



Received 5 June 1978  
Final 2 November 1978

# Dependency of Concordance Probability on Gene Frequencies in Genetic Systems for the Diagnosis of Twin Zygosity

## *A Graphical Presentation Enabling the Rapid, Optimal Choice of Genetic System*

Seppo Sarna, Jaakko Kaprio

*Department of Public Health Science, University of Helsinki*

---

The dependency of probabilities of phenotypic concordance of gene frequencies in three-allele genetic systems is presented. A graphical display enables the rapid comparison of the relative effectiveness of different systems, taking into account dominance relationships within each genetic system. Four or more allele systems can also be approximated, while two-allele systems are considered to be special cases of three-allele ones.

**Key words:** Twin zygosity diagnosis, Genetic markers, Multiple alleles, Concordance probabilities

---

### INTRODUCTION

The determination of simple mendelian markers is the most effective method of accurately determining twin zygosity. Because the gene frequencies of blood markers vary in different populations, the same markers or combinations of markers are not necessarily the most effective in zygosity diagnosis in every population. In a previous paper [3] general formulas for determining concordance probabilities, relative odds of dizygosity, and probabilities of misclassification in twin zygosity diagnosis using mendelian markers were presented. In this paper these formulas are applied to derive a set of graphs enabling the rapid comparison of the relative effectiveness of different blood markers. A general case covering three or more alleles with varying dominance relationships is presented. The case for two alleles is a special case, in which the frequency of the third allele is zero, which has been described by Gaines and Elston [2].

### GRAPHICAL DISPLAY

Graphs showing the dependency of concordance probability on gene frequency in markers with three alleles will be presented.

Let us consider a marker  $M$  with three codominant alleles  $A_1$ ,  $A_2$ , and  $A_3$ , that have gene frequencies  $p_1$ ,  $p_2$ , and  $p_3$ , so that  $p_1 + p_2 + p_3 = 1$ . To be able to study in graphic form the effect of varying gene frequencies on the total concordance probability, the frequency of one allele must be kept fixed. The allele the gene frequency of which is to be varied, and the allele with the fixed gene frequency, may be chosen freely in the following consideration.

Let us choose  $p_1 = c$  and  $p_2 = p$ , so that  $p_3 = 1 - p - c$ . The concordance probability for a random DZ pair is in this case [3]:

$$\begin{aligned}
 P_M(\text{Conc}; \text{DZ}) &= P(A_1A_1 \cap A_1A_1; \text{DZ}) + P(A_1A_2 \cap A_1A_2; \text{DZ}) + \\
 &P(A_1A_3 \cap A_1A_3; \text{DZ}) + P(A_2A_2 \cap A_2A_2; \text{DZ}) + \\
 &P(A_2A_3 \cap A_2A_3; \text{DZ}) + P(A_3A_3 \cap A_3A_3; \text{DZ}) \\
 &= \frac{1}{4}c^2(1+c)^2 + \frac{1}{2}pc[1+p+c+2pc] + \\
 &\frac{1}{2}(1-p-c)c[1+(1-p-c)+c+2(1-p-c)c] + \frac{1}{4}p^2(1+p)^2 + \\
 &\frac{1}{2}p(1-p-c)[1+p+(1-p-c)+2p(1-p-c)] + \\
 &\frac{1}{4}(1-p-c)^2[1+(1-p-c)]^2 \\
 &= (\frac{3}{2}c^4 - 3c^3 + \frac{7}{2}c^2 - 2c + 1) + \\
 &(3c^3 - 5c^2 + 4c - 2)p + (\frac{1}{2}c^2 - 7c + \frac{7}{2})p^2 + \\
 &(3c - 3)p^3 + \frac{3}{2}p^4
 \end{aligned} \tag{1}$$

For different values of  $c$  we obtain the family of graphs A shown in the Figure. For each of these curves the restricted minimum, with side-condition  $p_1 = c$ , occurs when  $p_2 = p_3$ , ie, the maximum discriminating power is obtained when the gene frequencies of the other two alleles are equal. The maximum possible discriminating power of a marker with three codominant alleles occurs when  $p_1 = p_2 = p_3$ , obtaining a value of 0.53704 at this point.

When considering a marker with alleles  $A_1$ ,  $A_2$ , and  $A_3$ , with gene frequencies  $p_1$ ,  $p_2$ , and  $p_3$ , such that  $A_2$  and  $A_3$  are codominant and  $A_1$  is recessive with respect to  $A_2$  but codominant with  $A_3$ , a modification of the situation with three codominant alleles is necessary. Then the function in Equation 1 is corrected by adding the function:

$$C_M(p, c) = p^2(1-p-c)(1+p), \tag{2}$$

where  $c = p_1$  and  $p = p_2$ .

The correction function is shown as the family of graphs B in the Figure for different values of  $c$ . For such a marker, the concordance probability  $P_M$  is calculated for given values of  $p$  and  $c$  by adding the values obtained from the graphs in parts A and B in the Figure.

The formulas and graphs for the situation with two codominant or one dominant and one recessive allele genes are special cases of the three allelic genes presented above. They can be obtained by giving the frequency of the third allele as zero. These have previously been presented by Gaines and Elston [2]. An extension to a marker with four alleles can be made by summing the gene frequencies of two codominant alleles and considering this fixed. Then one may proceed as for three allele markers, because then graphical estimation of concordance probabilities is possible.

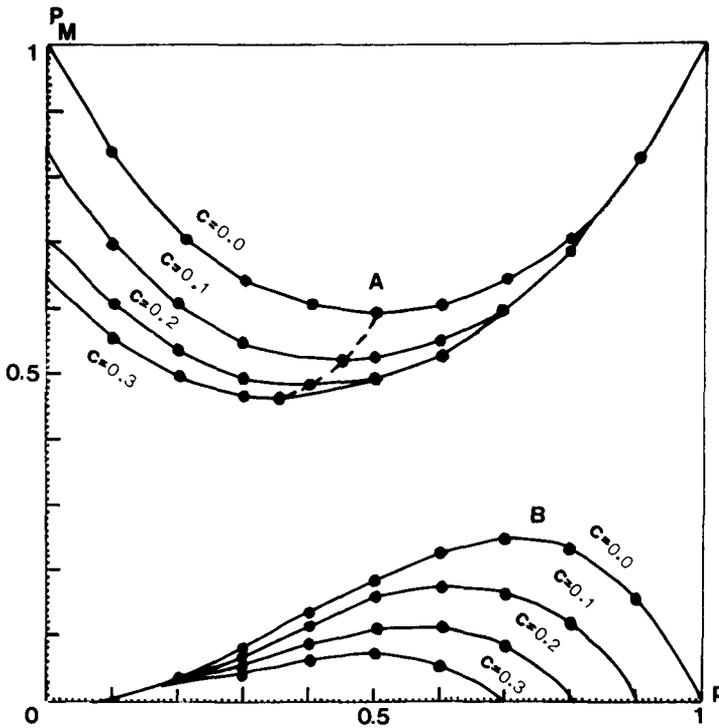


Figure 1. Two families of graphs showing the concordance probabilities for a marker  $M$  with three alleles. A) Family of graphs of concordance probabilities ( $P_M$ ) as a function of the gene frequency ( $p$ ) of one allele with different fixed values of the gene frequency  $c$  of another allele for a marker  $M$  with three codominant alleles. B) Family of graphs of the additive correction functions needed, when one allele is recessive with respect to another, the alleles otherwise being codominant. ---) Graph of optimum values for fixed values of  $c$ .

**NUMERICAL EXAMPLES**

Two examples of the use of the described graphs for estimating concordance probabilities are given using data from the Finnish Twin Registry zygosity determination studies [4]. The acid phosphatase (AP) and Gm systems are used.

**Example 1: AP System**

Three codominant alleles,  $A_1, A_2, A_3$ , with gene frequencies  $p_1, p_2, p_3$ .

Gene frequency	Phenotypes			
	First allele	Second allele		
		$A_1$	$A_2$	$A_3$
$p_1 = 0.330$	$A_1$	$A_1$		
$p_2 = 0.594$	$A_2$	$A_1A_2$	$A_2$	
$p_3 = 0.076$	$A_3$	$A_1A_3$	$A_2A_3$	$A_3$

Let us choose in this case  $c = 0.076$ ,  $p = 0.594$ . Using the curve family A in the Figure and interpolating roughly between  $c = 0.0$  and  $c = 0.1$ , we obtain the approximate value  $P_M \approx 0.56$ . By substituting the values of  $p$  and  $c$  into Equation 1 a value of  $P_M = 0.559$  (to three significant figures) is obtained.

### Example 2: Gm System

Three codominant alleles,  $a$ ,  $ax$ ,  $b$ ; linkage between  $a$  and  $ax$ ; gene frequencies  $p_1$ ,  $p_2$ ,  $p_3$ .

Gene frequency	Phenotypes			
	First allele	Second allele		
		$a$	$ax$	$b$
$p_1 = 0.265$	$a$	$a$		
$p_2 = 0.135$	$ax$	$ax$	$ax$	
$p_3 = 0.600$	$b$	$ab$	$axb$	$b$

Because the correction curves B in the Figure are computed for the allele that is recessive or linked with some other allele, we must choose in this case  $c = p_1 = 0.265$ ,  $p = p_2 = 0.135$ . With this value of  $p$  we obtain, by rough interpolation between  $c = 0.2$  and  $c = 0.3$ , the value 0.53 from the curve family A and the value 0.02 from the curve family B. Adding these two values gives a rough estimate  $P_M \approx 0.55$  for the concordance probability in this case. Substituting the values of  $p$  and  $c$  into Equations 1 and 2 we obtain  $P_M = 0.547$  (to three significant figures).

## DISCUSSION

Zygoty diagnosis is necessary in nearly all twin studies. The determination of simple genetic markers, with known inheritance patterns, is the most accurate method of determining zygoty. At present, increasing numbers of blood groups and other markers are becoming available for zygoty diagnosis. As their gene frequencies vary in different populations, concordance probabilities and associated values must be calculated using the gene frequencies of the population from which the study series is drawn. The choice of markers will depend on the relative efficiency of the markers as well as on the resources available.

Gaines and Elston [2] have presented a family of graphs showing the effect of varying gene frequency on the relative chance of dizygoty in the case of a two-allele marker. In this study the effect of gene frequency variation on concordance probabilities has been expressed for situations with three alleles and the possibility of more alleles has been considered. In addition to this, the effect of having recessive or codominant alleles can be seen. These enable the rapid comparison of marker systems with two or three alleles. Thus, the concordance probability of a given marker system can be compared in different populations. This graphic display can help in the selection of marker systems for zygoty diagnosis.

Selvin [5] discussed the effect of the number of alleles in a genetic system on the efficiency of genetic systems in twin zygoty diagnosis. He shows that the minimum probability of concordance is a function of the number of alleles in a genetic system, provided that the genotypes are known. In practice, however, when parental phenotypes are not known, zygoty diagnosis is decided by determining phenotypic concordance or discordance. Thus the relative efficiency of different markers will depend on phenotypic concordance prob-

abilities. If for any phenotype in the genetic system there exists more than one genotype, the efficiency of the marker will differ from that presented by Selvin, because both the theoretical minimum probability of concordance and the actual probability of concordance will differ. A detailed critique of the pitfalls in Selvin's article has been given by Chakraborty [1]. The minimum value of concordance probability as computed by Chakraborty can also be directly estimated from the Figure when the number of alleles and their dominance relations are known.

The examples illustrate the use of graphical estimation of concordance probability. After obtaining a rough estimate graphically, the final choice of markers for zygoty diagnosis can be done by considering the resources available. The choice of an optimal set of markers is presented and discussed in a further article.

**Acknowledgments.** This study was supported by a grant from the Council for Tobacco Research, USA, Inc.

## REFERENCES

1. Chakraborty R (1979): Twin zygoty diagnosis by genetic systems: An efficiency evaluation. *Acta Genet Med Gemellol* 28:77–79, 1979.
2. Gaines RE, Elston RC (1969): On the probability that a twin pair is monozygotic. *Am J Hum Genet* 21:457–465.
3. Sarna S (1979): Probabilities of concordance of twins with respect to genetic markers. *Acta Genet Med Gemellol: A general formulation*. 28:125–132, 1979.
4. Sarna S, Kaprio J, Sistonen P, Koskenvuo M (1978): Diagnosis of twin zygoty by mailed questionnaire. *Hum Hered* 28:241–254.
5. Selvin S (1977): Efficiency of genetic systems for diagnosis of twin zygoty. *Acta Genet Med Gemellol* 26:81–82.

**Correspondence:** Seppo Sarna, Department of Public Health Science, Haartmaninkatu 3, SF-00290 Helsinki 29, Finland.