To the Editor

Disease concordance and sex similarity in twins: application of a multifactorial model with latent structure

SIR—Marshall and Knox¹ proposed a model for disease concordance in twins when the zygosity of the twin pairs was not available in the data. I think that Marshall and Knox may have been somewhat incautious in the conclusions they drew from their results. For instance, they claimed that the results for infantile pyloric stenosis were incompatible with polygenic inheritance. But the model used was very different from the multifactorial (polygenic) model,² and it is not possible to interpret the parameters of one model in terms of the other.

On applying the multifactorial model (see appendix) to the same data, I have found the results given in the table. For four of the diseases the fit is good, with χ^2 similar to that in the Marshall-Knox (MK) model. But many of the parameter estimates are incompatible with other evidence. Nevertheless, plausible parameter values² also give quite a good fit to the data. Is, therefore, the quality of fit of the MK model also fairly insensitive to the parameter values? If so then many of the conclusions of Marshall and Knox about the modes of transmission of these diseases are questionable also.

The MK model and that given in the appendix both exhibit "latent structure"—that is, the observable proportions derive from a mixture of two different populations⁵ (monozygotic and dizygotic twins). I conclude that the twin data considered is analysable

by such models, but caution is needed in drawing conclusions, as a wide range of parameter values may give a good fit with datasets of sizes met in practice.

T P HUTCHINSON

Department of Statistics and Operational Research,
Coventry (Lanchester) Polytechnic,
Priory Street,
Coventry CV1 5FB.

Appendix

NOTATION

$\Phi(x)$:	Univariate Normal integral—that is, the probability of a variable being less than x if it has a Normal distribution with mean 0 and standard deviation 1.
$L(x,y,r)$:	Bivariate Normal integral—that is, the probability that x and y are exceeded given that the variables have a bivariate Normal distribution with means 0, standard deviations 1, and correlation r .
$T_M (T_F)$:	Threshold such that males (females) manifest the disease if their disease liability exceeds it. The corresponding incidences of disease in males and females are I_M and I_F .
$r_1 (r_2)$:	Correlation between disease liabilities of monozygotic (dizygotic) twins.
$p (q)$:	Proportion of twin pairs that are monozygotic (dizygotic), $q = 1-p$.
$m (f)$:	Proportion of male (female) fertilised ova, $f = 1-m$.

Results from fitting the multifactorial model to the five datasets of table 5 of Marshall and Knox¹

		I_M	I_F	r_1	r_2	χ^2	$\chi^2(MK)$
		%	%				
Down's syndrome	(a)	0.35	0.35	1.00	0.00	4.8	4.7
	(b)	0.10	0.10	1.00	0.00	4.8	
Schizophrenia	(a)	7.9	15.1	0.98	0.00	4.6	1.8
	(b)	1.0	1.0	0.90	*	47.2	
	(c)	0.80	1.22	0.98	*	8.5	
Coronary occlusion (Notes d and e below)	(a)	0.94	0.40	0.85	0.85	21.7	26.4
Neural-tube defects	(a)	0.40	0.62	0.00	0.00	2.8	2.0
	(b)	0.30	0.30	0.76	*	14.6	
	(c)	0.25	0.34	0.70	*	9.5	
Infantile pyloric stenosis (Note e below)	(a)	23.6	5.9	0.07	0.07	2.3	2.3
	(b)	0.50	0.10	0.74	*	13.1	
	(c)	0.47	0.11	0.78	*	9.7	

(a) Parameters fitted by minimising $\chi^2 = \sum (O-E)^2/E$. χ^2 interpretable as χ^2 (2 degrees of freedom).

(b) Parameters based on independent evidence. χ^2 interpretable as χ^2 (6 df).

(c) Parameters close to those in (b), but giving a much reduced χ^2 . χ^2 not strictly interpretable as having a χ^2 distribution, but note that the 5% critical value for χ^2 (6 df) is 12.6.

(d) Fitting of plausible parameter values was not attempted because even the best fit was very poor, as also for the MK model. According to reference 3, there were 176 like-sexed dizygotic twin pairs, 72 of unlike sex, and 104 monozygotic. Possibly the failure of both models stems from these proportions being very different from those expected.

(e) There are minor discrepancies between two of the datasets in table 5 of Marshall and Knox (and as used by me) and the original publications.^{2,4}

PROPORTIONS

A bar—that is, $\bar{M}$, $\bar{F}$ —is used to indicate the presence of the disease. Details of the derivation of the following expected proportions are available from the writer.

$\bar{M} \bar{M}$	$mpL(T_M, T_M, r_1) + m^2qL(T_M, T_M, r_2)$
$\bar{M} M$	$2mp[1 - \Phi(T_M) - L(T_M, T_M, r_1)]$ $+ 2m^2q[1 - \Phi(T_M) - L(T_M, T_M, r_2)]$
$\bar{F} \bar{F}$	$fpL(T_F, T_F, r_1) + f^2qL(T_F, T_F, r_2)$
$\bar{F} F$	$2fp[1 - \Phi(T_F) - L(T_F, T_F, r_1)]$ $+ 2f^2q[1 - \Phi(T_F) - L(T_F, T_F, r_2)]$
$\bar{M} \bar{F}$	$2mfqL(T_M, T_F, r_2)$
$\bar{M} F$	$2mfq[1 - \Phi(T_M) - L(T_M, T_F, r_2)]$
$M \bar{F}$	$2mfq[1 - \Phi(T_F) - L(T_M, T_F, r_2)]$

REMARKS

As did Marshall and Knox, I assumed $m = f = 0.5$, $p = 0.38$ (or 0.263 for Down's syndrome). In fitting this model the condition $r_2 = \frac{1}{2}r_1$ was not imposed (this is so if the correlations are entirely genetic and the genetic variation entirely additive⁶), but it was required that $0 \leq r_2 \leq r_1 \leq 1$. Thus there are four parameters, as in the MK model. Fitting both correlations independently permits even the data on Down's syndrome to be successfully predicted, even though the polygenic hypothesis is not credible for Down's syndrome.

References

- ¹Marshall T, Knox EG. Disease concordance and sex similarity in twins. *J Epidemiol Community Health* 1980; **34**: 1-8.
- ²Curnow RN, Smith C. Multifactorial models for familial diseases in man. *Journal of the Royal Statist Society Series A* 1975; **138**: 131-69 (with discussion).
- ³Harvald B, Hauge M. Coronary occlusion in twins. *Acta Genet Med Gemellol* 1970; **19**: 248-50.
- ⁴MacMahon B, McKeown T. Infantile hypertrophic pyloric stenosis: data on 81 pairs of twins. *Acta Genet Med Gemellol* 1955; **4**: 320-9.
- ⁵Fielding A. Latent structure models. In: O'Muircheartaigh CA, Payne C, ed. *The analysis of survey data. I. Exploring data structures*. London: Wiley, 1977: 125-57.
- ⁶Curnow RN. The multifactorial model for the inheritance of liability to disease and its implications for relatives at risk. *Biometrics* 1972; **28**: 931-46.

Dr Marshall and Dr Knox reply

Hutchinson is quite right. We were rash to "disprove" all polygenic/multifactorial models. Indeed, we have difficulty in regarding such a pleomorphic proposition as a model at all, but rather as a general frame of reference within which *all* etiological studies are conducted. A proposition in such broad terms is incapable of disproof. We should have referred more specifically to an "additive" polygenic/multifactorial model—that is, one with "latent structure," in Hutchinson's terms. If there are differences between us we think that they hinge upon Popperian philosophy—that is, establishment of a fit between a model and the facts says only a little for the validity of the model, whereas a non-fit destroys it. We therefore feel unsympathetic towards a demonstrated fit between the data for IHPS and a polygenic/multifactorial model if another model (our own) implies a non-fit.

So far as the sensitivity of our model to parameter variations is concerned, we were unable to pursue this through analytical means, but did so through iterative fitting of different parameter values. So far as was possible within the space available, we indicated these limits in our tables and text in relation to the particular problems examined.

We accept that our model has a "latent structure"—that is, a set of premises on which the validity of its applications depend. In particular, we supposed the existence of both precleavage and postcleavage determining events, both of them necessary, the frequency of whose combination is the product of their individual frequencies. In circumstances where it could be shown that these premises were non-valid—as in the case of a genetic interaction between mother and fetus, for example—our numerical conclusions would collapse. The same would be true of any other model whose premises did not envisage such events. These problems apply, of course, to all inductive processes.