ANNALS OF HUMAN BIOLOGY, 1981, VOL. 8, NO. 6, 529–541

# A twin study of skin reflectance

P. CLARK, A. E. STARK and R. J. WALSH

School of Human Genetics, University of New South Wales, Sydney

and R. JARDINE and N. G. MARTIN

Department of Population Biology, Research School of Biological Sciences, Australian National University, Canberra

*Received 22 October 1980; revised 19 May 1981*

**Summary.** Skin colour has been measured by reflectance spectrophotometry on 134 pairs of twins at three sites, forehead, forearm and upper arm, each at three wavelengths, 425, 545 and 685 nm. Tanning is measured most reliably at 685 nm and at this wavelength the heritability is high at the least exposed upper arm site, intermediate on the forearm, while on the forehead variation is entirely environmentally determined. The same gradient is observed, though less strikingly at 545 nm, but at 425 nm, where haemoglobin is reflecting most of the light, the degree of genetic determination is the same at all sites.

## 1. Introduction

Following the introduction of reflectance spectrophotometry, Harrison and Owen (1964) examined the genetics of skin colour in hybrids between Europeans and West Africans. Because there was no overlap in the parental distributions, they were able to assume that they were dealing with pure-breeding lines, and so to estimate the potence ratio and number of effective factors (Mather and Jinks 1971). While this was a unique opportunity in human quantitative genetics, it is of considerable interest to examine their necessary assumption that variance within the two parental races and the $F_1$ is all environmental.

In this paper the classical twin method is used to dissect the causes of variation in skin reflectance in a population of European descent. The importance of environmental factors in skin colour has been inferred from different rates of tanning between parts of the body differentially exposed to the sun and during maturation (e.g. Harrison, Owen, da Rocha and Salzano 1967, Walsh 1963). Given that tanning is an important environmental influence on skin colour, the relative importance of genetic and environmental variation will clearly depend on differences in exposure to the sun. The patterns of variation described below this apply to only one set of environmental conditions.

According to Harrison and Owen (1964), melanin content is measured most accurately at 685 nm and it can be predicted that environmental factors would have most influence on an exposed site (e.g. forehead) measured at this wavelength and least at an unexposed site (e.g. upper arm). Haemoglobin has its maximal absorption at 545 nm, so melanin is measured less accurately and a less pronounced, but similar pattern of heritabilities might be expected. At 425 nm tanning is measured least accurately because melanin skin and blood all absorb light strongly at this wavelength. Therefore any differences in genetic determination between sites at this wavelength would not be influenced by tanning.

0301–4460/81/0806 0529 $02.00 

Table 1. Age, sex and zygosity composition of the sample.

| | MZ Males | MZ Females | DZ Males | DZ Females | DZ Opposite sex |
|---|---|---|---|---|---|
| Number of pairs | 23 | 45 | 20 | 33 | 13 |
| Mean age (years) | 24·39 | 26·16 | 20·95 | 25·52 | 20·85 |
| Age range | 17–53 | 15–64 | 17–37 | 16–47 | 17–31 |

## 2. Sample and measurements

A sample of 134 twin pairs was obtained in the Sydney area by appeals in the press. The sex, zygosity and age distribution of the sample are shown in table 1. There is the familiar excess of female pairs over male observed in all volunteer twin studies, so there will be better discrimination between alternative models for variation in females. Most of the twins are in their twenties but there are sufficient older pairs to create some problems in the analysis of variables heavily dependent on age. Inferences about the causes of variation must take account of the age distribution of the sample. The twins are nearly all of Northern European ancestry, but the inclusion of a small proportion of Southern European ancestry may tend to inflate genetic variance between pairs out of proportion to within-pair variance.

Zygosity was diagnosed by blood typing with the following antisera: anti A, $A_1$, B, C, c, D, E, e, M, N, S, s, $Fy^a$, K and $P_1$. In addition, twin pairs were typed for HLA using up to 29 antisera and for the serum protein haptoglobin, the red cell enzyme acid phosphatase and the Gc group. Secretor, colour vision and PTC tasting status were also determined. Eye colour, hair colour and earlobe form were noted. It is very unlikely, therefore, that any twins have been misclassified with respect to zygosity. Anthropometric data collected on the same sample have been analysed elsewhere (Clark, Jardine, Martin, Stark and Walsh 1980).

Skin reflectance was measured with an EEL reflectometer as described by Weiner and Lourie (1969). Reflectance was measured at three sites: (1) on the forehead midway between the hairline and the bridge of the nose; (2) the centre of the medial aspect of the left forearm midway between elbow and wrist; (3) the centre of the medial aspect of the left upper arm midway between elbow and shoulder. Measurements were made at three wavelengths at each site: (*a*) 425 nm—filter 601; (*b*) 545 nm—filter 605; (*c*) 685 nm—filter 609. Subjects were measured during the course of a year so the measurements should not be taken as reflecting a predominantly summer tanning nor a predominantly winter paleness, but rather a cross-section of seasonal states for Sydney. Both members of a pair were generally measured on the same day so any seasonal heterogeneity should contribute only to between-pair environmental variation ($E_2$).

In the few cases where individuals were not measured on all wavelengths, these missing observations can be seen from the degrees of freedom attached to the observed mean squares in table 6.

## 3. Methods of analysis

*Scaling*

Harrison and Owen (1964) spent considerable effort in search of suitable scales of measurement. Their criteria were: (i) to minimize genotype–environment interaction, as judged by the inequality of the parental and $F_1$ variances, and (ii) to minimize

directional dominance as measured by the departure of the $F_1$ mean from the mid-parent. They found that reflectance at 545 nm needed no transformation, but the most suitable transformation at 425 nm was $\log_{10} R$ and at 685 nm, $\text{antilog}_{10} R$ and these transformations will be employed here. The ensuing analysis has been performed on both the transformed and untransformed scales, but in most cases there was very little difference in the results.

The problem of genotype–environment interaction was considerable in the wide range of values which Harrison and Owen were dealing with. There is a much narrower range in this study, so to check the necessity and efficacy of these transformations the test of genotype–environment interaction proposed by Jinks and Fulker (1970) was used. These authors showed that certain types of genotype–environment interaction proposed by Jinks and Fulker (1970) was used. These authors showed that certain types of genotype–environment interaction ($G \times E$) could be detected by regressing the absolute differences on MZ pairs (a measure of individual environmental differences, $E_1$) on their pair-sums (a measure of genotype ($G$) and/or family environment ($E_2$)). Martin and Eysenck (1976) showed that such interactions could be detected with great sensitivity, but that they could nearly always be removed by a transformation of the scale of measurement which lessened departures from normality. It was also found that, in most cases, such transformations had a negligible effect on the results of model-fitting to variance components and this experience is borne out here.

The proportions of variance in absolute differences within MZ pairs accounted for by variation in MZ sums are shown in table 2. Both linear and quadratic regression components are shown and it can be seen that for both male and female MZ pairs there is no significant evidence of systematic $G \times E$ interaction in the raw data at either 425 or 685 nm. When the appropriate transformations are carried out on the raw data, in some cases $G \times E$ is decreased and in others it is increased. In two cases transformation introduces levels of $G \times E$ which are just significant, but in view of the number of significance tests employed these may not be reliable. In general it appears that $G \times E$ is not a significant problem in the untransformed measurements within the range of values found here, but since they do not appreciably worsen the situation, for consistency the transformations suggested by Harrison and Owen will be employed.

Table 2. Genotype–environment interaction before and after transformation at wavelengths 425 and 685 nm. Shown are the proportions of variance in absolute pair differences accounted for by regression on pair sums.

| | Untransformed variable | | | | Transformed variable | | | |
|---|---|---|---|---|---|---|---|---|
| Wavelength (nm) | MZ Male | | MZ Female | | MZ Male | | MZ Female | |
| | Linear | Quadratic | Linear | Quadratic | Linear | Quadratic | Linear | Quadratic |
| *Forehead* | | | | | | | | |
| 425 | 0·13 | 0·13 | 0·02 | 0·02 | 0·03 | 0·03 | 0·00 | 0·00 |
| 685 | 0·04 | 0·04 | 0·05 | 0·05 | 0·01 | 0·01 | 0·01 | 0·01 |
| *Upper arm* | | | | | | | | |
| 425 | 0·00 | 0·05 | 0·00 | 0·00 | 0·07 | 0·09 | 0·06 | 0·07 |
| 685 | 0·04 | 0·08 | 0·00 | 0·04 | 0·01 | 0·17 | 0·03 | 0·04 |
| *Forearm* | | | | | | | | |
| 425 | 0·01 | 0·06 | 0·03 | 0·03 | 0·01 | 0·06 | 0·10* | 0·10 |
| 685 | 0·02 | 0·02 | 0·00 | 0·00 | 0·05 | 0·09 | 0·01 | 0·22* |

* $0{\cdot}01 < P < 0{\cdot}05$.
** $0{\cdot}001 < P < 0{\cdot}01$.
*** $P < 0{\cdot}001$.

On these scales of measurement, and within this range of genotypes and environments, it can be concluded that different genotypes do not tan differentially on exposure to sunlight. From a given starting point of paleness, different genotypes should darken by the same amount (but not to the same end-point) on exposure to a given amount of sunlight. This conclusion could be tested critically by taking a range of identical twin pairs showing differing pre-tanning reflectances and then subjecting one member of each pair to a fixed regime of sun-tanning while the other remained unexposed.

*Correction for sex differences and regression on age*

Before attempting to fit models to explain trait variation, it is important to ensure that MZ and DZ groups have been sampled in the same range, i.e. that the subgroup means and variances are comparable. No significant differences between MZ and DZ means or total variances in males nor between these zygosity means and variances in females were found.

The raw data to which models of variation are fitted are the between- and within-pairs mean squares from an analysis of variance of each separate group of $n$ twin pairs:

| Source | Degree of freedom | Expected mean squares |
|---|---|---|
| Between pairs | $n-1$ | $\sigma_w^2+2\sigma_b^2$ |
| Within pairs | $n$ | $\sigma_w^2$ |

If a variable is strongly age-dependent, heterogeneity between age structures of subgroups will inflate the between-pairs mean squares to different degrees. Thus it is important to correct the between-pairs sum of squares by subtracting the sum of squares for regression of pair means on age. Table 3 shows the linear age correlations with the appropriately transformed variables. All the correlations, and many of these are significant, suggest that older people are darker. This is particularly true for the more exposed sites on the forehead and forearm and presumably reflects the cumulative effects of exposure to the sun. The same pattern with ageing has been found by Walsh (1964) in New Guineans.

For all measurements showing significant age correlations, the between-pairs mean squares have been corrected for regression on age and the corresponding degrees of freedom (table 5) reduced by one to $n-2$. The only variables not corrected were upper

Table 3. Correlations between skin reflectance (appropriately transformed) and age.

| Wavelength | Males | Females |
|---|---|---|
| *Forehead* | | |
| 425 | −0·25* | −0·25** |
| 545 | −0·24* | −0·30*** |
| 685 | −0·14 | −0·22** |
| *Upper arm* | | |
| 425 | −0·07 | −0·05 |
| 545 | −0·09 | −0·10 |
| 685 | −0·21* | −0·03 |
| *Forearm* | | |
| 425 | −0·21* | −0·20* |
| 545 | −0·22* | −0·16* |
| 685 | −0·36*** | −0·16* |

arm at 425 and 545 nm, where neither sex shows a significant correlation between skin reflectance and age.

A large difference in the means of males and females will inflate the within pairs mean squares (WMS) of DZ opposite sex pairs by an amount $\frac{1}{2}n(\bar{M}-\bar{F})^2$, where there are $n$ pairs, $\bar{M}$ is the male mean and $\bar{F}$ is the female mean. Because there were only 13 DZ opposite-sex pairs, it was decided that a more reliable correction would be obtained by using the total sample values of $\bar{M}$ and $\bar{F}$. The residual WMS (now with $n-1$ d.f.) is given by $[n/(n-1)][\text{WMS}-\frac{1}{2}(\bar{M}-\bar{F})^2]$.

Male and female means and variance for the total sample are given in table 4 and in every case females show a significantly higher reflectance (i.e. are lighter) than males. This is a general finding (Harrison 1973) reported in Brazilians (Harrison and Salzano 1966, Harrison *et al.* 1967), New Guineans and Pacific peoples (Walsh 1963, 1964), Sikhs (Kahlon, 1976) and Tibetans (Kalla and Tiwari 1970).

The DZ opposite sex within-pairs mean square has been corrected for sex differences in all variables.

Table 4. Total sample means and variances of skin reflectance (appropriately transformed) for males and females.

| | Males | | Females | |
|---|---|---|---|---|
| Wavelength | Mean | Variance | Mean | Variance |
| *Forehead* | | | | |
| 425 | 1·355 | 0·00467 | 1·402 | 0·00394 |
| 545 | 27·0 | 11·731 | 30·1 | 15·249 |
| 685 | 0·399 | 0·00111 | 0·412 | 0·00086 |
| *Upper arm* | | | | |
| 425 | 1·402 | 0·00880 | 1·444 | 0·00660 |
| 545 | 31·3 | 22·034 | 34·2 | 19·820 |
| 685 | 0·391 | 0·00088 | 0·401 | 0·00066 |
| *Forearm* | | | | |
| 425 | 1·357 | 0·00789 | 1·436 | 0·00539 |
| 545 | 28·8 | 19·928 | 34·0 | 18·054 |
| 685 | 0·376 | 0·00126 | 0·405 | 0·00058 |

Table 5. Age-corrected correlations between skin reflectance measures (appropriately transformed) for males (upper triangle) and females (lower triangle). Decimal points omitted.

| | Males ($N=96$) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Forehead | | | Upper arm | | | Forearm | | |
| Females ($N=161$) | 425 | 545 | 685 | 425 | 545 | 685 | 524 | 545 | 685 |
| *Forehead* | | | | | | | | | |
| 425 | — | 78 | 66 | 47 | 44 | 44 | 45 | 39 | 31 |
| 545 | 74 | — | 64 | 29 | 32 | 31 | 31 | 29 | 29 |
| 685 | 61 | 61 | — | 32 | 34 | 31 | 32 | 35 | 32 |
| *Upper arm* | | | | | | | | | |
| 425 | 55 | 46 | 36 | — | 86 | 79 | 76 | 67 | 51 |
| 545 | 55 | 48 | 37 | 84 | — | 79 | 74 | 77 | 56 |
| 685 | 42 | 41 | 39 | 77 | 76 | — | 69 | 64 | 62 |
| *Forearm* | | | | | | | | | |
| 425 | 57 | 44 | 35 | 81 | 78 | 71 | — | 89 | 72 |
| 545 | 49 | 41 | 39 | 70 | 79 | 66 | 86 | — | 74 |
| 685 | 44 | 43 | 39 | 67 | 71 | 81 | 74 | 73 | — |

One would expect the forehead to be the most tanned site and the upper arm the least tanned and this difference is significant in both males and females at 425 and 545 nm but not at 685 nm.

Age-corrected correlations between the measurements are shown in table 5. The correlations are all highly significant and are largest between different wavelengths at the same site rather than between the same wavelengths at different sites. Correlations between the two sites on the arm are higher than between these sites and the forehead.

*Fitting models*

For each variable there is now a set of ten mean squares, corrected for age and sex differences where appropriate and these are listed in table 6. Models of variation can be fitted to these mean squares using the method of weighted least squares which has been discussed extensively in the literature (Martin 1975, Eaves, Last, Young and Martin 1978).

A simple model for variation in MZ and DZ mean squares is shown in table 7. $E_1$ is environmental variance within families and as such it is specific to the individual and will include error variance. $E_2$, on the other hand, includes sources of environmental variance shared by members of a family but differing between families. It will thus embrace the lasting effects of cultural and class differences and parental rearing practices. In the present case it may include those habits and experiences influencing tanning which both members of a pair of twins share in common, but which differ between pairs. $V_A$ is that part of the genetic variation due to the additive effects of genes in the absence of assortative mating. If assortative mating occurs for skin colour, it will increase the additive genetic variance between pairs but in proportions such that it is completely confounded with estimates of $E_2$. Baldwin and Damon (1973) report assortative mating for skin colour among the heterogeneously coloured people of the Solomon Islands, but we known of no data to suggest that it occurs to any degree within European populations and it will not be considered further in this paper.

The model omits mention of dominance variance, $V_D$. In practice it is extremely difficult to detect in classical twin studies, even under ideal conditions and where it occurs it will usually be confounded with $V_A$ and $E_2$ (Martin, Eaves, Kearsey and Davies 1978). One of the criteria for transformation used by Harrison and Owen (1964) was the removal of the directionality of dominance. This does not usually affect the estimate of dominance variance, but in their analysis of European–African hybrids the estimate of $V_D$ was negligible and it is likely that it will be even less important within populations.

A sensible hierarchy of models is to first fit $E_1$ alone. Failure of this most simple model will indicate that there is significant between-families variation. A model incorporating $E_1$ and $E_2$ will test whether the between-families variation is entirely environmental in origin, while the $E_1 V_A$ model will test whether it is entirely genetic. If both two-parameter models fail, a model incorporating all three sources of variation must be considered. Since this model specifies that the total variances of MZ and DZ twins are equal, the model matrix is not of full rank and a maximum of three parameters can be estimated. Thus, when a model including $E_1$, $E_2$ and $V_A$ is fitted the remaining degree of freedom will simply test the equality of MZ and DZ variances.

There is no necessary reason why the components of variation will be the same in males and females, so models are first fitted to the sexes separately and then the eight statistics together. At this stage a heterogeneity chi-square for $k$ degrees of freedom can be calculated by adding the two male and female chi-squares for $4-k$ d.f. and

Table 6. Observed mean squares used for model fitting corrected for age† and sex.

| | Forehead | | | | | | Upper arm | | | | | | Forearm | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 425 | | 545 | | 685 | | 425† | | 545† | | 685 | | 425 | | 545 | | 685 | |
| Statistic | d.f. | Observed MS | d.f. | Observed MS | d.f. | Observed MS | d.f. | Observed MS | d.f. | Observed MS | d.f. | Observed MS | d.f. | Observed MS | d.f. | Observed MS | d.f. | Observed MS |
| $MZM_b$ | 20 | 0·007740 | 21 | 15·61 | 21 | 0·001480 | 21 | 0·017238 | 22 | 45·13 | 21 | 0·001970 | 20 | 0·015360 | 21 | 35·07 | 21 | 0·001870 |
| $MZM_w$ | 22 | 0·000774 | 23 | 3·67 | 23 | 0·000881 | 22 | 0·002706 | 23 | 6·06 | 23 | 0·000170 | 22 | 0·002173 | 23 | 6·99 | 23 | 0·000274 |
| $MZF_b$ | 42 | 0·006965 | 43 | 23·86 | 43 | 0·001550 | 43 | 0·010333 | 43 | 30·19 | 43 | 0·001120 | 42 | 0·007750 | 42 | 21·49 | 43 | 0·000715 |
| $MZF_w$ | 44 | 0·001360 | 45 | 6·31 | 45 | 0·000440 | 44 | 0·001853 | 44 | 5·51 | 45 | 0·000111 | 44 | 0·000953 | 44 | 5·80 | 45 | 0·000205 |
| $DZM_b$ | 18 | 0·006930 | 18 | 10·46 | 18 | 0·001450 | 19 | 0·013388 | 19 | 22·39 | 18 | 0·001135 | 18 | 0·010675 | 18 | 23·19 | 18 | 0·001425 |
| $DZM_w$ | 20 | 0·002074 | 20 | 7·33 | 20 | 0·000800 | 20 | 0·003103 | 20 | 11·44 | 20 | 0·000503 | 20 | 0·002495 | 20 | 10·71 | 20 | 0·001208 |
| $DZF_b$ | 29 | 0·005085 | 31 | 20·49 | 31 | 0·001025 | 30 | 0·010460 | 32 | 36·55 | 31 | 0·001385 | 29 | 0·009595 | 31 | 34·17 | 31 | 0·000910 |
| $DZF_w$ | 31 | 0·001901 | 33 | 7·92 | 33 | 0·000314 | 31 | 0·004118 | 33 | 11·29 | 33 | 0·00347 | 31 | 0·003048 | 33 | 10·26 | 33 | 0·000327 |
| $DZO_b$ | 11 | 0·001795 | 11 | 7·39 | 11 | 0·000690 | 12 | 0·009528 | 12 | 24·95 | 11 | 0·000860 | 11 | 0·007195 | 11 | 23·56 | 11 | 0·001475 |
| $DZO_w$ | 12 | 0·002093 | 12 | 8·03 | 12 | 0·000660 | 12 | 0·001912 | 12 | 14·48 | 12 | 0·000261 | 12 | 0·000920 | 12 | 8·21 | 12 | 0·000316 |

† Between pairs mean squares not corrected for regression on age.

Table 7. Simple model for mean squares of a classical twin study.

| | $E_1$ | $E_2$ | $V_A$ |
|---|---|---|---|
| $MZM_b$ | 1 | 2 | 2 |
| $MZM_w$ | 1 | 0 | 0 |
| $MZF_b$ | 1 | 2 | 2 |
| $MZF_w$ | 1 | 0 | 0 |
| $DZM_b$ | 1 | 2 | $1\frac{1}{2}$ |
| $DZM_w$ | 1 | 0 | $\frac{1}{2}$ |
| $DZF_b$ | 1 | 2 | $1\frac{1}{2}$ |
| $DZF_w$ | 1 | 0 | $\frac{1}{2}$ |
| $DZO_b$ | 1 | 2 | $1\frac{1}{2}$ |
| $DZO_w$ | 1 | 0 | $\frac{1}{2}$ |

subtracting from the chi-square ($8-k$ d.f.) for the corresponding model of $k$ parameters fitted to all eight statistics. The heterogeneity chi-square for $k$ d.f. will indicate whether the same parameters are appropriate for both sexes. If it is not significant, the DZ opposite-sex data may be added and the same model fitted to all ten statistics.

## 4. Results of model fitting

Except in one variable, forehead at 685 nm in males, the $E_1$ model fails badly in all cases, indicating that there is significant between-families variation to be accounted for. Tanning is best measured at 685 nm and the exceptional case ($\chi^2_3 = 3{\cdot}08$) is consistent with a view that variation in reflectance at this wavelength on the forehead in males is due entirely to differences in exposure to the sun and that these are individual environmental differences which are not shared by members of a pair. However, the number of male twins is small so that the discrimination between alternative models is poor and effects which are substantial may not be estimated as significant. Martin *et al.* (1978) have shown that much larger numbers than are available for this analysis are required to obtain reliable discrimination between models for traits of intermediate heritability.

This can be seen when the alternative two-parameter models, $E_1E_2$ or $E_1V_A$, are fitted to the data. In most cases either model is adequate to explain the data for the sexes separately. Although the chi-squares attached to the $E_1E_2$ model are generally higher, they are only significant for the upper arm at 685 nm in both sexes, forearm at 685 nm in females and 425 nm in males. It is reassuring that the simple environmental model fails at the sites and wavelengths where environmental influences would be expected to be weakest and genetic influences most important.

Chi-squares for heterogeneity of fit between sexes were calculated as described above. There was no significant heterogeneity in six of the nine variables, so the models with the same parameters for males and females were fitted to all ten statistics. It is not appropriate to do this in the cases of heterogeneity (forehead at 685 nm, forearm at 425 and 685 nm) and these cases will be discussed below. However, for completeness the results of this first stage of model-fitting to the ten statistics for all nine variables are shown in table 8. The $E_1$ model failed badly in every variable and results of fitting this model are omitted. The significance of both the parameter estimates and the chi-squares which test the adequacy of the model is indicated. Heritability is calculated as $h^2 = \hat{V}_A/(\hat{E}_1 + \hat{V}_A)$ for the $E_1V_A$ model and $h^2 = \hat{V}_A/(\hat{E}_1 + \hat{E}_2 + \hat{V}_A)$ for the $E_1E_2V_A$ model and standard errors or these estimates are given.

At the upper arm site the $E_1V_A$ model gives a better fit than the $E_1E_2$ model at all wavelengths, particularly 545 and 685 nm. The $E_1E_2$ model fails at 685 nm, the only

Table 8. Results of model-fitting to ten statistics for all variables.

| Wavelength | $E_1$ | $E_2$ | $V_A$ | $\chi^2$† | $h^2$ |
|---|---|---|---|---|---|
| *Forehead* | | | | | |
| 425 | 0·00157*** | 0·00230*** | | 9·98 | |
| | 0·00110*** | | 0·00265*** | 7·34 | 0·71±0·06 |
| | 0·00114*** | 0·00085 | 0·00184* | 5·67 | 0·48±0·21 |
| 545 | 6·561*** | 5·827*** | | 9·51 | |
| | 5·166*** | | 7·056*** | 8·06 | 0·58±0·07 |
| | 5·289*** | 1·440 | 5·570* | 7·38 | 0·45±0·25 |
| 685‡ | 0·00056*** | 0·00038*** | | 13·02 | |
| | 0·00052*** | | 0·00041*** | 14·07 | |
| | 0·00055*** | 0·00036* | 0·00002 | 13·00 | |
| *Upper arm* | | | | | |
| 425 | 0·00274*** | 0·00458*** | | 10·03 | |
| | 0·00199*** | | 0·00512*** | 8·47 | 0·72±0·05 |
| | 0·00213*** | 0·00269* | 0·00249* | 5·81 | 0·34±0·19 |
| 545 | 8·762*** | 11·97*** | | 12·22 | |
| | 5·564*** | | 14·982*** | 3·84 | 0·73±0·05 |
| | 5·708*** | 2·747 | 12·308** | 3·27 | 0·59±0·21 |
| 685 | 0·00025*** | 0·00053*** | | 24·54** | |
| | 0·00013*** | | 0·00063*** | 7·80 | 0·83±0·03 |
| | 0·00013*** | 0·00016 | 0·00049*** | 6·73 | 0·63±0·17 |
| *Forearm* | | | | | |
| 425‡ | 0·00190*** | 0·00398*** | | 18·47* | |
| | 0·00127*** | | 0·00436*** | 17·01* | |
| | 0·00136*** | 0·00233* | 0·00219* | 13·25 | |
| 545 | 8·085*** | 9·677*** | | 7·27 | |
| | 5·960*** | | 11·666*** | 6·41 | 0·66±0·06 |
| | 6·359*** | 4·924 | 6·606* | 4·15 | 0·37±0·22 |
| 685‡ | 0·00041*** | 0·00036*** | | 55·35*** | |
| | 0·00024*** | | 0·00055*** | 30·21*** | |
| | 0·00024*** | −0·00009 | 0·00064*** | 27·48*** | |

† Degrees of freedom are eight for the $E_1E_2$ and $E_1V_A$ models and seven for the $E_1E_2V_A$ model.
‡ Significant heterogeneity of fit between sexes.

model to do so for reasons other than heterogeneity over sexes. Only in this case can we say unequivocally that variation cannot be explained by environmental agents alone. The necessity for larger sample sizes to enable clear discrimination between models is obvious. For the other variables, judgement of which is the more appropriate model will have to reply upon which parameter causes the largest reduction in chi-square. It can be seen that in all variables but forehead at 685 nm (where the data are heterogeneous), it is the $V_A$ parameter which causes a larger reduction in chi-square. Apart from this case, when all three parameters are fitted, the estimate of $V_A$ is always significant, but the estimate of $E_2$ is significant only at 425 nm in the upper arm and forearm. Because the importance of $E_2$ is equivocal heritability estimates have been listed for both the $E_1V_A$ and $E_1E_2V_A$ models. These can probably be taken as bounds of the true proportion of genetic variation. It should be realized that estimates of $V_A$ and $E_2$ are highly negatively correlated and not too much significance can be attached to the partition of variation between these two when both are present, particularly with small numbers of observations (Martin *et al.* 1978).

The three cases of heterogeneity of fit over sexes must now be considered. The first of these, forehead at 685 nm (heterogeneity $\chi^2_2 = 6{\cdot}83$ for $E_1V_A$ model) might have been predicted from the observation that an $E_1$ model alone is adequate to account for

variation in males, while $E_2$ appears to be important in females. In the other two cases, forearm at 425 and 685 nm (heterogeneity $\chi^2_2 = 7{\cdot}67$ and $\chi^2_2 = 22{\cdot}86$ respectively for $E_1 V_A$ model), the relative sizes of the parameter estimates are similar for males and females. However, from table 4 it can be seen that for forearm measurements the total variance of males is greater than that for females, significantly so at these wavelengths, and this difference is sufficient to cause model failure. If some male subjects are outdoor manual workers while others work indoors in sedentary jobs, this might tend to produce greater variance at the forearm site (at the critical wavelengths 425 and 685 nm) in males than in females who may not have such occupational variety. This greater variance in males is also seen to a lesser degree at the other sites.

A full model incorporating different sized $E_1$, $V_A$ and $E_2$ effects for males and females has been developed by Eaves (1977), illustrated in Eaves *et al.* (1978) and Clark *et al.* (1980), and is shown in table 9. $V_{A_{mf}}$ is the covariance between the genetic effects in males and the genetic effects in females. If the genes affecting a trait in males are quite different from those affecting the trait in females then $\hat{V}_{A_{mf}}$ will be zero. If the genes acting in males and females are the same but produce scalar differences in the two sexes, the correlation between the effects

$$r_{V_A} = \frac{V_{A_{mf}}}{\sqrt{(V_{A_m} V_{A_f})}}$$

will be one. A similar argument applies to $E_{2_{mf}}$, the covariation between $E_2$ effects acting in males and females. If $\hat{V}_{A_{mf}}$ and $\hat{E}_{2_{mf}}$ are zero, the between- and within-mean squares for DZOS pairs will be equal, i.e. the intra-class correlation for DZOS pairs will be zero. Clearly, however, fairly large numbers of opposite-sex pairs will be needed to make reliable inferences about the size of $V_{A_{mf}}$ and $E_{2_{mf}}$ and there are only 13 such pairs in this study.

While the data for forehead at 685 nm are not inconsistent with low heritabilities for males ($h^2_m = 0{\cdot}28$) and females ($h^2_f = 0{\cdot}54$), earlier model-fitting and inspection of the data suggest that the only important effects are environmental. Since the DZOS between- and within-mean squares are equal, it suggests that there is no covariation of $E_2$ effects acting in males and females (i.e. $E_2$ effects in same-sex pairs are essentially $E_1$ effects in opposite sex pairs). Therefore any covariance parameter is omitted from our model and

Table 9. Model for the covariance of genetic and of environmental effects in mean squares of DZ opposite sex twin pairs.

| | $E_{1_m}$ | $E_{1_f}$ | $E_{2_m}$ | $E_{2_f}$ | $E_{2_{mf}}$ | $V_{A_m}$ | $V_{A_f}$ | $V_{A_{mf}}$ |
|---|---|---|---|---|---|---|---|---|
| $MZB_{males}$ | 1 | · | 2 | · | · | 2 | · | · |
| $MZW_{males}$ | 1 | · | · | · | · | · | · | · |
| $MZB_{females}$ | · | 1 | · | 2 | · | · | 2 | · |
| $MZW_{females}$ | · | 1 | · | · | · | · | · | · |
| $DZB_{males}$ | 1 | · | 2 | · | · | $1\frac{1}{2}$ | · | · |
| $DZW_{males}$ | 1 | · | · | · | · | $\frac{1}{2}$ | · | · |
| $DZB_{females}$ | · | 1 | · | 2 | · | · | $1\frac{1}{2}$ | · |
| $DZW_{females}$ | · | 1 | · | · | · | · | $\frac{1}{2}$ | · |
| $DZB_{m-f}$ | $\frac{1}{2}$ | $\frac{1}{2}$ | $\frac{1}{2}$ | $\frac{1}{2}$ | 1 | $\frac{1}{2}$ | $\frac{1}{2}$ | $\frac{1}{2}$ |
| $DZW_{m-f}$ | $\frac{1}{2}$ | $\frac{1}{2}$ | $\frac{1}{2}$ | $\frac{1}{2}$ | $-1$ | $\frac{1}{2}$ | $\frac{1}{2}$ | $-\frac{1}{2}$ |

$V_{A_m} = V_A$ effect for males.
$V_{A_f} = V_A$ effect for females.
$V_{A_{mf}}$ = covariance of additive genetic effects in males and females.
Similarly for $E_2$ and $E_1$.

separate $E_1$ and $E_2$ effects for males and females are fitted. This yields the following estimates:

$$\hat{E}_{1_f} = 0{\cdot}000383^{***}$$
$$\hat{E}_{2_f} = 0{\cdot}000814^{***}$$
$$\hat{E}_{1_m} = 0{\cdot}000455^{***}$$
$$\hat{E}_{2_m} = 0{\cdot}000281^{*}$$

and the model fits well with $\chi^2_6 = 4{\cdot}44$. Variation in skin reflectance on the forehead at the wavelength which best measures tanning, is thus explicable purely in terms of environmental influences. In males 50% of this environmental variation is of the type shared with the twin, while in females 63% is shared with the twin sister and this difference in proportions appears to be the principal cause of heterogeneity. These $E_2$ effects presumably reflect similarities between co-twins in work and recreational habits or may simply reflect the fact that twins were measured at the same time.

By contrast, reflectance on the forearm at 685 nm appears to be little influenced by $E_2$. Heterogeneity has arisen largely because the total male variance is double that of females. If a model with different $E_1$ and $V_A$ terms for males and females and also a genetic covariation term is fitted the following estimates are obtained:

$$\hat{E}_{1_f} = 0{\cdot}000202^{***}$$
$$\hat{V}_{A_f} = 0{\cdot}000327^{***}$$
$$\hat{V}_{1_m} = 0{\cdot}000312^{***}$$
$$\hat{V}_{A_m} = 0{\cdot}000976^{***}$$
$$\hat{V}_{A_{mf}} = 0{\cdot}001182^{***}$$

and a value of $\chi^2_5 = 5{\cdot}75$. The heritabilities for males and females are much the same ($h^2_f = 0{\cdot}62 \pm 0{\cdot}08$, $h^2_m = 0{\cdot}76 \pm 0{\cdot}08$) and the correlation between genetic effects acting in males and in females is not significantly different from one.

At 425 nm in the forearm the result is similar, although there is some evidence that $E_2$ may be more important in males. If the same model is fitted as before we obtain:

$$\hat{E}_{1_f} = 0{\cdot}000956^{***}$$
$$\hat{V}_{A_f} = 0{\cdot}004207^{***}$$
$$\hat{E}_{1_m} = 0{\cdot}001908^{***}$$
$$\hat{V}_{A_m} = 0{\cdot}005018^{***}$$
$$\hat{V}_{A_{mf}} = 0{\cdot}010196^{***}$$

and $\chi^2_5 = 5{\cdot}89$. Heritabilities are $h^2_f = 0{\cdot}81 \pm 0{\cdot}05$ and $h^2_m = 0{\cdot}72 \pm 0{\cdot}09$. As before, $r_{V_A}$ is not significantly different from unity.

## 5. Discussion

Despite the small numbers of twin pairs measured in this study and the consequent inability to discriminate unequivocally between alternative models, some general trends have emerged which are consistent with expectation.

At the upper arm site heritability is high, 0·83 at 685 nm, 0·73 at 545 nm and 0·72 at 425 nm, and the amount of $E_2$ appears negligible, except possibly at 425 nm. This is consistent with the prediction that environmental factors will be least important at this least exposed site. Harrison and Owen (1964) based their estimates of heritability in African–European hybrids on measurements of the upper arm at 685 nm because they predicted that this would have the highest heritability. Their prediction is supported by these results. They assumed that variation within a population was environmental, but the present resresults show that the heritability of his measurement in another population of European ancestry is 0·83. If there is as much genetic variation at the same loci in Africans as is found here then estimates of heritability quoted by Harrison and Owen would have to be revised upwards.

Harrison and Owen also showed that measurement at 685 nm is the best indication of tanning and the results here are certainly consistent with this. Heritability is highest on the upper arm and less at the forearm, while at the forehead there is no genetic variation at all and skin colour is entirely influenced by environmental factors. The fact that a greater proportion of environmental variation occurs between pairs in females than in males may reflect a greater propensity of sisters to share the same activities than brothers.

The same pattern is apparent, but not so pronounced at 545 nm where there is still substantial genetic variation on the forehead measurement. Maximum heritabilities are 0·73 at the upper arm, 0·66 at the forearm and 0·58 at the forehead. The overall lower level of genetic determination may reflect the fact that this is more a measurement of haemoglobin than of tanning.

At the wavelength least reliable as a measure of tanning, 425 nm, no such gradient is apparent. Maximum heritabilities are the same ($\sim$0·72) at all sites but it is clear that family environment may also be an important influence on these measurements. At this wavelength absorption by melanin, skin and blood are confounded and it is possible that the measurements are related to flushing rather than tanning.

## Acknowledgements

This work was supported in part by a grant from the National Health and Medical Research Council of Australia. The authors would like to acknowledge the assistance of Dr. L. Y. C. Lai, Dr. Helen Bashir, Mrs. P. Rosenthal and the willing cooperation of the twins. We thank Professor J. B. Gibson for helpful comments on the manuscript.

P. Clark is now at the Institute of Clinical Pathology and Medical Research, Westmead, N.S.W., Australia.

## References

BALDWIN, J. C., and DAMON, A., 1973, Some genetic traits in Solomon Island populations V. Assortative mating, with special reference to skin colour. *American Journal of Physical Anthropology*, **39**, 195–202.

CLARK, P., JARDINE, R., MARTIN, N. G., STARK, A. E., and WALSH, R. J., 1980, Sex differences in the inheritance of some anthropometric characters in twins. *Acta Geneticae Medicae Gemellologiae* (in the press).

EAVES, L. J., 1977, Inferring the causes of human variation. *Journal of the Royal Statistical Society* A, **140**, 324–355.

EAVES, L. J., LAST, K. A., YOUNG, P. A., and MARTIN, N. G., 1978, Model-fitting approaches to the analysis of human behaviour. *Heredity*, **41**, 249–320.

HARRISON, G. A., 1973, Differences in human pigmentation: measurement, geographic variation, and causes. *Journal of Investigative Dermatology*, **60**, 418–426.

HARRISON, G. A., and OWEN, J. J. T., 1964, Studies on the inheritance of human skin colour. *Annals of Human Genetics, London*, **28**, 27–37.

HARRISON, G. A., OWEN, J. J. T., DA ROCHA, F. J., and SALZANO, F. M., 1967, Skin colour in Southern Brazilian populations. *Human Biology*, **39**, 21–31.

HARRISON, G. A., and SALZANO, F. M., 1966, The skin colour of the Caingang and Guarani Indians of Brazil. *Human Biology*, **38**, 104–111.

JINKS, J. L., and FULKER, D. W., 1970, Comparison of the biometrical genetical, MAVA and classical approaches to the analysis of human behavior. *Psychological Bulletin*, **73**, 311–349.

KAHLON, D. P. S., 1976, Age variation in skin colour: a study in Sikh immigrants in Britain. *Human Biology*, **48**, 419–428.

KALLA, A. K., and TIWARI, S. C., Sex differences in skin colour in man. *Acta Geneticae Medicae Gemellologiae*, **19**, 472–476.

MARTIN, N. G., 1975, The inheritance of scholastic abilities in a sample of twins. *Annals of Human Genetics, London*, **39**, 219–229.

MARTIN, N. G., and EYSENCK, H. J., 1976, Genetic factors in sexual behaviour. In *Sex and Personality*, Chapter 6, edited by H. J. Eysenck (London: Open Books).

MARTIN, N. G., EAVES, L. J., KEARSEY, M. J., and DAVIES, P., 1978, The power of the classical twin study. *Heredity*, **40**, 97–116.

MATHER, K., and JINKS, J. L., 1971, *Biometrical Genetics*. (London: Chapman and Hall).

WALSH, R. J., 1963, Variations of melanin pigmentation of the skin in some Asian and Pacific peoples. *Journal of the Royal Anthropological Institute*, **93**, 126–132.

WALSH, R. J., 1964, Variation in the melanin content of the skin of New Guinea natives at different ages. *Journal of Investigative Dermatology*, **42**, 261–265.

WEINER, J. S., and LOURIE, J. A., 1969, *Human Biology, A Guide to Field Methods* (Oxford: Blackwell Scientific Publishers).

Address correspondence to: N. G. Martin, Box 475 G.P.O., Canberra City, A.C.T., 2601, Australia.

**Zusammenfassung.** Die Hautfarbe wurde durch Reflex-Spektrophotometrie bei 134 Zwillingspaaren an drei Stellen—Stirn, Unterarm und Oberarm—gemessen, jede für drei Wellenlängen, 425, 445 und 685 nm. Bräunung wird am verläßlichsten bei 685 nm gemessen, und bei dieser Wellenlänge ist die Heritabilität für die am wenigsten ausgesetzte Stelle des Oberarms hoch, von mittler Größe am Unterarm, während die Variation bei der Stirn ganz durch Umwelt bestimmt ist. Der selbe Gradient wird, wenn auch weniger klar, bei 445 nm beobachtet, aber bei 425 nm, wo das Haemoglobin das meiste Licht reflektiert, ist der Grad der genetischen Bestimmung an allen Stellen der Gleiche.

**Résumé.** La couleur de la peau a été mesurée par spectrophotométrie de réflectance sur 134 paires de jumeaux à trois sites: le front, l'avant-bras et le bras, chacun à trois longueurs d'onde: 425, 545 et 685 nm. Le bronzage est mesuré le plus fidèlement à 685 nm, et à cette longueur d'onde l'héritabilité est élevée au site du bras, le moins exposé et intermédiaire à l'avant-bras tandis qu'au front la variation est déterminée entièrement par le milieu. Le même gradient est observé à 545 nm, quoique de façon moins tranchée, mais à 425 nm, où l'hémoglobine réflète la plus grande part de la lumière, le degré de détermination génétique est le même à tous les sites.