



Acta Genet Med Gemellol 40: 159-164 (1991)  
© 1991 by The Mendel Institute, Rome

Received 18 April 1990  
Final 14 September 1990

## Efficient Sequential Search of Genetic Systems for Diagnosis of Twin Zygosity

A.B. Das Chaudhuri

*Human Genetics Laboratory, Department of Anthropology, University of Calcutta, India*

---

**Abstract.** Conventionally, the zygosity of twins into monozygotic (MZ) and dizygotic (DZ) is determined by utilizing several blood groups or other genetic markers. Existence of many blood groups or other genetic markers raises the crucial problem of choosing minimum number of blood groups and applying them in a sequence with minimum cost offering maximum effectiveness in classifying the twins by zygosity. In this note, a statistical approach to this problem is made and a solution provided. The efficient sequential search procedure utilizing the gene frequency of four blood groups from the German population of Nordrhein-Westfalen province and applying the results to a twin set of 72 pairs (32 MZ and 40 DZ) of the same population indicates that efficient sequences with 3 and 4 blood groups are  $ABO \rightarrow P \rightarrow Rh$  and  $ABO \rightarrow P \rightarrow Rh \rightarrow MNSs$ , respectively. Further, it is understood that the efficient sequence for zygosity determination will be differing from population to population. As gene frequencies of blood systems differ in the various populations, this will effect the probability that a DZ twin pairs is concordant for a specific marker. Moreover, the relative cost may also turn out to be a decisive factor in the determination of the best sequence. Hence, indiscriminate use of several conventional blood group systems in the determination of zygosity of twins from a particular population should be discontinued. The formulae developed in this note are quite general and the procedure explained can easily be generalized to other situations.

**Key words:** Zygosity, Efficient sequential search, Genetic markers

---

### INTRODUCTION

The determination of zygosity is a prior task in twin research before its utilization in any kind of human genetic investigation. The usual method in the zygosity diagnosis of a

set of twins is to analyze several conventional blood groups so as to classify the twins as monozygotic (MZ) or dizygotic (DZ). The existence of many blood groups or other genetic markers raises the problem of choosing an optimal number and applying them in a sequence with minimum cost and maximum efficiency.

In this note we consider a statistical approach to this problem and provide a solution.

The probability that a DZ twin is concordant for a specific genetic marker is

$$P(\text{cc/DZ}) = \left\{ \left[ 1 + \sum_{i=1}^N p_i \right]^2 + \sum_{i \neq j}^N (p_i p_j)^2 \right\} / 4$$

where  $p_i$  is the frequency of the  $i^{\text{th}}$  allele in a system containing  $N$  alleles [4]. This probability holds good for a randomly mating population.  $P(\text{cc/DZ})$  has been derived by several authors [3,5] for specific cases of  $N=2$  and  $N=3$  alleles. The derivation for  $N=2$  alleles with frequency  $p$  and  $q=1-p$  is obtained [3] as  $P(\text{cc/DZ})=1-(1/2)pq(4-3pq)$ . Similar results were achieved by other workers [5]. Utilizing Lagrange multipliers it was found [4] that  $P(\text{cc/DZ})$  takes its minimum value when  $p_i=1/N$  for  $i=1,2,3,\dots,N$ . From this, a definition of efficiency for a specific genetic system is suggested [4]:

$$\begin{aligned} \text{Efficiency} &= P(\text{discordance/DZ})/\text{maximum } P(\text{discordance/DZ}) \\ &= \{1 - P(\text{concordance/DZ})\} / \{1 - \text{minimum } P(\text{concordance/DZ})\} \end{aligned}$$

and

$$\text{minimum } P(\text{concordance/DZ}) = \{N(N+1)^2 + N-1\} / 4N^3$$

By comparing 8 common blood groups, it has been found that the more complex, multiple allele systems are not necessarily more efficient for determining zygosity [4].

For our purpose, we have taken into consideration the population gene frequencies of five common blood groups conventionally used in twin diagnosis, ie: AB0 (3), Rh (8), MNSs (4), P (2) and Kell (2), of Germans of Nordrhein-Westfalen province [2] (the quantities in parenthesis indicate the number of alleles in the respective blood group systems). Selection of this sample is due to the simultaneous availability of different blood group gene frequencies and of zygosity-determined twin series [1] with the same blood groups in the same population.

Accordingly,  $P(\text{cc/DZ})$  has been calculated following the general formula mentioned earlier, and this is shown in Table 1 for different systems. The results are in conformity with findings based on different populations [4]. It is readily seen that blood group type K has very little detecting power, since it can detect a DZ twin pair only in about 6% of cases. In view of this, we will not consider using K in this diagnostic twin study. The costs ( $C_1, C_2, C_3, C_4$  and  $C_5$ ) involved for performing tests ( $T_1, T_2, T_3, T_4$  and  $T_5$ ) of five blood groups concerned for a pair of twins have been calculated in US\$ based on the current price list of Ortho-Diagnostic, Indian agent for Ethnor, USA (Table 1).

In the following section, we will develop a sequential search procedure for the four blood tests, designated as  $T_1-T_4$  in Table 1.

**THEORETICAL CONSIDERATIONS**

Consider a population of  $N$  twin pairs of which  $N\alpha$  and  $N\beta$  are MZ and DZ twins, respectively, where  $\alpha + \beta = 1$ . With reference to the German population mentioned above, it is roughly known that  $\alpha = 0.30$  and  $\beta = 0.70$ . It is clear that any blood test will inevitably show concordance when a MZ twin pair is under investigation. On the other hand, a blood test may or may not show discordance when a DZ twin pair is under investigation. Clearly,  $P(cc/DZ)$  is the chance of observing concordance when dealing with a DZ twin pair, and the smaller its value, the better its predicting capacity.

When we apply a specific blood type, say,  $T_{i_1}$ , to the group of  $N$  twin pairs, the cost involved is  $NC_{i_1}$ , irrespective of what results are derived from the test. Next, if  $n_{i_1}$  of the  $N$  pairs have been diagnosed as DZ, then these are left out of further tests and only the remaining  $N - n_{i_1}$  pairs are carried forward and subjected to the next blood test, say,  $T_{i_2}$ . If, again,  $n_{i_2}$  of these  $N - n_{i_1}$  pairs are diagnosed as DZ, then the remaining  $(N - n_{i_1} - n_{i_2})$  pairs are carried forward, and so on. Therefore, the cost involved in administering the sequential search procedure is given by

$$(1) \quad C_{i_1 i_2 i_3 i_4} = NC_{i_1} + (N - n_{i_1}) C_{i_2} + (N - n_{i_1} - n_{i_2}) C_{i_3} + (N - n_{i_1} - n_{i_2} - n_{i_3}) C_{i_4}$$

On the other hand, the number of pairs identified as DZ following the above sequential search procedure is given by

$$(2) \quad n_{i_1 i_2 i_3 i_4} = n_{i_1} + n_{i_2} + n_{i_3} + n_{i_4}$$

It is evident that both the cost  $C_{i_1 i_2 i_3 i_4}$  and the number of identified DZ pairs  $n_{i_1 i_2 i_3 i_4}$  are random variables. It can be shown that the average cost is:

$$(3) \quad \bar{C}_{i_1 i_2 i_3 i_4} = N\alpha(C_{i_1} + C_{i_2} + C_{i_3} + C_{i_4}) + N\beta A_{i_1 i_2 i_3 i_4}$$

where

$$(4) \quad A_{i_1 i_2 i_3 i_4} = C_{i_1} + P_{i_1}(cc/DZ)C_{i_2} + P_{i_1}(cc/DZ)P_{i_2}(cc/DZ)C_{i_3} + P_{i_1}(cc/DZ)P_{i_2}(cc/DZ)P_{i_3}(cc/DZ)C_{i_4}$$

$$(5) \quad \text{Average number of DZ twin pairs identified} = \bar{n}_{i_1 i_2 i_3 i_4} = N\beta B_{i_1 i_2 i_3 i_4}$$

where

$$(6) \quad B = [1 - P_{i_1}(cc/DZ)P_{i_2}(cc/DZ)P_{i_3}(cc/DZ)P_{i_4}(cc/DZ)]$$

In the above,  $P_{i_1}(cc/DZ)$  refers to  $P(cc/DZ)$  when the test applied is  $T_{i_1}$ . Similar interpretation holds for other  $P$  values.

In any case, the average number of DZ pairs identified per unit cost following the sequence  $i_1 i_2 \dots$  is given by  $\bar{n}_{i_1 i_2 \dots} / \bar{C}_{i_1 i_2 \dots}$  and our task would be to choose a particular sequence for which the ratio is a maximum. This would amount to maximum turn over per unit cost.

**RESULTS AND DISCUSSION**

In the next section, we show the computations using the numerical values in Table 1.

**Table 1 - Relevant information concerning different tests**

Tests	Blood group systems	P(cc/DZ)	Cost in US\$ per pair
T <sub>1</sub>	AB0	0.5808	0.06
T <sub>2</sub>	MNSs	0.4400	14.66
T <sub>3</sub>	Rh	0.5027	12.64
T <sub>4</sub>	P	0.5940	2.34
T <sub>5</sub>	K	0.9431	2.36

The available population gene frequencies of five blood group systems of Germans of the Nordrhein-Westfalen Province [2] used for calculation of P(cc/DZ) were the following:

AB0:  $p=0.267, q=0.082, r=0.651$ ;  
 MNSs:  $MS=0.240, Ms=0.289, NS=0.073, Ns=0.398$ ;  
 Rh:  $CDE=0, CDe=0.437, CdE=0, Cde=0, cDE=0.103, cDe=0.021, cdE=0.006, cde=0.433$ ;  
 P:  $p_1=0.486, p_2+p=0.514$ ;  
 K:  $p(K)=0.030, p(k)=0.970$ .

**Case 1. Choosing the Best 3-Test Sequence**

The number of ways of selecting 3 out of 4 blood test types in an ordered form  $4 \times 3 \times 2 = 24$ . For each such ordered sequence, we can compute  $\bar{n} / \bar{C}$  using the formulae in (3) and (5). Thus, for example, for the sequence, T<sub>1</sub>, T<sub>2</sub> and T<sub>3</sub> (say),

$$i_1 = 1, i_2 = 2, i_3 = 3 \quad P_{i_1}(cc/DZ) = P_{T_1}(cc/DZ) = 0.58$$

$$P_{i_2}(cc/DZ) = P_{T_2}(cc/DZ) = 0.44$$

$$P_{i_3}(cc/DZ) = P_{T_3}(cc/DZ) = 0.50$$

$$C_1 = 0.06, C_2 = 14.66 \text{ and } C_3 = 12.64$$

$$\alpha = 0.30, \beta = 0.70;$$

$$\bar{n}_{123} = 61.068, \bar{C}_{123} = 1645.9950; \quad \bar{n}_{123} / \bar{C}_{123} = 0.037$$

For all the 24 ordered triplets, we have calculated the values of  $\bar{n} / \bar{C}$  and compared them. The largest of them corresponds to the sequence 1,4,3, ie, (AB0→P→Rh). We have

$$\bar{n}_{143} = 58.023, \bar{C}_{143} = 853.182; \quad \bar{n} / \bar{C} = 0.068, \text{ as against}$$

$$\bar{n}_{123} / \bar{C}_{123} = 0.037$$

**Case 2. Choosing the Best 4-Test Sequence**

Similar calculations indicate that the 4-test sequence is 1,4,3,2, ie, (AB0→P→Rh→MNSs). Here  $\bar{n}_{1432} = 64.73$  and  $\bar{C}_{1432} = 1468.563$  and therefore,  $\bar{n} / \bar{C} = 0.044$ .

A twin set consisting of 72 pairs (32 MZ and 40 DZ) from the German Nordrhein-Westfalen province, whose zygosity was determined with the help of 10 genetic markers including the blood groups considered here, were studied at the Institute of Human Genetics, University of Munster [1]. The 40 DZ pairs were all identified by using a collection of genetic markers AB0, Rh, MNSs, P, K, Gc, Gm, Hp, Fy and Jk (Table 2).

It may be mentioned that the genetic marker K enables us to identify only 1 out of 40 DZ pairs. This is in agreement with the little theoretical efficiency of this genetic marker. When AB0, P, Rh and MNSs are taken, they collectively distinguish 32 such pairs. Following the sequence AB0→P→Rh→MNSs, we have come up with the findings shown in Table 2.

**Table 2 - Performance of ten genetic markers including AB0, P, Rh and MNSs in a sequence**

Type	Total number of pairs identified	Total number of pairs identified in four genetic markers in a sequence
AB0	17	17 <sup>a</sup>
P	6	2 <sup>a</sup>
Rh	21	11 <sup>a</sup>
MNSs	18	2 <sup>a</sup>
K	2	
Gc	16	
Gm	16	
H <sub>p</sub>	12	
Fy	2	
Jk	10	

<sup>a</sup> In the sequence, at every stage, only the additional pairs identified are shown.

If we decide to use only the popular blood group markers AB0, Rh and MNSs in the sequence AB0→Rh→MNSs, then it is readily observed that we have distinguished 30 out of 40 DZ pairs identified on the whole. Thus, 75% empirical probability is obtained against the theoretical probability of 87% deduced as  $1 - P_{AB0(cc/DZ)} P_{Rh(cc/DZ)} P_{MNSs(cc/DZ)}$ .

The observed proportions of DZ pairs identified by the above sequence of 4 genetic markers is 80%. On the other hand, the theoretical probability is given by the expression B in (6), where we take  $i_1 = AB0$ ,  $i_2 = P$ ,  $i_3 = Rh$  and  $i_4 = MNSs$ . Computations, using P(cc/DZ) values for Table 1, yield a probability of detection of a DZ pair using the sequence AB0→P→Rh→MNSs = 92%. The above findings may be regarded as in agree-

ment with the theoretical results. The observed difference may be due to the small size of the sample.

The above results concerning efficient sequential search procedure with four blood group systems indicate that the relative position of the AB0 blood group in the best sequence is quite consistent with what is commonly done in zygosity determination, ie, it ranks first. The Rh blood group system, also widely used, is preferred to MNSs. We find that the P group ranks second both when choosing three or four blood groups. Therefore, use of the P blood group should be encouraged in zygosity determination. It is to be considered that the efficient sequences of AB0→P→Rh and AB0→P→Rh→MNSs found for our German twins may not be equally applicable in other populations, as the gene frequencies of blood group systems differ from one population to another. It is obvious that the population gene frequency of a blood group system will affect P(cc/DZ) and subsequently the proportion of MZ and DZ twins. Further, the relative cost may also turn out to be a decisive factor in the determination of the best sequences.

It is hoped that this note will be helpful in the efficient application of blood tests to zygosity determination in any population. The formulae given here are quite general and the procedure explained can easily be generalized to other situations.

**Acknowledgments.** I am thankful to Professor B. K. Sinha, Indian Statistical Institute, Calcutta for many useful comments and for critical reading of the manuscript. The financial support of CSIR and UGC, Government of India is acknowledged with thanks.

## REFERENCES

1. Banerjee AR (1974): On the metrical character of the ear of twins. *Endocrinol Sci Con* 31:297-312.
2. Mourant AE, Kopec AC, Domaniewska-Sobczak K (1976): *The Distribution of the Human Blood Groups and Other Polymorphisms*. London: Oxford University Press.
3. Neel JV, Schull WJ (1954): *Human Heredity*. Chicago: The University of Chicago Press.
4. Selvin S (1977): Efficiency of genetic systems for diagnosis of twin zygosity. *Acta Genet Med Gemellol* 26:81-82.
5. Smith SM, Penrose LS (1955): Monozygotic and dizygotic twin diagnosis. *Ann Hum Genet* 19:273-289.

**Correspondence:** Dr. A.B. Das Chaudhuri, Department of Anthropology, University of Calcutta, 35 Ballygunge Circular Road, Calcutta 700019, India.