1+k)(1-f)}.$$

For lethal X linked diseases ($f=0$), information from the maternal side, such as healthy brothers, could lower the estimation of the risk by at most a factor of 2, while when $f>0$, the maximum weight of information from the parents increases. A high value of $\max(A)$ corresponds to a low frequency of sporadic cases.

How important are uncles of the consultand? If their contribution to the risk of their mother is $R_u$, the factor for the risk of their sister is $2R_u/(C_m R_u + 1)$, while for their niece it is:

$$R = \frac{2(2R_u/(C_m R_u + 1))}{C_m \times 2R_u/(C_m R_u + 1) + 1}.$$

In order to calculate the maximum influence of uncles, the last relation corresponds to the case where no other information besides the father’s phenotype is known. If $C_m R_u \gg 1$, we have:

$$R = \frac{\frac{4}{C_m}}{2 + \frac{4}{3C_m}} = 2/3 \max(A).$$

When $f=0$, the existence of healthy uncles and a normal CK test in the grandmother of the person seeking advice may contribute to the risk estimation by a factor of only 1.33.

**Germinal mosaicism**

If the possibility of germinal mosaicism is taken into consideration when computing the risk, the basic relation for the risk, written as $1/p$, of a mother of a carrier being a carrier herself is:

$$d = 1 + L_1 R_1 + L_2 R_2 + L_3 R_3.$$

where the $L_1$ variables are the ratios of the prior probability for each hypothesis (non-carrier, maternal mosaicism, and paternal mosaicism) to the prior probability of being a carrier, and the $R_1$ variables are the corresponding likelihood ratios. Experimental estimation of prior probabilities and likelihood ratios for a mitotic origin of a mutation is imprecise. This last formula is cumbersome to use and not necessary for most risk estimations. This problem deserves full treatment but even so it should be noted that if germinal mosaicism is not frequent.

**FIG 4** Example of family illustrating the application of the relations described here. Computations are performed (1) when the disease is lethal and the relative mutation rate in male to female gametes is 1 and (2) when male fitness is 0-70 and the relative mutation rate in gametes is 4.
(less than 10% of the non-carriers) and none of the likelihood ratios as computed in the former sections is larger than 20 or so, the ignorance of mosaicism is of little consequence when counselling families before DNA analysis. This is not the case after RFLP studies where the likelihood ratios, that is \((1-\theta)/\theta\), may be very high.

**Discussion**

There are two basic strategies for hand calculation of genetic risk: one can calculate the odds for each possible combination of genotypes, usually in a tabular form,\(^9\) to\(^{11}\) or use an algebraic formula.\(^{12}\)\(^{13}\) This latter choice could be seen as a set of recipes, but it can be more than this since an algebraic formula can be analysed with the usual mathematical tools and give results such as the notion of a maximum for one component. The strategy for risk calculation developed here stems only from practical considerations: since a risk of 0.045 is expressed as 1/22, why not simply calculate the denominator of a fraction whose numerator is 1? If needed, the alternative probability of not being a carrier may be easily computed from \(d\) as \(Pr (\text{not carrier}) = (d-1)/d\). The concern with easy calculation also resulted in the choice of likelihood ratio of not being a carrier as a way of including pedigree data in the computation. The relative weight of each piece of information may be easily compared: 2 for each healthy son, 1·6 for a daughter with two normal sons, 1·78 for three unaffected brothers. The relations between likelihood ratios of a mother and daugher are easily remembered as they have the same expression when fitness of male hemizygotes is null: \(2R/(R+1)\).

Where the mutation is less severe and affected males reproduce, the general expression for the likelihood of being a healthy homozygote when the mother has a likelihood \(R\) is \(2R/(C_m R+1)\), given that the father of the consultand is known to have a normal phenotype. The values of \(C_m\) and the odds for the mother of an isolated case being a healthy homozygote may be either calculated from the estimations of \(k\) and \(f\), or obtained from tables 1 and 3. Fitness of male carriers and relative mutation rates may thus be used as parameters, but these variables are obtained from experimental data and may not be known with certainty when the disease is not lethal. Since different mutations lead to distinct phenotypes, neither their probability of occurrence nor the associated fitness are constant features, even if a few mutations are recurrent, as in haemophilia A.\(^{14}\) Analysis of tables 1 to 3 shows that the probability of a mother of an isolated case being a carrier is strongly dependent on both parameters, while the probability that a mother of a carrier is herself a carrier, when the father of the carrier is unaffected, is dependent only on fitness. The problem of the relative mutation rate in male and female gametes may thus be ignored whenever the risk is calculated in one branch of a multigene pedigree.

Risk estimations after DNA analysis may not be very different in some families, but this should not prevent us from initiating molecular studies in most cases, though one should bear this fact in mind when requesting blood samples. Molecular analysis of the grandparents of affected children may also lead to fruitful epidemiological studies.\(^{15}\)

**References**


Correspondence and requests for reprints to Dr M Jeanpierre, Service de Biochimie Génétique et Unité INSERM U129, CHU Cochin, 24 rue du Faubourg Saint-Jacques, 75014 Paris, France.