Journal of Medical Genetics 1986, **23**, 40–45

A computer programme for estimation of genetic risk in X linked disorders, combining pedigree and DNA probe data with other conditional information

MANSOOR SARFARAZI AND HYWEL WILLIAMS

From the Section of Medical Genetics, Department of Medicine, University of Wales College of Medicine, Heath Park, Cardiff CF4 4XN.

SUMMARY A computer programme is presented for calculating the recurrence risk in X linked disorders, combining pedigree and DNA probe data with other conditional information such as carrier detection tests. The methods of computation are shown in the given examples. The programme can be used with either a single DNA probe or two 'flanking' DNA probes for both familial and isolated case pedigrees. For isolated case families the mutation rate at the disease locus can be taken into account in conjunction with the DNA probe data.

In recent years more than 200 DNA polymorphisms have been characterised in man,[1] and there is no doubt that the number required[2 3] to span the human genome will soon be reached. This has increased the power of linkage analysis for prediction of genetic diseases, especially X linked diseases, such as Duchenne (DMD) and Becker (BMD) muscular dystrophy.[4] Before linkage was shown in these diseases with various DNA markers,[5 6] estimation of recurrence risk for subjects at risk was limited to analysis of the pedigree structure together with any tests of carrier status available. The estimation of risk in large families can be complex and prone to error. A computer programme[7] (PEDIG) has helped to overcome the problem and can handle the additional information produced when carrier detection tests (for example, CK levels) are used. Recently, the use of linked DNA markers in genetic counselling has also been shown to be a valuable means of altering the risk of individuals.[8–9] However, the combined use of pedigree information, carrier detection tests, and DNA analysis is complex and mistakes are made easily. A computer programme (LINKAGE) has recently been introduced,[10] which can handle complex pedigrees, using linked DNA markers and CK information. However, the use of this programme has proved to be difficult for clinicians with no computing experience, and the programme still needs to be tested further. The present paper describes a simple and easy to use interactive computer programme (RISKDNA), which can be used by workers without computer knowledge.

Received for publication 3 June 1985.
Revised version accepted for publication 28 June 1985.

Methods

The programme RISKDNA is written in 'Basic' language and can be used for any X linked disease genetically linked to either a single DNA marker (two alleles only) or to two flanking markers. In principle, the programme combines the pedigree data with DNA and conditional information by the Bayesian methods described by Murphy and Mutalik.[11] In isolated cases, the conditional probabilities (that is, CK levels and normal sons) from aunts, uncles, grandmother, and mother are combined to estimate the prior probability of the consultand. Then, the DNA information from grandparents, parents, and sibs is used to derive the probabilities for possible genotypes of the consultand. The estimated risk for the consultand can be further modified by conditional information, such as carrier detection tests and her normal sons. To illustrate the scope of the programme, the following examples are given for X linked disorders such as Duchenne or Becker muscular dystrophy. The letters B and C represent two 'flanking' DNA markers each with two codominant alleles. If only one DNA marker is available, then θ is the recombination frequency between that marker and the disease locus. In the case of 'flanking' markers, θ_1 and θ_2 are the recombination frequencies between

the disease locus and the distal and proximal markers respectively.

EXAMPLE 1

In this case the phase of the disease in the mother 'M' with both DNA markers is known. Assuming that $\theta_1=0{\cdot}17$ and $\theta_2=0{\cdot}12$, the estimated recurrence risk in consultand 'C' is $\theta_1\theta_2/[\theta_1\theta_2+(1-\theta_1)(1-\theta_2)]$ or (2·72%), as she has inherited alleles 2 for both markers from her mother. In the absence of a double crossover between the disease and the DNA markers, she would have inherited the normal gene at the disease locus from her mother. In this example the information from the other sib is not used because the phase of the disease with respect to both markers is known. Here, the prior risk for the consultand of 50% has been changed to 2·72%, which is substantially lower. This result can be further modified by information from carrier detection tests. The CK testing for the consultand provides a likelihood ratio of 0·12:1 of being a carrier at her age and CK level. Combining this with DNA analysis would give her a final risk of 0·33% for being a carrier.

EXAMPLE 2

A more complicated situation arises when the phase of the DNA markers with the disease gene is not known. The phase of the disease in the mother with the marker alleles can be calculated, using the genotypes of her offspring. The odds of the disease and marker alleles being in coupling in the mother 'M' are first estimated, and then the probability of the consultand 'C' being a carrier is computed. Again this estimation can be modified by additional information from carrier detection tests (CK) and her normal son. For $\theta_1=0{\cdot}16$ and $\theta_2=0{\cdot}12$ and CK likelihood of 0·44:1 for being a carrier, the calculations are given in table 1.

EXAMPLE 3

In this example the phase of the disease with the DNA marker is unknown, and there are no sibs from which to estimate the odds of coupling in the consultand's mother. However, information for the uncle of the consultand is available and can be used to determine the likelihood of her being a carrier. The recurrence risk in the consultand will vary, depending on the gene frequencies of the DNA marker (with rare allele frequency of 'f'). In this situation all the possible genotypes for the grandparents have to be considered; from this the odds of coupling in the mother 'M' are estimated and applied to compute the final risk in the consultand 'C'. A summary of this approach is given in table 2 for $\theta=0{\cdot}15$ and $f=0{\cdot}11$.

When the grandfather's DNA is 1, the odds of the mother 'M' having the disease gene and allele 2 (of the DNA marker) on the same chromosome are (0·0630 + 0·0020 + 0·0108)/0·0786=0·9644, and, similarly, when the grandfather's DNA genotype is 2, the odds of the mother having the disease gene and allele 1 on the same chromosome are (0·0014 + 0·0014)/0·0786=0·0356.

The probability of recurrence risk in the consultand is now equal to $(1-\theta) \times 0{\cdot}9644 + \theta \times 0{\cdot}0356=0{\cdot}8251$.

EXAMPLE 4

This is an isolated case of Duchenne muscular dystrophy (DMD) with four normal sons and two

TABLE 1 *Risk calculation for the consultand in example 2.*

	Coupling of disease gene in mother 'M'			
	B1C1	*B1C2*	*B2C1*	*B2C2*
DNA analysis of				
Affected son (S1)	$(1-\theta1)\theta2=0{\cdot}1008$	$(1-\theta1)(1-\theta2)=0{\cdot}7392$	$\theta1\ \theta2=0{\cdot}0192$	$\theta1\ (1-\theta2)=0{\cdot}1408$
Affected son (S2)	$\theta1\ \theta2=0{\cdot}0192$	$\theta1\ (1-\theta2)=0{\cdot}1408$	$(1-\theta1)\theta2=0{\cdot}1008$	$(1-\theta1)(1-\theta2)=0{\cdot}7392$
Joint	0·0019	0·1041	0·0019	0·1041

Consultand	*Carrier*				*Not carrier*			
In mother 'M' disease is coupled with:	*B1C1*	*B1C2*	*B2C1*	*B2C2*	*B1C1*	*B1C2*	*B2C1*	*B2C2*
Prior	0·0019	0·1041	0·0019	0·1041	0·0019	0·1041	0·0019	0·1041
DNA analysis	0·7392	0·1008	0·1408	0·0192	0·0192	0·1408	0·1008	0·7392
Joint	0·0014	0·0105	0·0003	0·0020	0·00004	0·0147	0·0002	0·0770
Total	0·0142				0·0919			
Posterior	0·0142/(0·0142+0·0919)=0·1340				0·0919/(0·0142+0·0919)=0·8660			
Normal son	0·50				1			
CK result	0·44				1			
Joint	0·0295				0·8660			
Final risk	0·0329				0·9671			

TABLE 2 *Risk calculation for the consultand in example 3.*

Grandfather	*Grandmother*		*DNA analysis on*		*Joint probability*	*Numerical value of probability*
DNA type	*DNA type*	*Phase with DNA*	*Mother*	*Uncle*		
		2 (½)	1-θ	1-θ	(1-f)×f(1-f)×(1-θ)×(1-θ)	0·0630
	2-1					
1 (1-f)	2f(1-f)	1 (½)	θ	θ	(1-f)×f(1-f)×θ×θ	0·0020
	2-2 (f^2)	2	1	1	(1-f)×f^2×1×1	0·0108
2		2 (½)	θ	(1-θ)	f×f(1-f)×θ×(1-θ)	0·0014
	2-1					
(f)	2f(1-f)	1 (½)	(1-θ)	θ	f×f(1-f)×(1-θ)×θ	0·0014
Total						0·0786

In mother 'M' odds of coupling with the DNA allele 2 is: (0·0630+0·002+0·0108)/0·0786=0·9644
odds of coupling with the DNA allele 1 is: (0·0014+0·0014)/0·0786=0·0356
The probability of the consultand being a carrier is: 0·85×0·9644+0·15×0·0356=0·8251

f=rare allele frequency of the DNA marker (f=0·11).
θ=recombination frequency between the disease and the DNA marker loci (θ=0·15).

TABLE 3 *Risk calculation for the consultand in example 4.*

Grandmother	*Carrier*	*Not carrier*
Prior	4 μ	1
4 normal sons	$(0·5)^4$	1
2 daughters with normal son	3/4×3/4	1
Grandmother's CK	199	1
Joint	27·984 μ	1

Mother	*Carrier*	*Carrier*	*Not carrier*
Prior	1/2×27·984 μ	2 μ	1
Affected son	0·5	0·5	μ
Mother's CK	0·28	0·28	1
Joint	1·959 μ	0·28 μ	μ
Posterior	0·605	0·086	0·309

Consultand	*Carrier*	*Not carrier*	*Carrier*	*Not carrier*	*Not carrier*
Prior	0·605	0·605	0·086	0·086	0·309
DNA analysis	0·78	0·22	0·343	0·657	1
Consultand's CK	0·12	1	0·12	1	1
Joint	0·057	0·133	0·004	0·057	0·309
Posterior	0·102	0·237	0·007	0·102	0·552

Final probability that consultand is a carrier=0·102+0·007=0·109

daughters in the second generation each with one normal son. The likelihood ratios of being carriers for the grandmother's, mother's and consultand's CK were 199:1, 0·28:1, and 0·12:1, respectively. Here the information anterior to the consultand 'C' can be used to predict the likelihood for the mother 'M' being a carrier, incorporating pedigree data and CK levels from both mother and grandmother. This information is then combined with the DNA analysis and the CK likelihood of the consultand to achieve the final risk estimate.

The frequency of carrier females in a given population for DMD is 4 μ. In this example, the DNA genotype information on the grandmother and mother cannot affect the likelihood of either of them being a carrier, and gives additional information only for the consultand. The summary of calculations for θ=0·22 is given in table 3.

Probability that consultand is a carrier:
0·102 + 0·007=0·109
and the probability that she is not a carrier:
0·237 + 0·102 + 0·552=0·891.

EXAMPLE 5

This is a pedigree in which the consultand is the aunt of the living affected male. DNA information from

TABLE 4 *Risk calculation for the aunt of the affected male.*

In mother 'M'		*Disease is coupled with*							
		B1C1		*B1C2*		*B2C1*		*B2C2*	
Carrier daughter	Affected son	0·981		0·981		0·981		0·981	
	Normal son	0·614		0·614		0·614		0·614	
	Her DNA analysis	0·015		0·135		0·085		0·765	
	Joint	0·00904		0·08132		0·0512		0·4608	
		Carrier	*Not carrier*	*Carrier*	*Not carrier*	*Carrier*	*Not carrier*	*Carrier*	*Not carrier*
Normal daughter	Son (S1)	0·614	1	0·614	1	0·614	1	0·614	1
	Son (S2)	0·386	1	0·386	1	0·386	1	0·386	1
	Her DNA analysis	0·015	0·765	0·135	0·085	0·085	0·135	0·765	0·015
	CK result	0·16	1	0·16	1	0·16	1	0·16	1
	Joint	0·00057	0·765	0·00512	0·085	0·00322	0·135	0·02901	0·015
	Total	0·76557		0·09012		0·13822		0·04401	
Normal son		0·015		0·135		0·085		0·765	
Joint		0·000104		0·000989		0·000602		0·015514	

Consultand		*Carrier*				*Not carrier*			
Phase in mother 'M'		*B1C1*	*B1C2*	*B2C1*	*B2C2*	*B1C1*	*B1C2*	*B2C1*	*B2C2*
Prior		0·000104	0·000989	0·000602	0·015514	0·000104	0·000989	0·000602	0·015514
DNA analysis of	Herself	0·015	0·135	0·085	0·765	0·765	0·085	0·135	0·015
	Son (C1)	0·386	0·386	0·386	0·386	1	1	1	1
	Son (C2)	0·614	0·614	0·614	0·614	1	1	1	1
	Son (C3)	0·981	0·981	0·981	0·981	1	1	1	1
CK result		3·28	3·28	3·28	3·28	1	1	1	1
Joint		0·000001	0·000101	0·000038	0·009050	0·000080	0·000084	0·000081	0·000233
Total			0·009191				0·000478		
Posterior			0·9506				0·0494		

The numerical values in this table are derived from the following equations (θ_1=0·15 and θ_2=0·10):

$$(1-\theta_1)(1-\theta_2)=0{\cdot}765$$
$$(1-\theta_1)\,\theta_2=0{\cdot}085$$
$$\theta_1(1-\theta_2)=0{\cdot}135$$
$$\theta_1\,\theta_2=0{\cdot}015$$

$$\frac{(1-\theta_1)(1-\theta_2)}{(1-\theta_1)(1-\theta_2)+\theta_1\,\theta_2}=0{\cdot}981 \qquad \frac{(1-\theta_1)\,\theta_2}{(1-\theta_1)\,\theta_2+\theta_1(1-\theta_2)}=0{\cdot}386 \qquad \frac{\theta_1(1-\theta_2)}{\theta_1(1-\theta_2)+(1-\theta_1)\,\theta_2}=0{\cdot}614 \qquad \frac{\theta_1\,\theta_2}{\theta_1\,\theta_2+(1-\theta_1)(1-\theta_2)}=0{\cdot}019$$

the affected male's mother is combined with that obtained from the aunt's sibs (and their children) to derive a provisional risk for the consultand. Finally, the information from carrier detection tests and DNA information from the consultand's children is used to derive the final risk for her. The summary of this approach for $\theta_1=0{\cdot}15$ and $\theta_2=0{\cdot}10$ is shown in table 4.

Discussion

The calculation of recurrence risks, using pedigree data and conditional information from CK levels and healthy sons, can present difficulties. The computation is even more complex when DNA data are included and especially when a new mutation at the disease locus is a possibility. An obvious solution to this problem is a computer programme.

This paper describes an interactive programme (RISKDNA) which is easy to use and does not require computing experience to understand or to operate (see appendix for illustrations of data entry and output for example 1). This programme prompts the user for each item of information and, depending on the type of pedigree data, the logic and order of the questions will vary. The programme will then call for the appropriate subroutine and at the end displays the final risk on the screen. The details of the analysis can also be displayed. This programme can be used for any X linked disease, using either one single DNA marker, or two 'flanking' DNA markers each with two codominant alleles. The programme can incorporate information from pedigree data (for both familial and isolated cases), DNA analysis, and carrier detection tests (such as CK levels). The RISKDNA programme can be used for one or two 'flanking' DNA markers in phase known (see example 1) and phase unknown pedigrees (see example 2), or when the only information available is on the uncle of the consultand (see example 3).

It is also possible to incorporate DNA data into the estimation of risk for sisters (see example 4), mothers, aunts, and grandmothers of isolated cases. In the case of 'flanking' markers, if the phase is only known for one probe, the programme uses the information on sibs to estimate the odds of coupling in the mother of the consultand for the second marker.

If the consultand is the aunt of the affected male, then the programme combines the DNA information from the affected male's sibs with the DNA information from the consultand's sibs (and their children) to estimate the recurrence risk for the consultand. This is further modified by DNA information from her children and the result of any carrier detection tests (see example 5).

It is hoped that the approach and the programme described here will help those involved in genetic counselling, who have little or no experience in computing, by enabling them to estimate recurrence risks in families with X linked disorders without having to go through lengthy mathematical calculations. The programme RISKDNA is freely available to the interested reader, and can be supplied on IBM 5¼" floppy disks. Suggestions for improvement of the programme from users will be particularly welcome.

References

1 Human Gene Mapping. 7th International Workshop. *Cytogenet Cell Genet* 1984;**37**:1–4.

2 Botstein D, White RL, Skelnick M, Davis RW. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* 1980;**32**:314–31.

3 Lange K, Boehnke M. How many polymorphic genes will it take to span the human genome? *Am J Hum Genet* 1982;**34**:842–5.

4 Zatz M, Itskan SB, Sanger R, Frota-Pessoa O, Saldanha PH. New linkage data for the X-linked types of muscular dystrophy and G-6-PD variants, colour blindness, and Xg blood groups. *J Med Genet* 1974;**11**:321–7.

5 Davies KE, Pearson PL, Harper PS, *et al.* Linkage analysis of two cloned DNA sequences flanking the Duchenne muscular dystrophy locus on the short arm of the human X chromosome. *Nucleic Acids Res* 1983;**11**:2303–12.

6 Kingston HM, Sarfarazi M, Thomas NST, Harper PS. Localisation of the Becker muscular dystrophy gene on the short arm of the X chromosome by linkage to cloned DNA sequences. *Hum Genet* 1984;**67**:6–17.

7 Heuch I, Li FHF. PEDIG—a computer programme for calculation of genotype probabilities using phenotype information. *Clin Genet* 1972;**3**:501–4.

8 Harper PS, O'Brien T, Murray JM, Davies KE, Pearson PL, Williamson R. The use of linked DNA polymorphisms for genotype prediction in families with Duchenne muscular dystrophy. *J Med Genet* 1983;**20**:252–4.

9 Kingston HM, Sarfarazi M, Newcombe RG, Willis N, Harper PS. Carrier detection in Becker muscular dystrophy using creatine kinase estimation and DNA analysis. *Clin Genet* 1985;**27**:383–91.

10 Lathrop GM, Lalouel JM, Julier C, Ott J. Strategies for multilocus linkage analysis in humans. *Proc Natl Acad Sci USA* 1984;**81**:3443–6.

11 Murphy EA, Mutalik GS. The application of Bayesian methods in genetic counselling. *Hum Hered* 1969;**19**:126–51.

Correspondence and requests for reprints to Dr M Sarfarazi, Section of Medical Genetics, University of Wales College of Medicine, Heath Park, Cardiff CF4 4XN.

APPENDIX *Input and output of programme RISKDNA for example 1. (Input data are shown in bold.)*

RISK calculation combining DNA, CK, and pedigree information.
(1) Procedure using one marker only.
(2) Procedure using two flanking markers.
(3) Procedure using information from uncle only.
(4) To exit this programme.

Select the procedure no? **2**

Enter the recombination fraction between the 1st marker and disease locus? **0·17**
Enter the recombination fraction between the 2nd marker and disease locus? **0·12**
How many generations have been typed (2/3)? **3**
Enter consultand's relationship to the most recent affected male, as below:

G for the grandmother of the affected male.
M for the mother of the affected male.
S for the sister of the affected male.
A for the aunt of the affected male.

Enter the code? **S**

Is this an isolated case pedigree (Y/N)? **N**

Calculation for 3 generation families, 'phase known'.

Information:

Distal marker is referred to as marker B.
Proximal marker is referred to as marker C.

Enter phase of markers with disease in consultand's mother as:
B1C1, B1C2, B2C1, B2C2
B0C1, B0C2, B1C0, B2C0 (0000 if unknown)? **B1C1**

Enter the result of the consultand as below:
(B1C1, B1C2, B2C1, B2C2, B1C2–1, B2C2–1, B2–1C1, B2–1C2, B2–1C2–1)? **B2–1C2–1**

Enter the marker result for the father of this female as:
B1C1, B1C2, B2C1, B2C2, B0C1, B0C2, B1C0, B2C0, (0000 if unknown)? **B1C1**

Does consultand have any children (Y/N)? **N**
Is there any CK information on consultand (Y/N)? **Y**
Enter consultand's CK likelihood of being a carrier (0 if not known)? **0·12**
Do you want the screen display of this analysis (Y/N)? **Y**

Consultand	*Carrier*	*Not carrier*
DNA analysis	0·0272	0·9728
Children	*****	*****
CK result	0·1200	1·0000
Joint	0·0033	0·9728
Posterior	0·0033	0·9967

The probability of consultand being [B2–1C2–1] and:

Carrier/affected is 0·00 (or 0·33%)
Not carrier/normal is Δ1·00 (or 99·67%)

Hit any key to start again