

# USER-FRIENDLY PROGRAMS FOR PATERNITY CALCULATIONS AND KINSHIP DETERMINATIONS

**Wing K. Fung**

*Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam Road, Hong Kong*

## ABSTRACT

Based on the ideas of conditional probability, Bayes Theorem and pedigree analysis, four computer programs are developed for calculations of likelihood ratios for various paternity and kinship determinations. These programs are flexible to use and user-friendly. The pull-down manual is employed for the calculations and so users can print and check the results easily. This article explains the theory and describes the nice features of the software.

## INTRODUCTION

There are four different programs for handling various common paternity and kinship determinations namely (a) alleged father, (b) alleged father but without DNA typing, (c) incest case, and (d) missing person/kinship determination. Both civil paternity and criminal paternity can be handled by the programs. The programs are powerful and very easy to use. For example, in program (a) alleged father, it can handle the calculations for the standard trio case: mother-child-alleged father, and the motherless paternity case: child-alleged father, with or without DNA typing for the relatives of the mother. It can also handle the alternative hypothesis that, for example, the true father of the child is a relative (such as brother, half-brother, father etc) of the alleged father. Below we are going to explain the theory and methods of the programs.

## ALLEGED FATHER

We first illustrate with a standard trio problem with DNA profiles of the mother (M), the child (C) and the alleged father (AF) at two STR loci TH01 and D7S820 (this can be many loci say 20 or 30 loci). Table 1 gives the population allele frequencies for the two loci (Wong *et al.*, 2000). They are given in 2 decimal places so that readers can easily check the calculations manually. The usual hypotheses for this problem are

$H_0$ : The alleged father is the true father (TF),

$H_1$ : The true father (TF) is an unrelated random man.

The genotypes at TH01 are C: (7, 8), M: (7, 9) and AF: (8, 10) which are all heterozygous. At D7S820, C and M have the same genotypes (7, 9), and AF is homozygous with the genotype (9, 9). Exhibit 1 evaluates the paternity index (PI) for TH01. The result was obtained easily by clicking in values (7, 8) to the box *Child*, (7, 9) to *Mother*, and (8, 10) to *H0: Alleged Father (AF)*, and then clicked *Calculate*. The value 12.5 would appear immediately in the box *Paternity Index*. This was done very quickly by using the built-in pull down manual. Readers can verify the result that  $PI = 1/(2p_8)$  in this case.

At locus D7S820, similar can be done and we obtain the  $PI = 14.29$  (Exhibit 2). For checking, readers can refer to the formula  $1/(p_7 + p_9)$  in this case. The overall PI is equal to 178.6, which is rather high for only testing two loci. A nice feature of our program is that the inputted values and the outputted findings can be saved in a file. They are shown in Chart 1.

Sometimes, the alleged father would argue that he is not the TF but his brother is the TF of C. In this case,  $H_0$  remains unchanged and  $H_1$  becomes

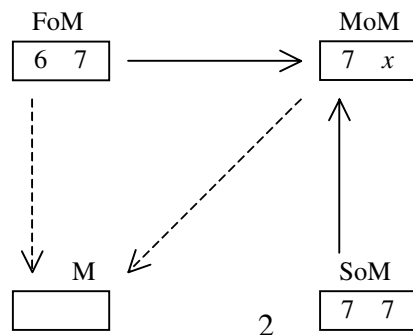
$H_1$ : A brother of the AF is the TF of C.

The calculation can be done easily by clicking *Brother of AF* (there are other choices that one can choose) into the box *H1: True Father*, and then click *Calculate*. The PI's can be obtained immediately and they are equal to 1.85 (TH01) and 1.87 (D7); see Exhibit 3 and Chart 2 for the output. These results can also be derived manually using formulas given in the literature,  $1/[2\{p_8(1-2\times 0.25)+0.25\}]$  for TH01 and  $1/[(p_7+p_9)(1-2\times 0.25)+2\times 0.25]$  for D7. The overall PI drops to 3.46 which is 50 times smaller than that when the common alternative hypothesis  $H_1$ : the TF is an unrelated random man is chosen. Notice that we are only using 2 loci here, and the magnitude of drop would be much higher if we use more loci.

Sometimes, the mother's (M) information may be missing. In such a motherless case, we can do similarly by just inputting the genotypes of C and AF. The associated results for the hypotheses given in (1) are obtained as 6.25 for TH01 and 8.33 for D7. For brevity, the exhibit results are omitted. One can refer to Fung, Wong and Tsui (1996) for the calculation formulas:  $1/(4f_8)$  for TH01 and  $1/(2f_9)$  for D7.

Although the genotype of M is missing, sometimes relatives of M can provide the genetic information. Suppose that the mother of mother (MoM) and the father of mother (FoM) have genotypes (6, 7) and (7, 9) at TH01 respectively. We expect that M would inherit, on average, one "7" from her parents and then pass it to C. Thus, the PI in this case is the same as that if the genotype of M is known to be (7, 9). This is, however, not the case for D7. The derivations for D7 is more complicated, and we only give the result in Table 3. For brevity, the exhibits for both loci are omitted.

Sometimes the genotypes of the parents of M may not be both available. Instead the sibling(s) of M (SoM) is available for typing. There are various combinations of relatives under this circumstance. Table 3 gives the PIs for some of these cases. The idea in the calculations is to infer, from the genotype information of other relatives, the possible genotypes of the *parents of M* by enumeration, which may possible pass on to M and then to C. For example, in the C-FoM-SoM-AF case at TH01, we have to use the genotypes of FoM and SoM to infer the genotype of MoM, which must be (7, x) where x can be any allele at TH01. As a result, the genotype of M can be either one of (6, 7), (7, 7), (6, x) or (7, x). From that, we can compute the PI based on all possible values of x. In the literature, there is little discussion about the evaluations for such complicated situations in general, but our computer program can handle them easily. Exhibit 4 shows the calculations using the software and these and other results are summarized in Table 3. The diagram below shows the concept of the calculation in this situation, which can be easily generalized to deal with other situations.



### **ALLEGED FATHER NOT TYPED**

In case when the AF cannot be typed, the relatives of the AF have to be typed. A computer program which has a more complicated pedigree diagram was written for such a purpose. Previous ideas in Section 2 can be generalized, though the theory would be more involved. Consider an example that the AF is not available and siblings 1 and 2 of the AF have genotypes (9, 10) and (6, 9) at TH01. The PI for this C-FoM-SoM-S1oAF-S2oAF case is astonishing small, 0.133, as compared to 11.0 for C-FoM-SoM-AF if the AF is typed (Table 3). For brevity, the result of the exhibit for this no AF case is omitted.

The new PI is much smaller than one and so it does not support the null hypothesis. Such a low value is not surprising since the two siblings of the AF do not even have allele 8 of which the C is inherited from his TF.

### **INCEST CASE**

Consider a criminal paternity incest case with the genotype information at TH01 given in Table 4. A child ("child 1 of F and M") and his mother are accused to have an incest relationship for giving birth to another child (C). The hypotheses of interest are

$H_0$ : The AF of C is "child 1 of F and M",  
 $H_1$ : The TF of C is an unrelated random man.

This is not a difficult problem and the PI is obtained as 16.67; see Exhibit 5. Suppose the accused puts up an alternative explanation  $H_1$ : The TF of C is actually F, and there is no incest relationship, how should we compute the PI?

If the genotype of F is available, the problem can be solved easily. If it is not, the problem is non-trivial. Consider the case that the genotype of F is unavailable, and his mother's and his 2 siblings are (Table 4). There is little discussion about the calculation for such a case in the literature. However, our software can deal with this problem easily by just clicking in the information of relatives to the software. By choosing the appropriate option in the box *H1*, the PI can be obtained immediately. The PI in this case is reduced dramatically to 0.975, which is even smaller than 1; see Exhibit 6. The figure suggests that the genetic evidence is more in favour of the accused. Of course, the genetic evidence at other loci as well as the non-genetic evidence has to be considered in order to come up with a conclusion.

### **MISSING PERSON**

A person (X) was missing and his family members reported it to the police. A few weeks later, a dead man was found. The police wanted to see whether the dead man was X. The police obtained the DNA profiles of the family members of X including his two children, two siblings, parents, and his spouse (S) as well as his spouse's family members: parents and a sibling. The DNA profiles at D7S820 of this pedigree are presented in Exhibit 7. The hypotheses of interest are  $H_0$ : The dead man was X, and  $H_1$ : The dead man was not X. Although the problem looks complicated, our computer program can handle it easily. The resulted LR is calculated as 13.59 and the generated output is presented in Chart 3.

Exhibit 8 gives the LR by only inputting the genotypes of the spouse and the two children. The LR remains the same; see also the output generated in Chart 4. Actually, at this particular locus only information shown in Exhibit 8 is necessary and the other information is redundant. A nice feature of the software is that it still gives the correct answer even though one has inputted some unnecessary information. This is especially good if one is not sure which information is necessary and which is not. Notice that in this example, the genotypes of the parents and the

sibling of S, and the genotypes of the siblings of X are not needed at all at any loci. But the genotypes of the parents of X are generally needed for other loci. Also, the sibling of S does not look like a child of the father of S, but this does not affect our calculations.

The “missing person” program can also be used for kinship determination. Consider the following situation: Two siblings (S1 and S2) have alleles (8, 9) and (8, 13) at TH01. A man (AS) with alleles (9, 13) alleges that he is a sibling of the other two. A common method used by paternity testing laboratories is to compute the sibling indices (SI's, or avuncular indices in general) for the pairwise comparisons S1-AS and S2-AS. Standard formulas exist for such calculations that are equal to  $(1+2p_9)/(8p_9)$  and  $(1+2p_{13})/(8p_{13})$  for the S1-AS and the S2-AS comparisons, respectively; see Li and Sacks (1954). The corresponding SI's are 2.33 and 4.42. These pairwise comparison figures are often reported by paternity laboratories. However, if (S1, S2) are known true siblings, why shouldn't we compute the SI for (S1, S2)-AS; little discussion is found for this calculation in the literature. Our computer software gives the SI as 12.13 which is even higher than the product of the two pairwise-comparison SI's. Actually, the program we developed is general and it can calculate the SI's for comparisons with more than three siblings.

### CONCLUDING REMARKS

The programs we developed use the pull down manual approach and so they are very user-friendly. They are flexible to use and can handle any number of STR (or RFLP) loci. The pedigree diagrams are provided so that the relationship of the people involved can be visualized easily. The output can be generated and saved for clients and for legal purposes. In addition, the cross checking of the inputting figures can be done easily. The programs can handle cases for civil and criminal paternity, missing persons and kinship determinations. Different sets of hypotheses are built in the software for easy selection. When forensic scientists are not sure what genotype information is needed for the calculations and they input some unnecessary information, the programs can still handle it and give the correct answer. Enquiries or further explanations of the software can refer to hrntfwk@hku.hk.

### ACKNOWLEDGEMENT

I am most grateful to the Symposium Director Carol Zabit for her kindest invitation. This project is partly supported by the Hong Kong RGC Earmarked Research Grant HKU7136/97H.

### REFERENCES

1. Fung W.K., Wong D.M. and Tsui P. (1996) Determination of both parents using DNA profiling, *Jurimetrics Journal of Law, Science and Technology* **36**:337-342
2. Li C.C. and Sacks L. (1954) The derivation of joint distribution and correlation between relatives by the use of stochastic matrices, *Biometrics* **10**:347-360
3. Wong D.M., Law M.Y., Fung W.K., Chan K.L., Li C., Lun T.S., Lai K.M., Cheung K.Y. and Chiu C.T. (2000) Population data for 12 STR loci in Hong Kong Chinese. To appear in *International Journal of Legal Medicine*.

**Table 1 A Paternity Case**

	C	M	AF	FoM	MoM	SoM
TH01	7, 8	7, 9	8, 10	6, 7	7, 9	7, 7
D7S820	7, 9	7, 9	9, 9	7, 7	9, 10	7, 9

**Table 2 Allele Frequencies for Two Loci**

TH01 Allele $i$	Frequency $p_i$	D7S820 Allele $i$	Frequency $p_i$
6	0.10	7	0.01
7	0.32	8	0.13
8	0.04	9	0.06
9	0.44	10	0.16
9.3	0.03	11	0.37
10	0.06	12	0.23
11	0.01	13	0.03
		14	0.01

**Table 3 PI with  $H_0$ : TF of C is AF**

$H_1$ : TF of C is	Locus	C-M- AF	C-AF	C-FoM- MoM-AF	C-FoM- AF	C-FoM- SoM-AF	C-SoM- AF
A random man	TH01	12.5	6.25	12.5	8.99	11.0	10.1
	D7	14.3	8.33	15.4	16.5	15.3	14.1
	Overall	179	52.1	192	148	168	142
A brother of AF	TH01	1.85	1.72	1.85	1.80	1.83	1.82
	D7	1.87	1.79	1.88	1.89	1.88	1.87
	Overall	3.46	3.08	3.48	3.39	3.44	3.40

**Table 4 Alleles of Persons (at TH01) for an Incest Case**

Person	C	M	“Child 1”	MoF	S1oF	S2oF
Allele	10, 13	10, 12	10, 13	9, 13	9, 10	10, 13

```

                                CHART 1
                                Results of paternity testing (alleged father typed)
-----
                                Genotype information for locus : TH01
-----
                                Tested Person      |      Genotype
=====
Mother                          | ( 7, 9)
Child                            | ( 7, 8)
-----
Alleged Father (AF)             | ( 8, 10)
Alternative Father (H1) : A random man
Paternity Index = 12.500000
-----
                                Genotype information for locus : D7S820
-----
                                Tested Person      |      Genotype
=====
Mother                          | ( 7, 9)
Child                            | ( 7, 9)
-----
Alleged Father (AF)             | ( 9, 9)
Alternative Father (H1) : A random man
Paternity Index = 14.285714
-----
                                Overall Summary
-----
Locus tested      |      Paternity Index
=====
TH01              |      12.500000
D7S820           |      14.285714
-----
Overall          |      178.571425
-----
```

CHART 2  
Results of paternity testing (alleged father typed)

-----  
Genotype information for locus : TH01

Tested Person		Genotype
Mother		( 7, 9)
Child		( 7, 8)
Alleged Father (AF)		( 8, 10)

Alternative Father (H1) : Brother of AF

Paternity Index = 1.851852

-----  
Genotype information for locus : D7S820

Tested Person		Genotype
Mother		( 7, 9)
Child		( 7, 9)
Alleged Father (AF)		( 9, 9)

Alternative Father (H1) : Brother of AF

Paternity Index = 1.869159

-----  
Overall Summary

Locus tested		Paternity Index
TH01		1.851852
D7S820		1.869159
Overall		3.461406

CHART 3  
Results of missing person determination

-----  
Genotype information for locus : D7S820

Tested Person		Genotype
=====		
Father of X		( 10, 12)
Mother of X		( 10, 13)
Sibling 0 of X		( 10, 12)
Sibling 1 of X		( 12, 13)
Father of Spouse of X		( 8, 9)
Mother of Spouse of X		( 7, 11)
Spouse of X		( 9, 11)
Sibling 0 of Spouse of X		( 11, 13)
Child of X		( 9, 10)
Sibling 0 of Child of X		( 11, 12)
-----		
Alleged X		( 10, 12)

Likelihood ratio = 13.586957

-----  
Overall Summary

Locus tested		Likelihood Ratio
=====		
D7S820		13.586957
-----		
Overall		13.586957

CHART 4  
Results of missing person determination

-----  
Genotype information for locus : D7S820

Tested Person		Genotype
=====		
Spouse of X		( 9, 11)
Child of X		( 9, 10)
Sibling 0 of Child of X		( 11, 12)
-----		
Alleged X		( 10, 12)

Likelihood ratio = 13.586957

-----  
Overall Summary

Locus tested		Likelihood Ratio
=====		
D7S820		13.586957
-----		
Overall		13.586957



Exhibit 1. Standard Trio at TH01

Calculate Paternity Index (alleged father typed)

LOCUS: TH01

True Father

Mother: 7, 9

Child: 7, 8

H0: Alleged Father (AF): 8, 10

H1: True Father: A random man

Paternity Index: 12.500000

TH01: 12.500000  
Overall: 12.500000

Buttons: Calculate, Reset, Exit, Generate Output

Exhibit 3. Standard Trio with a Different H1

Calculate Paternity Index (alleged father typed)

LOCUS: D7S820

True Father

Mother: 7, 9

Child: 7, 9

H0: Alleged Father (AF): 9, 9

H1: True Father: Brother of AF

Paternity Index: 1.869159

TH01: 1.851852  
D7S820: 1.869159  
Overall: 3.461406

Buttons: Calculate, Reset, Exit, Generate Output

Exhibit 2. Standard Trio at D7S820

Calculate Paternity Index (alleged father typed)

LOCUS: D7S820

True Father

Mother: 7, 9

Child: 7, 9

H0: Alleged Father (AF): 9, 9

H1: True Father: A random man

Paternity Index: 14.285714

TH01: 12.500000  
D7S820: 14.285714  
Overall: 178.371425

Buttons: Calculate, Reset, Exit, Generate Output

Exhibit 4. Motherless Case with Relatives

Calculate Paternity Index (alleged father typed)

LOCUS: TH01

True Father

Mother

Sibling of Mother: 7, 7

Child: 7, 8

H0: Alleged Father (AF): 8, 10

H1: True Father: A random man

Paternity Index: 10.984848

Buttons: Calculate, Reset, Exit, Generate Output

Exhibit 5. An Incest Case with a Random Man H1

Calculate paternity index for incest case

LOCUS: D7S820

H0: Alleged Father of C  
Child 1 of F and M

H1: True father of C  
A random man

Paternity Index: 16.666667

Buttons: Calculate, Reset, Exit, Generate Output

Exhibit 7. Missing Person, Input A11

Calculate likelihood ratio for missing person

LOCUS: D7S820

Likelihood Ratio: 13.586957

Buttons: Calculate, Reset, Exit, Generate Output

Exhibit 6. An Incest Case with a Different H1

Calculate paternity index for incest case

LOCUS: D7S820

H0: Alleged Father of C  
Child 1 of F and M

H1: True father of C  
F

Paternity Index: 0.975410

Buttons: Calculate, Reset, Exit, Generate Output

Exhibit 8. Missing Person, Input Needed

Calculate likelihood ratio for missing person

LOCUS: D7S820

Likelihood Ratio: 13.586957

Buttons: Calculate, Reset, Exit, Generate Output