Empirical and theoretical studies on "seroanalysis"

K. Hummel and N. Fukshansky

Prof. Dr. K. Hummel, Institut für Medizinische Mikrobiologie und Hygiene, Abteilung Blutgruppenserologie, Hermann-Herder-Str.11, D-7800 Freiburg/Brsg.

"Seroanalysis" is a mathematical method by which the population or - in the case of mixed blood - populations to which a person belongs can be established from his or her serotype (Hummel 1980, 1985, 1986). The person in question belongs either to population I or population II, or owns properties from both. If f(I) is the frequency of the phenotype of the person in population I and f(II) that in population II, then K is the likelihood ratio:

$$K = f(II)/f(I).$$

If the person belongs to population I, K will be very small, if to population II, K will be large. -
In practice, the great advantage of seroanalysis lies in the possibility it offers of determining the racial portions, R_I and R_{II}, in a person of mixed blood. This is done by calculating the probabilities

$$W_{RI} = 1/(1+K) \quad \text{(= blood portion from population I), and}$$

$$W_{RII} = 1/(1+1/K) = 1-W_{RI} \quad \text{(= blood portion from population II).}$$

The formulae are related to those of Essen-Möller (1938) and correspond to Bayes' Theorem (1763) with a neutral prior probability.

The greater the distance between the two populations, the more realistic are the statements W_{RI} and W_{RII} for a I-II-mixed-blood person.

To obtain mean lgK values one can use either the findings for typed individuals or calculate theoretical values from phenotype frequencies:

$$\overline{\lg K(I)} = \sum_{i=1}^{n} f_i(I) \cdot \lg[f_i(II)/f_i(I)] ;$$

$$\overline{\lg K(II)} = \sum_{i=1}^{n} f_i(II) \cdot \lg[f_i(II)/f_i(I)] .$$

$f_i(I)$ and $f_i(II)$ $(i=1,2,\ldots,n)$ are the phenotype frequencies of a system in populations I and II respectively. If - as is usual - several systems are included one sums the mean values. -
Given the mean lgK values for both populations one is able to define the "phenotype distance":

$$D = \overline{\lg K(II)} - \overline{\lg K(I)} = \sum_{i=1}^{n} [f_i(II)-f_i(I)] \cdot \lg[f_i(II)/f_i(I)] .$$

Advances in Forensic Haemogenetics 2
Edited by W. R. Mayr

The "phenotype distance" can be more clearly demonstrated by complete lgK curves. One proceeds from a normal distribution and calculates the variance with the following formulae:

$$\text{var lgK(I)} = \sum_{i=1}^{n} f_i(\text{I}) \left\{ \lg \left[f_i(\text{II})/f_i(\text{I}) \right] - \overline{\text{lgK(I)}} \right\}^2;$$

$$\text{var lgK(II)} = \sum_{i=1}^{n} f_i(\text{II}) \left\{ \lg \left[f_i(\text{II})/f_i(\text{I}) \right] - \overline{\text{lgK(II)}} \right\}^2.$$

Figs 1 and 2 compare the distributions of the lgK values of 119 Germans and 51 Italians respectively as well as of 2000 simulated Germans and 2000 simulated Italians with the respective theoretical lgK distribution curves. The agreement between the theoretical curves and, in particular, the simulated values is so close that it appears permissible to use the relatively easily constructed theoretical curves in comparable studies (their mean values and variance are, in any case, reliable).

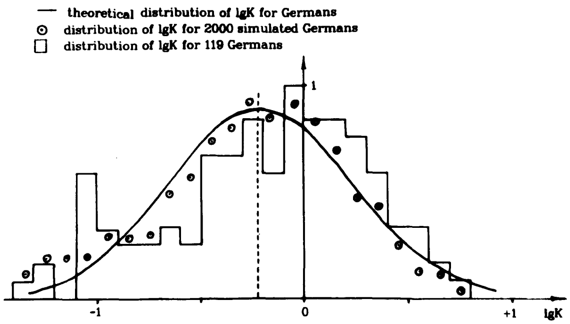

Fig. 1. Theoretical lgK distribution and mean value for Germans (in comparison with Italians), distribution of the lgK values for 2000 simulated Germans and for 119 random Germans (findings in 19 systems).

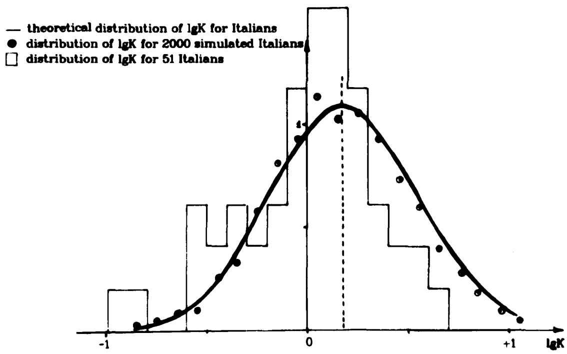

Fig. 2. Theoretical lgK distribution and mean value for Italians (in comparison with Germans), distribution of the lgK values for 2000 simulated Italians and for 51 random persons of Italian origin (findings in 19 systems).

The following list shows the phenotype distances between Germans and other European and non-European populations:

popul.:	Yugo.	Ital.	Span.	Pers.	Turk.	Arab	urban Mex.	USA Black	W-Af. Black
no. of systems	17	19	20	17	22	17	18	19	23
phenot. dist.	0.223	0.380	0.407	0.557	0.658	2.003	2.090	11.300	17.844

The greater the phenotype distance, the less overlap there is between the lgK curves for the two populations, and hence the greater the certainty that an individual belongs to one or other population or, in the case of ethnically mixed blood, the more realistic are the seroanalytically determined blood portions. The smaller the phenotype distance between two populations the closer will W_{RI} and W_{RII} be to the given prior probability of 0.5.

The usefulness of W_{RI} and W_{RII} depends on how realistic their information is on the quantities of the mixed blood portions. To test the degree of reality of W_{RI} and W_{RII} we examined the influence of different gene systems on these values and the effects of the different phenotypes. For our model we used children of ethnically different parents.
As there were not enough cases from practice available for comprehensive studies we calculated theoretical and simulated lgK_m values for the filial generation.
One calculates children's phenotype frequencies m_i as follows: If an ethnically mixed person is homozygous AA, the one gene is inherited from the mother from population I [frequency: $p_1(A)$], and the other from the father from population II [frequency: $p_2(A)$]. The frequency of the genotype AA will then be $p_1(A)\ p_2(A)$. Correspondingly, the frequency of the heterozygous genotype AB will be

$$p_1(A) \cdot p_2(B) + p_2(A) \cdot p_1(B).$$

A phenotype can have one or more genotypes. The sum of their frequencies is the frequency of their phenotype. Accordingly, the mean lgK_m value in a specific gene system with n phenotypes will be

$$\overline{lgK_m} = \sum_{i=1}^{n} m_i \cdot lg\left[(f_i(II)/f_i(I)\right].$$

As the figures in Table 1 (see below) show, the mean lgK value for Black-White children (with findings in 17 systems) lies almost exactly between the mean lgK values for their parents. If one includes the findings in the Duffy system - because of Fy^O this system differentiates effectively between black and white, as shown in the greater phenotype distance between the parents - the mean lgK value for the mixed blood children shifts in the direction of the white parent. If one includes, in addition, the findings in the Gm system - because of the negroid complex gene Gm*1b - this further increases the phenotype distance between Black and White (= "parents") - the mean lgK value for the children shifts markedly towards the black parent.

These results encouraged an attempt to establish some rules in accordance with which the different gene systems influence the position of $\overline{lgK_m}$. The mathematical analysis - which is beyond the scope of this paper - produced the following:

1. If gene systems are codominant, then $\overline{lgK_m}$ will lie halfway between $\overline{lgK(I)}$ and $\overline{lgK(II)}$. -
2. If a system includes recessive-dominant genes, then $\overline{lgK_m}$ will shift towards that population in which the dominant gene is more common. -
3. If a system includes a silent gene, then $\overline{lgK_m}$ will shift towards that population in which this gene is rarer. -
4. If a system includes "complex" genes (e.g. Gm*1,2; Gm*1,b), then $\overline{lgK_m}$ will shift towards that population in which non-complex genes are rarer. -
5. In a system with several non-codominant genes shifts in the same direction are cumulative, shifts in different directions neutralize one another.

The implications for seroanalysis are twofold:

I. Before a person can be assigned to one or other ethnic group one must examine all the available serological information. -

II. Statements about the portions of mixed blood in any specific person will be most realistic if only findings in codominant systems are included. Findings in systems with recessive-dominant and silent genes should be avoided unless these genes are more or less equally rare or common in both populations.

Table 1. Mean lgK values (theoretical) for Germans (G), USABlacks (USB) and their (mixed-blood) children (ch), from the findings in 17, 17+1 and 17+2 gene systems

systems	lgK(G)	lgK(USB)	lgK(ch)	mean dist.of ch. from the Europ.parent	mean dist.of ch. from the Black parent
n=17	-1.666	1.610	0.073	1.739	1.537
" +Fy(a,b,0)	-2.412	3.563	-0.173	2.239	3.736
" +Fy(a,b,0) +Gm(1,2,b)	-7.303	3.997	-0.148	7.155	4.145

References

Bayes Th (1763) Philos Transact 53:376-398, zit Timerding HE (Hrsg) Versuch zur Lösung eines Problems der Wahrscheinlichkeitsrechnung von Thomas Bayes. Engelmann, Berlin 1908

Essen-Möller E (1938) Die Beweiskraft der Ähnlichkeit im Vaterschaftsnachweis. Theoretische Grundlagen. Mitt Anthrop Ges Wien 68:9-53

Hummel K (1980) Biostatistik, HLA-Gutachten, Rassezugehörigkeit, Fremdstämmige. Puerto Rico, Spanien. DAVorm p 195-196

Hummel K (1983) Selection of gene frequency tables. In: Walker RH (ed) Inclusion probabilities in parentage testing. Am Ass Blood Banks.Arlington (Virg) p 231-243

Hummel K (1986) Serotype frequencies in different human populations; racial composition of individuals; statistical evaluation with a frequency mix. In: Brinkmann B, Henningsen K (eds) Advances in forensic haemogenetics. Springer, Berlin Heidelberg New York p 491-496