Genet. Res., Camb. (1989), **53**, *pp.* 207–214 *Printed in Great Britain*

Age differences between distributions of genotypes among pregnant women: evidence of fertility selection

A. GIMELFARB[1]* AND E. BOTTINI[2]

[1] *Department of Ecology and Evolution, The University of Chicago, Chicago, IL 60637, USA*
[2] *Division of Biometry and Human Development, School of Medicine, 2nd University of Rome, Rome, Italy*

(*Received 14 March 1988 and in revised form 21 October 1988*)

Summary

The number of children produced by a modern woman is usually below her total reproductive capacity and is determined by circumstances other than natural selection. It is, therefore, practically impossible to detect differences in natural fertilities associated with different types (e.g. phenotypes, genotypes) of women. This does not mean, however, that natural selection at the reproductive level cannot at all be detected today. If women of a particular type have high natural fertility, this usually means that they reproduce (become pregnant) at a higher rate than women of a type with lower natural fertility. Hence, when there is a limit on the number of children, women of the first type will reach the limit at an earlier age than women of the second type. As a result, types that have a higher natural fertility should be overrepresented among pregnant women of younger ages and, consequently, underrepresented among older ones, as compared to types with a lower natural fertility. Based on this notion, a model of age-related differences between distributions of types among pregnant women is suggested. The model is applied to data on MNSs-blood group and PGM_1 (phosphoglucomutase) types in a sample of pregnant women and an evidence of natural selection at the reproduction level associated with these genetic markers is obtained.

1. Introduction

The notion that natural selection played an important role in the evolution of man hardly needs a proof. However, except for cases of extremely high mortality and extremely low fertility, it may seem practically impossible to find today a direct evidence of natural selection. Indeed, mortality in a modern society is often controlled not as much by natural selection as by advantages in medicine and health care. The number of children produced by a modern woman is usually below her reproductive capacity and it is also not controlled by natural selection. For example, according to data by the Istituto Centrale di Statistica, Italy (Indagine sui Nuclei Familiari, 1982), the average number of children among married Italian women between 45 and 49 years of age with at least one child, i.e. not sterile, is 1·92. It is a safe bet that this number is determined by circumstances other than natural selection on the woman's genotype.

There is a way, however, to detect even relatively weak natural selection at the reproductive level associated with the genotype (or any other characteristic of a woman) even in our days. Consider, for example, women of two types, A and B. Imagine that under conditions when reproduction is not limited by factors other than natural selection, a woman of type A produces more children than a woman of type B. This means, of course, that women of the first type reproduce faster than those of the second type. In reality, however, women have a limit on the number of children that they are going to have. Assuming that the limit is the same for both types, women reproducing faster will reach the limit at younger ages, whereas slower-reproducing women will reach the limit at older ages. Therefore, the relative proportion of type A with respect to type B should be higher among younger pregnant women than among older ones. Thus, differences between distributions of types among pregnant women of different ages may represent an evidence of natural selection at the reproductive level. To make not only verbal but also quantitative arguments, we propose the following model.

2. The model of pregnancy age differences

The purpose of the model is to estimate differences in the natural fertility between women of diverse types

* Corresponding author.

on the basis of the proportions of the types among pregnant women of different ages. The natural fertility is defined as the expected number of children that a woman of a particular type would have produced during her lifetime if she did not have a limit on the number of children. The model is based on the following assumptions:

- There are several types of women. The types can be phenotypes, genotypes or any other distinctive characteristics. The distribution of types in the general population of women, i.e. the probability that a woman is of a particular type i, will be denoted $z(i)$.
- Women have a limit on the number of children that they are going to have. The limiting number need not be the same for all women, but it does not depend on the woman's type. The proportion among women of those who are going to have N children will be denoted $f(N)$. Since the number of children that a woman is going to have is usually below her natural reproductive capacity, it can be assumed that the total number of children by a woman during her lifetime is equal to the limit on the number of children that she was going to have. Therefore, $f(N)$ should be well approximated by the distribution of the number of children among women beyond the age of reproduction.
- The distribution of ages among women in the general population is independent of the woman's type.
- The 'fertility period' during which a woman is capable of producing a child is the same for all women. This period is divided into equidistant 'reproduction intervals' (e.g. 1 year, 2 years) during which a woman can become pregnant, but no more than once. The age of a woman expressed in such intervals will be called her 'reproductive age', t. The minimum reproductive age is always 1, whereas the maximum, t^*, depends on the length of the fertility period and the reproduction interval. If, for example, the fertility period is between 18 and 41 years of age and the reproduction interval is 2 years, the maximum reproductive age is 12 and the reproductive age of, say, a 25-year-old woman is 4. The distribution of reproductive ages, $x(t)$, can be calculated from the actual age distribution among women.
- The probability of becoming pregnant during a reproduction interval, which we shall call the 'reproductive rate', does not depend on the woman's age. It may depend, however, on her type, and the reproductive rate of type i women will be denoted p_i. The natural fertility of a type i woman, i.e. the expected number of children she would have had during her lifetime if there were no limit on the number of children, is computed as $p_i t^*$.
- Abortions as well as multiple births are not considered and mortality among children is neglected. Hence, any pregnancy is assumed to result in one living child and the number of pregnancies that a woman has had is equal to the number of children produced by her.

Based on the above assumptions, we shall now proceed to developing the model. Let $P_{tN}(t)$ be the probability for a woman of reproductive age t to be pregnant, given also that she is of type i and is going to have N children:

$$P_{tN}(t) = Pr\{\text{pregn} \mid t, i, N\}. \tag{1}$$

Under the assumptions of the model, this probability is equal to

$$P_{tN}(t) = p_i \sum_{n=0}^{N-1} C_{t-1}^{n} p_i^n (1-p_i)^{t-1-n}, \tag{2}$$

where C denotes a binomial coefficient. Indeed, a woman who is going to have N children will be pregnant at a reproductive age t, if she has no more than $N-1$ children (pregnancies) by the age $t-1$ (the probability of this is the sum of the binomial probabilities) and becomes pregnant during the reproduction interval t (the probability of this is p_i).

We shall consider only two age groups among pregnant women: 'younger' combining women of ages below a specified age T and 'older' consisting of women whose age is equal or above T. The model can be easily extended to more than two age groups, but it should be kept in mind that the statistical significance of results goes down when the number of groups increases. Let Q_i' and Q_i'' denote the probabilities that a woman of type i is pregnant and that she belongs to the first or the second age group, respectively:

$$Q_i' = Pr\{< T, \text{pregn} \mid i\}, \tag{3a}$$

$$Q_i'' = Pr\{\geqslant T, \text{pregn} \mid i\}. \tag{3b}$$

These probabilities represent the average of $P_{tN}(t)$ over the corresponding reproductive ages as well as over all limiting numbers of children, and the following expressions for them are straightforward:

$$Q_i' = \sum_N (N) \sum_1^{T-1} P_{tN}(t)\, x(t), \tag{4a}$$

$$Q_i'' = \sum_N (N) \sum_T^{t^*} P_{tN}(t)\, x(t). \tag{4b}$$

Let there be I different types among women and assume that data on the distribution of types among pregnant women are available in the form of a $I \times 2$ table with entries in ith row, c_{i1} and c_{i2}, representing the proportions among pregnant women of those whose type is i and who belong to 'younger' or 'older' age group, respectively:

$$c_{i1} = Pr\{i, < T \mid \text{pregn}\}, \tag{5a}$$

$$c_{i2} = Pr\{i, \geqslant T \mid \text{pregn}\}. \tag{5b}$$

According to the probability theory, these proportions can be expressed as

$$c_{i1} = Q'_i z(i)/\sum_i (Q'_i + Q''_i) z(i) \quad (i = 1, 2, \ldots, I), \qquad (6a)$$

$$c_{i2} = Q''_i z(i)/\sum_i (Q'_i + Q''_i) z(i) \quad (i = 1, 2, \ldots, I), \qquad (6b)$$

where $z(i)$ is the distribution of types in the general population of women (pregnant or not). Combined together, equations (2), (4) and (6) represent a system of equations for calculating the reproductive rates of women from a data set containing proportions of different types and ages among pregnant women.

It should be noted that, since $\sum_i (c_{i1} + c_{i2}) = 1$, the number of independent values among c_{i1} and c_{i2} is $2I-1$ and, hence, there are only $2I-1$ equations for determining parameters of the model. The age distribution among women as well as the distribution of the limiting number of children among them are assumed to be independent of their types, and, hence, obtainable from a data source external to the model. Therefore, as far as the model is concerned, $x(t)$ and $f(N)$ are fixed, and the total number of unknown parameters depends on whether the distribution of types in the general population of women can be estimated independently of the model or not. If it can, then $z(i)$ is also fixed and the model has only I unknown parameters: the rates of reproduction, p_i. Since the number of equations in this case is greater than the number of variables, a least square fitting of the system of equations (2), (4) and (6) can be employed to obtain the rates of reproduction.

It happens quite often, however, that a woman's type cannot be detected, unless she is pregnant and finds herself in a hospital where she is given a necessary test. In such a case, the distribution of types in the general population of women cannot be estimated and, hence, $z(i)$ are unknown parameters of the model. Given that $\sum z(i) = 1$, the total number of variables in the system of equations in this case is $2I-1$. Since this is equal to the number of equations, the rates of reproduction together with the distribution of types among women are obtained as a solution of the system of equations (2), (4) and (6). Finding the solution can be facilitated by dividing c_{i1} over c_{i2} in (6):

$$c_{i1}/c_{i2} = Q'_i/Q''_i. \qquad (7)$$

It is seen that (7) does not contain $z(i)$. It is also seen from (4) and (2) that parameters Q'_i and Q''_i for a particular type i are completely determined by the reproductive rate of this type, and, consequently, an equation in (7) for a given i includes only p_i and does not include the reproductive rates of other types. Thus, (7) represents a set of I separate equation in one variable, and the reproductive rate of any particular type can be obtained by solving just one of these equations for the corresponding i. Given that equations in (7) are highly nonlinear, it may not be possible to find an analytical solution, and numerical methods may have to be employed.

After having calculated the reproductive rates, the distribution of types in the general population of women, $z(i)$, is obtained by substituting p_i into equations (6) and solving them. The solution can be represented in the following form:

$$z(i) = a_i/\sum_{j=1}^{I} a_j \quad (i = 1, 2, \ldots I-1),$$
$$z(I) = 1 - \sum_{j=1}^{I-1} z(j), \qquad (8)$$

where

$$a_i = c_{i1}/Q'_i. \qquad (9)$$

Indeed, it follows from (6a) that for any i and j,

$$z(i)/z(j) = a_i/a_j, \qquad (10)$$

and it is not difficult to verify that (8) is the solution of this system of linear equations.

Let us point out in conclusion that the data required to fit the model are relatively easy to obtain. The distribution of ages and of the limiting number of children can be taken from a source which need not be related to the studied types of women (e.g. census). Only data on the distribution of types among pregnant women are required, whereas similar data for the general population are not necessary. Moreover, the model can predict the distribution of types in the general population of women based exclusively on the pregnancy data. This represents an important advantage, since, as has already been mentioned, it is often quite difficult (sometimes even impossible) to administer to women in a general population a test determining their type, especially if it is related to a genetic marker, whereas such tests are almost routinely given to pregnant women.

3. Results for MNS and PGM_1 genotypes

Let us apply the model to MNS and PGM_1 genetic markers in a sample of 212 pregnant women from Rome, Italy. All women in the sample had a normal (without complications) pregnancy, were admitted to a hospital for delivery and gave birth to a healthy child. For the purpose of applying the model, we divided the women in two age groups with 30 years being the cutting age between 'younger' and 'older' ones. The two genetic markers have been a subject of studies by our group for a number of years in connection with complications of pregnancy, such as diabetes and habitual abortions, and the sample of 212 normal pregnancies served as a control in these studies (Bottini *et al.* 1985, 1987; Gloria-Bottini *et al.* 1986).

In order to use the model, it is necessary to know the age distribution and the distribution of the limiting number of children for reproducing women in the general population from which our sample is derived.

Table 1. *Age distribution among 18–41 years old married women*

Age	x	Age	x
18	0·005	30	0·050
19	0·009	31	0·052
20	0·015	32	0·055
21	0·021	33	0·058
22	0·027	34	0·057
23	0·032	35	0·056
24	0·037	36	0·045
25	0·042	37	0·048
26	0·044	38	0·048
27	0·046	39	0·049
28	0·047	40	0·051
29	0·048	41	0·055

Table 2. *Distribution of the number of children among 45–49 years old married women with at least one child*

N	$f(N)$
1	0·417
2	0·348
3	0·136
$\geqslant 4$	0·099

Table 1 shows the distribution of ages among married women between 18 and 41 years of age residing in Italy in 1985. The table is adapted from 'Censimento Generale della Populazione, 25 Ottobre 1981' (1985). The distribution of reproductive ages, $x(t)$, is easily obtained from Table 1 for any given length of reproduction interval. It should be pointed out, however, that this length represents an additional parameter in the model that has not been discussed yet. Since we do not know the actual value of this parameter, we have fitted the model using three different reproduction intervals: 1, 2 and 3 years. The maximum reproductive age, t^*, was, respectively, 24, 12 and 8, and the reproductive age T corresponding to the 'cutting age' of 30 years was 13, 7 and 5, respectively.

Table 2 shows the distribution of the number of children among married Italian women between 45 and 49 years of age having at least one child. The table is adapted from 'Indagine sui Nuclei Familari' (1982). As we have already pointed out, such a distribution should well approximate the distribution of the limit on the number of children, $f(N)$. Only women with at least one child were included in Table 2, since a woman remaining childless by the age of 45 may for all purposes be considered as having zero fertility, whereas the model is applicable only to reproducing women. The distribution of the number of children among women having 4 or more children was not known to us because such women were grouped together in the original data source. For this reason, we considered 4 as the highest limit on the number of children, and the proportion of women having such limit was taken to be as in the last row of Table 2. Unmarried and divorced women as well as widows were not included in either Table 1 or Table 2 because their pattern of reproduction may differ from that of married women.

Table 3. *Distribution of MNS phenotypes among pregnant women of two age groups*

Phenotype	Younger	Older
MMSS	0·096	0·118
MMSs	0·184	0·053
MMss	0·096	0·224
MNSS	0·059	0·066
MNSs	0·154	0·184
MNss	0·154	0·158
NNSS	0·029	0·039
NNSs	0·051	0·039
NNss	0·176	0·118
Number	136	76

Table 3 shows the proportions of MNS blood group phenotypes among 'younger' and 'older' women in our sample. This blood group is controlled by two closely linked loci on chromosome 4, and exhibits 9 distinct phenotypes (Race & Sanger, 1975; Turner, 1969). Large differences are noticeable between the proportions of MMSs and MMss genotypes among younger and older women in the sample (similar differences, although less pronounced, are also present in two other samples of pregnant women: those with habitual abortions and with diabetes). The two first columns in Table 4 show the distributions of only three geneotypic classes; MMSS, MMSs and the 'other' which combines the rest of the genotypes. The χ^2 value for the reduced number of genotypes is only slightly lower than that for the total number (11·74 in Table 4 *vs.* 13·64 in Table 3) indicating that practically all differences in the age of pregnancy attributable to the MNS system are accounted for by only three genotypic classes: MMSs, MMss and other. For this reason, we shall apply the model to only these three classes.

Data in Table 4 can be rearranged into the following 3×2 table:

	< 30	$\geqslant 30$
MMSs	0·118	0·019
MMss	0·061	0·080
Other	0·462	0·259

(11)

where the entries are the proportions among all women in the sample of those having a corresponding genotype and belonging to a corresponding age group, i.e. they represent c_{t1} and c_{t2} in (5). Starting from this table and making use of data in Tables 1 and 2, calculations discussed in the previous section can be

Table 4. *Distribution of MMSs, MMss and 'other' genotypes among pregnant women of two age groups and among girls*

Genotype	Younger	Older	Girls
MMSs	0·184	0·053	0·150
MMss	0·096	0·224	0·123
Other	0·720	0·723	0·727
Number	136	76	187

carried out. Results of such calculations conducted with 1-, 2- and 3-year reproduction intervals are presented in Table 5. The reproductive rates of genotypes are given in the first column of the table. The second column shows the relative values of the reproductive rate of a genotype with respect to that of MMSs genotype. The column denoted $NF(i)$ presents the natural fertilities of corresponding genotypes, i.e. the expected number of children that women would have had during their lifetime if there were no limits on the number of produced children. It is computed as $NF(i) = p_i t^*$. The last column in Table 5 gives the proportions of genotypes predicted by the model for the general population of women.

It is seen that the reproductive rate of a genotype is strongly affected by the length of the reproduction interval. This should not come as a surprise, given that the reproductive rate represents the probability for a woman to be pregnant during a particular time interval and, therefore, it must depend on the length of the interval. Since the length of the reproduction interval for women in our sample is not known, we cannot estimate with a reasonable accuracy the reproductive rates. Recall, however, that our goal is not to estimate these rates, but rather to find an evidence of fertility selection associated with the

Table 5. *Parameters of MMSs, MMss and 'other' genotypes estimated from the model*

Genotype (i)	p_i	p_i/p_1	$NF(i)$	$z(i)$
Reproduction interval = 1 year				
(1) MMSs	0·331	—	7·946	0·157
(2) MMss	0·050	0·152	1·205	0·198
(3) Other	0·184	0·554	4·405	0·645
Reproduction interval = 2 years				
(1) MMSs	0·587	—	7·048	0·154
(2) MMss	0·100	0·170	1·195	0·199
(3) Other	0·349	0·594	4·188	0·647
Reproduction interval = 3 years				
(1) MMSs	0·786	—	6·292	0·150
(2) MMss	0·148	0·188	1·185	0·201
(3) Other	0·499	0·634	3·989	0·648

p_i, reproductive rate; $NF(i)$, natural fertility; $z(i)$, distribution of types.

Table 6. *Distribution of PGM_1 genotypes among pregnant women of two age groups and in the general population*

Genotype	Pregnant		General population
	Younger	Older	
1	0·461	0·595	0·513
1–2	0·432	0·391	0·399
2	0·107	0·014	0·087
Number	130	74	388

genotypes. For this, we only need to know the relative natural fertilities. Since the maximum reproductive age is assumed to be the same for women of all types, the relative values of natural fertilities are the same as those of the reproductive rates and are shown in the second column of Table 5. Even though these values are also affected by the length of the reproduction interval, the effect is much smaller than on p_i.

Pregnant women in our sample were also tested for phosphoglucomutase – locus 1 (PGM_1) genotypes. PGM_1 is one of the four loci that are known to determine distinct sets of phosphoglucomutase isozymes (Hopkinson & Harris, 1968; Spencer *et al.* 1964). It has two alleles, PGM_1^1, PGM_1^2, and exhibits three genotypes denoted as 1 for the first homozygote, 1–2 for the heterozygote and 2 for the second homozygote. The proportions of PGM_1 genotypes in the two age groups in our sample are shown in the first two columns of Table 6 (the number of women tested for PGM_1 was slightly below the number of those tested for MNS: 204 *vs.* 212).

The difference in the proportions between the two age groups is statistically significant ($P < 0{\cdot}03$). The table of elements c_{i1} and c_{i2} is obtained from Table 6 as:

	< 30	$\geqslant 30$
1	0·294	0·216
1–2	0·275	0·142
2	0·068	0·005

(12)

Since women tested for PGM_1 are the same individuals who were tested for MNS, the age distribution as well as the distribution of the limiting number of children are the same as those used in the calculations for MNS and shown in Tables 1 and 2. The results produced by the model for the PGM_1 data are presented in Table 7, the composition of which is the same as of Table 5. The second column represents the relative values of reproductive rates (natural fertilities) with respect to that of the first homozygote.

Even though we do not know the actual length of the reproduction interval for women in our sample, a good guess is that it is somewhere between 1 and 3 years, and, therefore, the relative values of the reproductive rates for the MNS and PGM_1 genotypes are between those corresponding to the shortest and

Table 7. *Parameters of PGM_1 genotypes estimated from the model*

Genotype (i)	p_i	p_i/p_1	$NF(i)$	$z(i)$
Reproduction interval = 1 year				
(1) 1	0·145	—	3·486	0·489
(2) 1–2	0·195	1·340	4·670	0·406
(3) 2	0·406	2·797	9·751	0·105
Réproduction interval = 2 years				
(1) 1	0·280	—	3·362	0·492
(2) 1–2	0·368	1·315	4·421	0·408
(3) 2	0·690	2·462	8·279	0·100
Reproduction interval = 3 years				
(1) 1	0·406	—	3·245	0·496
(2) 1–2	0·524	1·292	4·194	0·409
(3) 2	0·885	2·183	7·082	0·095

p_i, reproductive rate; $NF(i)$, natural fertility; $z(i)$, distribution of types.

the longest reproduction interval in the second column of Table 5 for MNS and Table 7 for PGM_1

4. Testing the validity of the model

There is no way, of course, to test directly predictions of the model concerning the reproductive rates or natural fertilities, since they will always remain unknown to us. There are, however, other ways to evaluate the validity of the model.

Notice that distribution $z(i)$ should not be affected by the length of the reproduction interval. Indeed, this is a distribution in the general population of women (pregnant or not), and, as such, it is not supposed to depend on how frequently women become pregnant. Thus, if the model is valid, it should predict similar distributions $z(i)$ for any length of the reproduction interval. It is seen from the last column in Tables 5 and 7 that this is, indeed, the case. The predicted distributions of genotypes corresponding to different reproduction intervals are very similar.

Another test of the validity of the model comes from a comparison between the predicted distribution $z(i)$ and the actual proportions of genotypes among women in the general population. Unfortunately, we do not have data on the proportions of MNS genotypes in the general population of women from which our sample is derived. We do have, however, data on the corresponding proportions in a sample of 187 girls of ages from newborn to 12 years old. The majority of girls in the sample were clinically normal and came to the attention of doctors because of diabetes or ischemia in one of their parents. The rest of the sample consisted of 24 girls with neonatal jaundice not related to ABO or Rh incompatibility, 52 with bronchial asthma and 11 with obesity. Neither of the listed conditions is known to be associated with MNSs blood group, and, therefore, the distribution of genotypes among the girls should not be very different from that among women in the general population. Comparing the last columns in Tables 4 and 5, it is seen that for any length of the reproduction interval the proportions of genotypes predicted by the model are not far from those among the girls. As for the predicted distribution $z(i)$ of PGM_1 genotypes, we can compare it with the proportions reported by Modiano *et al.* (1970) for a sample of 388 women from a population similar to the one from which our sample was derived that are shown in the last column of Table 7. The differences between the proportions predicted by the model and reported by Modiano *et al.* are quite small. It should be pointed out that there were only very few women with genotype 2 of PGM_1 in our sample. As a matter of fact, there was only one woman with such genotype among those of age $\geqslant 30$, meaning that the estimate $c_{32} = 0{\cdot}005$ in (12) is based on just one individual. It is remarkable that, in spite of that, the model predicts so well the distribution of PGM_1 genotypes in the general population of women.

Another test of the model is based on the following consideration. Since the same women in our sample were tested for both MNS and PGM_1 genotypes, any parameter of the general population of women not related to a particular genetic system should be the same whether it is predicted from MNS or PGM_1 data. One such parameter is the probability for a woman in the general population to be pregnant. It is given by the expression in the denominator in (6):

$$\sum_i (Q_i' + Q_i'')\, z(i). \tag{13}$$

The last two columns in Table 8 show the probability of pregnancy predicted by the model based on MNS and PGM_1 data. The correspondence between the two predictions is very good. It is useful to notice that the probability of pregnancy can, in principle, be directly estimated by the proportion of pregnancies among women of reproductive ages in the general population. Such an estimate can then be used to evaluate the actual length of the reproduction interval by comparing the probabilities of pregnancy predicted for different reproduction intervals with those estimated directly. The length of the interval for which the two probabilities are close to each other would

Table 8. *The probability of pregnancy in the general population of women estimated from MNS and PGM_1 data*

Reproductive interval	MNS	PGM_1
1 year	0·052	0·056
2 years	0·106	0·114
3 years	0·161	0·174

indicate the actual length of the reproduction interval. Unfortunately, we do not have the necessary data, and, therefore, the length of the reproduction interval for women in our sample remains unknown.

It should be pointed out in conclusion that we have presented in this paper only a deterministic model whose statistical properties have not been worked out, yet. Therefore, the validity tests discussed above are not statistical, they only verify that the deterministic model does not yield erroneous predictions.

5. Discussion and conclusions

The model presented in this paper provides means of estimating natural selection at the reproductive level in the case when the actual fertility of a woman, i.e. the number of children that she has during her lifetime, is below her total reproductive capacity due to causes other than natural selection. It stems from an observation that differential reproductive capacities (natural fertilities) of different types of women should be manifested in such case by differences between distributions of the types among pregnant women of different ages.

A possibility exists that the observed differences between genotypic distributions in two age groups of pregnant women are due to different mortality rates among genotypes, and, hence, may have nothing to do with pregnancy itself. It can be argued, however, that mortality of at least some of the genotypes must be unrealistically high in order to account for the differences between their proportions in the age groups. For example, if u_i and u_j denote the proportions of genotypes i and j in the younger group, whereas v_i and v_j denote the corresponding proportions in the older group, the following relations connect these variables:

$$v_i = u_i(1-m_i)/(1-M), \tag{14a}$$

$$v_j = u_j(1-m_j)/(1-M), \tag{14b}$$

where m_i and m_j are mortalities of the corresponding genotypes and M is the average mortality among women. Dividing (14a) by (14b) yields after some algebra

$$m_i = 1-(v_i u_j/v_j u_i)(1-m_j). \tag{15}$$

Let genotype i be MMSs and genotype j be MMss. Assuming that there is no mortality among MMss women, i.e. $m_j = 0$, and substituting the actual genotypic proportions from Table 4 into the right side of (15) results in $m_i = 0{\cdot}88$. Thus, in order to account for the observed difference in the genotypic proportions among younger and older women, the mortality among MMSs women must be at least as high as 88%. This, of course, is not very realistic, and differential mortality can be ruled out as an explanation. The model discussed in this paper provides another explanation.

As any model, it is based on a number of assumptions, some of which may be closer to reality than others. For example, the assumption that all pregnancies result in a living child is not very unrealistic, given the relatively low child mortality in developed countries. On the other hand, to assume that the reproductive rate of a woman, i.e. the probability for her to become pregnant during a reproduction interval, does not depend on the woman's age may seem much less realistic. Notice, however, that, even if in reality the reproductive rate changes with age, it would be quite difficult to obtain an adequate information about such changes. For example, many existing data show that the probability for a woman to be pregnant declines with age. This, however, does not represent an adequate description of changes in the reproductive rate. Indeed, the probability of pregnancy at a particular age is determined not only by the woman's reproductive rate at this age but also by the limit on the number of children that she is going to have. Since the probability to reach this limit increases with the age of a woman, the probability that she becomes pregnant will decline, even if her reproductive rate remains unchanged. Therefore, it seems practically impossible to obtain data on 'natural' age changes in the reproductive rate that are not confounded by the changes due to a limit on the number of children.

In spite of all the simplifying (and, perhaps, oversimplifying) assumptions, the validity of the model in instances when it could be tested has been confirmed, as the previous section demonstrates. Therefore, we may state with a sufficient confidence that the observed differences between the proportions of genotypes among pregnant women in two age groups reflect differential natural fertilities of the genotypes. In MNS genetic system, MMSs genotype has the highest natural fertility, whereas that of MMss genotype is the lowest. As for PGM_1 system, the natural fertility of the second homozygote is the highest and that of the first homozygote is the lowest.

Notwithstanding the demonstrated differences in the natural fertilities of MNS and PGM_1 genotypes, there is, obviously, no fertility selection on these genotypes in a population in which the number of children by a woman is limited by factors other than her natural fertility. This does not mean, however, that no selection at the reproductive level can be associated with MNS or PGM_1 systems in such a population. Genotypes with a higher reproductive rate will have a shorter generation length and, hence, will have a selective advantage over genotypes with a lower reproductive rate. Therefore, even though selection through differential fertility may not operate in the population from which our sample of women is derived, there still is selection associated with MNS and PGM_1 systems at the reproductive level through the differential generation length.

It should be pointed out in conclusion that the fact

of natural selection associated with MNS and PGM_1 genotypes does not necessarily mean that selection acts directly on either of these genetic systems. It is quite possible that the differences in the reproductive rates represent correlated responses to natural selection on other characterisics of a woman.

We wish to thank Professors L. Carenza and G. Maggioni for their support and Professor I. Barrai for useful discussions. The criticism and comments by anonymous referees were very helpful, and we are especially grateful to one of them for catching an important error. This work was supported by grants from the 1st University of Rome (60%) and from Italian CNR (No. 86 02040 11).

References

Bottini, E., Bonci, E., Gloria-Bottini, F., Nicotra, M., Carapella, E., Coghi, N., Lucarini, M., Martini, P. A., Paris, D., Bonci, M. & Finocchi, G. (1987). MNSs genetic polymorphism and intrauterine selection through habitual abortion. *Human Biology* **59**, 103–210.

Bottini, E., Gerlini, G., Fallucca, F., Lucarini, N., Bonci, E., Arachi, S. & Gloria-Bottini, F. (1985). Diabetic pregnancy: fetal development and maternal phosphoglucomutase phenotype. *Disease Markers* **3**, 83–89.

Dati sulle caratteristiche strutturali della popolazione e delle abitazioni. (1985). In Censimento Generale della Popolazione, 25 ottobre 1981. Istituto Centrale di Statistica, Roma.

Gloria-Bottini, F., Lucarini, N., Bonci, E., Pascasio, F., Bonci, M., Assumma, M. & Scalamandre, A. (1986). Genetic polymorphism and intrauterine development: the role of maternal PGM_1 and MNSs genotypes. *American Journal of Diseases of Children* **49**, 324–327.

Hopkinson, D. A. & Harris, H. (1968). A third phosphoglucomutase locus in man. *Annals of Human Genetics* **31**, 359–367.

Indagine sui nuclei familiari (1982). In *Collana d'Informazioni.* Istituto Centrale di Statistica, Roma.

Modiano, G., Scozzari, R., Gigliani, F., Santolamazza, C., Afeltra, P. & Frattoli, W. (1970). Red cell phosphoglucomutase polymorphism. I. Enzyme activity of different cell PGM phenotypes. *Human Heredity* **20**, 86–93.

Race, R. R. & Sanger, R. (1975). The MNSs blood groups. In *Blood Groups in Man*, 6th edn, pp. 92–138. Oxford: Blackwell.

Spencer, N., Hopkinson, D. A. & Harris, H. (1964). Phosphoglucomutase polymorphism in man. *Nature* **21**, 742–745.

Turner, J. G. R. (1969). Epistatic selection in the rhesus and MNSs blood groups. *Annals of Human Genetics* **33**, 197–206.