

**Estimating genetic and non-genetic components of variance  
for fasting glucose levels in pedigrees ascertained through non-insulin  
dependent diabetes**

BY T. H. BEATY

*Department of Epidemiology, Johns Hopkins University, School of Hygiene and Public Health,  
Baltimore, MD 21205*

AND S. S. FAJANS

*Department of Internal Medicine, Division of Endocrinology and Metabolism,  
The University of Michigan, Ann Arbor, MI 48109*

SUMMARY

Fasting glucose levels measured on 337 individuals in 14 pedigrees ascertained through a proband with non-insulin dependent diabetes were used to estimate genetic and non-genetic components of variance under a multifactorial model of inheritance. In this sample genetic factors were important in controlling variation in basal carbohydrate metabolism, as represented by age-adjusted log-fasting glucose. There was no evidence that arbitrary sib common environments or arbitrary parent common environments accounted for significant portions of the variability in fasting glucose in these data. An arbitrary environment shared by parent and offspring, however, had a marginally significant impact on the likelihood.

Parameter estimates obtained from multifactorial models analysed in this manner are sensitive to extreme phenotypic values, however, and caution must be exercised in estimating total genetic variation. While additive genetic factors did account for a significant proportion of the total variation in fasting glucose, a large proportion remained unexplained.

INTRODUCTION

Diagnosis of the disease 'diabetes' represents an arbitrary classification of continuously varying blood or plasma glucose levels, either fasting levels or in response to a glucose load, into discrete diagnostic categories. Such arbitrary grouping of individuals, while clinically necessary, can obscure patterns of familial aggregation which may shed light on etiological mechanisms. In this study, we will focus on fasting glucose as a representation of basal carbohydrate metabolism and examine a series of multifactorial models of inheritance to explain the distribution of fasting glucose in relatives. Since relatives share both genes and familial environments, these multifactorial models will assess the importance of both genetic and non-genetic factors. Fasting glucose levels on 337 members of 14 large pedigrees ascertained through a proband with non-insulin dependent diabetes mellitus (NIDDM) will be used to test hypotheses about the role of genetic and non-genetic factors controlling basal carbohydrate metabolism and contributing to observed correlations among family members.

Since this sample of pedigrees was ascertained through NIDDM, the inferences from this

analysis should not be applied to general populations. Rather this analysis is aimed at assessing the role of possible factors, both genetic and non-genetic, which could account for familial resemblance in basal carbohydrate metabolism in that portion of the population which could be ascertained through NIDDM. Given the high prevalence of NIDDM in many populations (Palumbo *et al.* 1976; West, 1978), this restriction in possible inferences from this sampling design is not as limiting as it might appear.

Although non-insulin dependent diabetes frequently shows strong familial aggregation, environmental factors such as diet and obesity also influence carbohydrate metabolism and risk to this disease. This combined influence of both genetic and environmental factors makes it difficult to discriminate among possible etiological mechanisms with any degree of confidence. While some forms of NIDDM do appear to follow simple autosomal dominant inheritance (Fajans, Cloutier & Crowther, 1978), the majority of NIDDM can only be described as 'familial' without being clearly determined by a single genetic or environmental mechanism. By examining fasting glucose levels directly it may be possible to gain insight into the role of genetic and non-genetic factors affecting carbohydrate metabolism in individuals at high risk for diabetes.

#### METHODS

*The sample.* The pedigrees used in this study were ascertained through a proband with non-insulin dependent diabetes and represent the largest kindreds among a series of relatives of diabetics studied over a period of more than 25 years described elsewhere (Beaty, Neel & Fajans, 1982). Diabetic probands were patients treated at the University of Michigan Hospital and were largely from southeastern Michigan, although some relatives from other states were tested. The extent of co-operation and participation among family members may well have been associated with the prevalence and severity of diabetes in the kindred. As mentioned above, such effects of sampling must be considered when making inferences about the genetic and non-genetic factors determining inheritance of fasting glucose levels. Estimated values for parameters of multifactorial models examined here should not be considered applicable to a general population.

The sample consists of 337 non-proband members of 14 pedigrees, 147 males and 190 females, ranging in age from 4 to 68. Each pedigree has at least ten members with fasting glucose available; the largest had fasting glucose samples on 91 individuals. Table 1 shows the number of tested individuals in each pedigree and their age ranges. The fasting glucose level was determined from a blood sample drawn after an overnight fast. Each individual had been instructed to follow a diet containing 250 gm of carbohydrate per day for three days prior to testing. All individuals tested were ambulatory, free of known disease, and not under medication at the time of testing. In this sample, males had a mean fasting glucose level of 101.18 mg/dl ( $\pm 1.396$  s.e.) and females had a mean of 102.62 mg/dl ( $\pm 2.368$  s.e.).

*The models examined.* A series of multifactorial models of inheritance was examined using data on fasting glucose in these 14 pedigrees. These models include both classical genetic factors and arbitrary family environmental factors which could account for similarities in fasting glucose among relatives. The non-genetic factors include common environments shared by full sibs, those shared by parents, etc. Although these parameters are usually estimated by least-squares

Table 1. 14 pedigrees ascertained through non-insulin dependent diabetes mellitus

Kindred number	Males	Females	Total	Mean age	Range in age	Mean fasting glucose	S.E. fasting glucose
3414	9	10	19	30.05	(11-62)	106.11	(± 1.91)
4084	7	8	15	27.53	(15-42)	95.60	(± 1.72)
4151	13	16	29	22.28	(4-68)	100.93	(± 4.95)
4299	20	35	55	29.44	(13-62)	97.81	(± 1.05)
4658	6	10	16	34.81	(15-54)	100.19	(± 2.18)
4883	9	4	13	39.54	(9-58)	99.38	(± 3.23)
4884	9	7	16	25.38	(4-59)	100.88	(± 2.76)
5577	40	51	91	23.52	(6-58)	106.81	(± 4.54)
7030	7	7	14	34.93	(11-64)	99.29	(± 2.93)
7163	4	15	19	25.00	(6-62)	97.37	(± 1.92)
7353	5	6	11	24.36	(7-40)	98.82	(± 2.27)
7364	6	7	13	25.46	(7-45)	115.46	(± 13.24)
15040	8	5	13	23.15	(7-45)	105.00	(± 5.77)
15750	4	9	13	19.46	(8-46)	91.77	(± 1.59)
Total	147	190	337	26.66	(4-68)	101.99	(± 1.47)

techniques using pairs or sets of relatives (e.g. pairs of sibs, parent-offspring trios, etc.), Lange, Westlake & Spence (1976) have shown that complete pedigrees can be used to partition the covariance in fasting glucose between any two relatives,  $i$  and  $j$ , as

$$\text{Cov}(G_i, G_j) = 2\phi_{ij}\sigma_A^2 + \Delta_{ij}\sigma_a^2 + r_{ij}\sigma_{CE}^2 + \delta_{ij}\sigma_R^2.$$

Here  $\phi_{ij}$  is the kinship coefficient describing the probability that relatives  $i$  and  $j$  share one allele at an autosomal locus identical by descent;  $\Delta_{ij}$  describes the probability that relatives  $i$  and  $j$  share both alleles at an autosomal locus which are identical by descent (Jacquard, 1974);  $r_{ij}$  is simply an index of an arbitrary common environment which  $i$  and  $j$  may share; and  $\delta_{ij} = 1$  if  $i = j$ , and  $= 0$  otherwise. The pedigree structure itself and the nature of the common environments being considered dictate these coefficients.

The parameters of the multifactorial model to be estimated are the components of variance representing the amount of variation in fasting glucose which can be attributed to variation in additive genetic alleles ( $\sigma_A^2$ ), to non-linear interactions between alleles, i.e. the dominance variance ( $\sigma_a^2$ ), to differences in common environments shared by various types of relatives ( $\sigma_{CE}^2$ ), and a residual variation which is not explained by any of the other factors in the model ( $\sigma_R^2$ ). The genetic parameters, additive genetic variance and dominance variance, are those of Fisher's classic polygenic model, although these genetic variance components may also reflect the action of alleles at a single locus. Making inferences about these parameters in any population, e.g. a population of individuals in pedigrees with NIDDM, requires assumptions of Hardy-Weinberg equilibrium, no linkage disequilibrium among loci, no assortative mating, and no gene-environment interactions.

The method for partitioning variation in a quantitative trait into genetic and non-genetic components, described by Lange *et al.* (1976) and later expanded by Moll, Powsner & Sing (1979), assumes the trait in question, here fasting glucose, follows a multivariate normal distribution within a given pedigree. To reduce the skewness and kurtosis observed in the raw fasting glucose levels, the natural logarithm of fasting glucose was used. Prior to examining different

multifactorial models of inheritance, the impact of covariates such as age and obesity (as represented by the Body Mass Index which is weight/height<sup>2</sup> in kilograms and metres, respectively) on log-fasting glucose was examined. In this sample, there was no significant effect of Body Mass Index on fasting glucose in a polynomial regression. Age, however, did significantly influence fasting glucose although it accounted for a modest proportion of the total variation: 3.43% and 3.08% of the variation in log-fasting glucose could be accounted for by linear effects of age among males and females, respectively. There was no evidence of any significant non-linear, i.e. quadratic or cubic, effects of age on log-fasting glucose.

By using the maximum likelihood technique described by Lange *et al.* (1976), a number of different variance components, both genetic and non-genetic in nature, may be estimated simultaneously on a set of pedigrees. Since this method is based on Fisher's scoring algorithm, the asymptotic covariance matrix for these estimators is also obtained. This not only provides approximate standard errors for each parameter estimate, but reveals which estimated components of variance are highly correlated with one another. For certain pedigree configurations, some parameters will be inextricably confounded with one another, making interpretation of the resulting parameter estimates difficult.

In addition, this technique provides test statistics for hypotheses about specific components of variance. Under the likelihood ratio criterion, an approximate chi-square statistic may be obtained to test the hypothesis that a specific component of variance is not significantly different from zero.

#### RESULTS

The age-adjusted log-fasting glucose values on all 337 members of 14 pedigrees ascertained through NIDDM were used to examine a series of different multifactorial models. This series included models with only genetic factors, only non-genetic factors (family environmental factors), and combinations of both genetic and non-genetic factors to account for the similarities in basal carbohydrate metabolism among relatives. The models, their log-likelihoods, and their estimated parameter values are listed in Table 2. From this Table, it is possible to obtain a number of approximate chi-square statistics for tests of hypotheses about the various genetic and non-genetic components of variance.

Several important patterns emerge from Table 2. The genetic components, additive genetic variance and dominance variance, do significantly contribute to correlations in log-fasting glucose among relatives. Comparing the genetic model with both additive and dominance variance to the marginal model gives an approximate chi-square statistic of 70.8 with 2 d.f. ( $P < 0.001$ ). Most of this improvement in the log-likelihood is attributable to the inclusion of additive genetic variation in the model, although there appears to be non-linear interaction among alleles. Including dominance variance does significantly improve the log-likelihood ( $\chi^2 = 6.14$ ). There was, however, substantial correlation ( $-0.83$ ) between the estimated values for these genetic parameters in these pedigrees. The interpretation of this role for dominance variance will be discussed further below.

When genetic components are in the model, there is no significant improvement in log-likelihood gained by considering either a common sibship environment shared by full sibs less than 20 years of age ( $\chi^2 = 0.002$ ) or a common environment shared by parents ( $\chi^2 = 0.05$ ).

Table 2. Estimated components of variance and log-likelihood of multifactorial models fitted to age-adjusted log-fasting glucose data from 14 pedigrees ascertained through NIDDM

Model	Number of parameters	Log likelihood	Parameter estimates					
			$\hat{\sigma}_A^2$	$\hat{\sigma}_D^2$	$\hat{\sigma}_{SibCE}^2$	$\hat{\sigma}_{ParentCE}^2$	$\hat{\sigma}_{Parent-offspringCE}^2$	$\hat{\sigma}_R^2$
Marginal model	1	-338.354	—	—	—	—	—	2.740
Genetic models:								
Additive only	2	-306.314	4.172	—	—	—	—	0.836
Additive and dominance	3	-302.946	0.603	2.714	—	—	—	0.799
Non-genetic models:								
Parent CE	2	-319.493	—	—	—	2.347	—	1.283
Juvenile sibship CE	2	-322.348	—	—	1.914	—	—	1.495
Parent-offspring CE	2	-325.354	—	—	—	—	1.355	1.915
Parent CE, Sib CE and parent-offspring CE	4	-302.897*	—	—	0.919	0.991	1.333	0.799
Non-genetic and genetic models:								
Additive and parent CE	3	-304.342	2.497	—	—	1.086	—	0.817
Additive, dominance and parent CE	4	-302.919	0.662	2.418	—	0.221	—	0.799
Additive and Sib CE	3	-305.819	3.386	—	0.5107	—	—	0.834
Additive, dominance and Sib CE	4	-302.945	0.599	2.686	0.030	—	—	0.799
Additive, parent-offspring CE	3	-304.360	3.179	—	—	—	0.843	0.840
Additive, dominance and parent-offspring CE	4	-301.933	0.372	2.178	—	—	0.709	0.795
Complete model	6	-302.006*	0.2517	1.538	0.0765	0.406	0.929	0.797

\* No convergence.

Table 3. *Estimated components of variance of a multifactorial model evaluated on the entire sample of 337 individuals and on a sample of 329 individuals, excluding 8 individuals with extreme phenotypic values*

	Number of pedigrees	Number of individuals	Parameter estimates ( $\pm$ S.E.)		
			$\hat{\sigma}_A^2$	$\hat{\sigma}_d^2$	$\hat{\sigma}_R^2$
Entire sample	14	337	0.603 ( $\pm 0.731$ )	2.714 ( $\pm 0.724$ )	0.799 ( $\pm 0.104$ )
Excluding eight individuals	14	329	0.228 ( $\pm 0.158$ )	0.051 ( $\pm 0.178$ )	0.619 ( $\pm 0.079$ )

When the multifactorial model did *not* include genetic components, there was a significant improvement in log-likelihood due to these common environments compared with the marginal model. This illustrates how variance due to genetic factors can be absorbed by familial environments when genetic factors are ignored. Overall, however, there is little evidence to suggest any role for either common sibship or parent environments in determining the distribution of log-fasting glucose once genetic factors are considered.

The only familial environmental component to show any evidence of influencing familial aggregation of log-fasting glucose was parent-offspring common environment. The significance of parent-offspring common environment in models considering genetic factors was marginal ( $\chi^2 = 2.03$ ). Nonetheless, this latter model gave the best overall fit to these data on age-adjusted log-fasting glucose. In this model, 9.2% of the total variation in age-adjusted log-fasting glucose is attributed to variation in additive genetic factors (both additive and dominance variance account for 62.9% of the total variation); 17.5% of the total variation is attributed to an arbitrary common environment shared by parent and juvenile offspring; and 19.6% is not explained by either of these factors.

In this model, all four components of variance are estimated with substantial error, however. For example,  $\hat{\sigma}_A^2 = 0.372 \pm 0.744$ ,  $\hat{\sigma}_d^2 = 2.178 \pm 0.646$ ,  $\hat{\sigma}_{POCE}^2 = 0.709 \pm 0.432$ , and  $\hat{\sigma}_R^2 = 0.795 \pm 0.327$ . The small estimate of additive genetic variance compared with variation due to interaction between alleles (dominance variance) reflects the high negative correlation between these two genetic parameters ( $-0.68$ ) in this model.

Estimates of variation due to non-linear interactions, or dominance variance, can also be influenced by outlying phenotypic values in a few individuals. This type of extreme value is common in many physiological measures, where clinically identifiable individuals have phenotypic values well removed from the sample mean. While all analyses discussed up to this point have dealt with *all* non-proband members of these pedigrees, regardless of their diagnostic status, eight individuals in four pedigrees had fasting glucose levels more than two standard deviations above the mean, i.e. a fasting glucose level of 160 mg/dl or greater. A multifactorial model including additive and dominance genetic variation was examined on the sample excluding these eight individuals. Table 3 compares the estimated variance components for the entire sample and for the sample excluding these eight individuals. The major difference in parameters estimated on these two samples lies in the magnitude of the estimated variation due to non-linear interaction between alleles (dominance variance). Including data on individuals with extreme phenotypic values leads to a grossly inflated estimate of dominance variance at the expense of

all other parameters in the model. This apparent lack of robustness to extreme phenotypic values requires cautious interpretation of 'total genetic variation' (i.e. the sum of additive and dominance variation). Variation due to additive genetic alleles does not appear to be as sensitive, accounting for 14.7% of the total variation in the total sample of 337 individuals and 25.4% of the total variation in the subset of 329 individuals. This parameter is still estimated with substantial error, however. It should be noted that, among the 329 individuals in 14 pedigrees, 68.9% of the total variation in age-adjusted log-fasting glucose could not be accounted for by genetic factors.

#### DISCUSSION

The multifactorial models of inheritance presented here have been used to quantify the role of genetic factors controlling the distribution of fasting glucose levels in 337 members of 14 pedigrees ascertained through non-insulin dependent diabetes mellitus (NIDDM). The analytical techniques used here to estimate a series of genetic and non-genetic components of variance are those originally described by Lange *et al.* (1976) and later expanded by Moll *et al.* (1979). The multifactorial models listed in Table 2 show that the genetic factors, additive genetic variance and dominance genetic variance, significantly influence the distribution of fasting glucose among relatives in these pedigrees.

Three non-genetic familial environments were examined: an environment common to full sibs under the age of 20, an environment common to parents, and an environment common to parents and young offspring (those under the age of 20). These common environments do not represent any specific factors influencing basal carbohydrate metabolism, e.g. diet, exercise habits, etc., rather they are arbitrary non-genetic factors which could mimic the correlations among relatives expected under genetic models. When all three common environmental factors were examined simultaneously, the log-likelihood was fairly high, illustrating how non-genetic common environments may mimic genetic factors when such exist. Once genetic factors were considered, however, the common environments shared by parents and those shared by full sibs did not significantly improve the multifactorial model.

The only common environment to suggest any degree of influence on the distribution of age-adjusted log-fasting glucose was parent-offspring common environment, although its impact was of marginal significance. Under this best-fitting model, the expected correlation between parent and offspring was 0.221, which was quite close to the observed correlation of 0.240. Similarly the expected correlation between full sibs was 0.180, which was also close to the observed correlation of 0.130 for these pedigrees.

*Sensitivity to extreme phenotypes.* The estimators for different components of variance can be quite sensitive to extreme phenotypic values, which are not uncommon in physiological traits such as fasting glucose. While it is statistically reasonable to simply discard such outliers, they may well be of great genetic and/or clinical interest.

The impact of such extreme phenotypes can be seen in comparing a multifactorial model analysed on the total sample of 337 individuals with that obtained from the sample of 329 individuals (eight individuals with fasting glucose levels of 160 mg/dl or greater were excluded). This comparison showed that the variation due to non-linear interactions between alleles (dominance variance) is quite sensitive to extreme phenotypes and can be grossly inflated when

such values are present. A conservative approach to interpreting multifactorial models such as those presented here would be to rely more on the estimated additive genetic variance as an indicator of the true role of genetic factors controlling a trait than on the combination of both additive and dominance genetic variance. While additive genetic factors do play a significant role in determining fasting glucose levels among individuals in NIDDM pedigrees, a majority of variation in fasting glucose remains as residual variation.

We wish to thank Drs P. P. Moll, G. A. Chase, and J. V. Neel for their helpful comments during the course of this project. We would also like to thank R. N. Diffenderfer for his excellent programming. This work was supported in part by US Public Health Service grants AM00888 and NIH 5-SO7-RR05445 Biomedical Research Support Program.

#### REFERENCES

- BEATY, T. H., NEEL, J. V. & FAJANS, S. S. (1982). Identifying risk factors for diabetes in first degree relatives of non-insulin dependent diabetic patients. *Am. J. Epidemiol.* **115**, 380-397.
- FAJANS, S. S., CLOUTIER, M. C. & CROWTHER, R. L. (1978). Clinical and etiological heterogeneity of idiopathic diabetes mellitus. *Diabetes* **27**, 1112-1125.
- JACQUARD, A. (1974). *The genetic structure of populations*, pp. 102-112. New York: Springer-Verlag.
- LANGE, K., WESTLAKE, J. & SPENCE, M. A. (1976). Extensions to pedigree analysis. III. Variance components by the scoring method. *Ann. Hum. Genet.* **39**, 485-491.
- MOLL, P. P., POWSNER, R. & SING, C. F. (1979). Analysis of genetic and environmental sources of variation in serum cholesterol in Tecumseh, Michigan. V. Variance components estimated from pedigrees. *Ann. Hum. Genet.* **42**, 343-354.
- PALUMBO, P. J., ELVEBACK, L. R., CHU, C. P., CONNALLY, D. C. & KURLAND, L. T. (1976). Diabetes mellitus: incidence, prevalence, survivorship, and causes of death in Rochester Minnesota 1945-1970. *Diabetes* **25**, 566-573.
- WEST, K. M. (1978). *Epidemiology of diabetes and its vascular lesions*, pp. 127-158. New York: Elsevier.