Twin Research (1998) 1, 71–77

http://www.stockton-press.co.uk/tr

A factor analysis of associations among self-reported immune related symptoms in a large twin sample

David L Duffy[1], Diana Battistutta[1], John D Mathews[2] and Nicholas G Martin[1]

[1]*Epidemiology Unit, Queensland Institute of Medical Research, Brisbane*
[2]*Menzies School of Health Research, Darwin, Australia*

We examined the cumulative prevalences of 22 symptoms thought to reflect immune system function reported in a questionnaire mailed to 7616 Australian twins. The associations between symptoms and demographic variables were expressed in terms of polychoric or polyserial correlations, and a principal components analysis performed. Factors representing underlying propensities respectively to allergic disease, various minor infections, diseases associated with aging such as arthritis, skin disease, and respiratory tract infection were extracted. Possible processes underlying these symptom clusters and the relative strengths and weaknesses of this type of analysis are discussed.

Keywords: allergy, epidemiology, questionnaires

Introduction

Exploratory epidemiological studies often throw up large numbers of possible interassociations between categorical variables which ideally should be looked at in an 'objective' manner. Usually, associations between such variables are expressed in terms of an odds ratio, and in the case of multiple variables can be examined by log-linear modelling or discriminant analysis. These approaches have advantages, but can be unwieldy where relationships between large numbers of variables are to be summarised. In this paper we present a different approach – that of factor analysis of tetrachoric correlations – which is more utilised in the social sciences. We have used this to try and simplify the relationships between the prevalences of 22 common symptoms as reported in a large questionnaire survey of Australian twins, with a view to performing later genetic epidemiological analyses. These symptoms were included in the questionnaire as possibly reflecting different aspects of immune system function.

Model

The threshold model is one of the older models for the analysis of count data.[1] It assumes that a given variable, recorded as a number of categories, in the simplest case present or absent, represents an underlying latent continuous variable (in the psychometric literature often called the response strength). When this latent variable exceeds a given threshold, the indicator or observed variable enters the associated category:

$$x=0 \text{ if } x^*<t;\ x=1 \text{ if } x^*\geq t \qquad (1)$$

where x = observed variable; x^* = underlying latent variable; t = threshold value estimable from $Pr(x = 1)$. In the multifactorial genetic model, the latent variable is the liability to a disease or trait, and can be shown to correspond to a continuous cumulative risk function rising to unity at the limit.[2] It is thus similar to the concept of frailty or susceptibility used in life table analysis.[3]

Relationships between the discrete observed variables are interpreted in terms of equivalent relationships between the (antecedent) continuous latent variables. In the traditional approach, which we have followed in this analysis, the latent variables are assumed to conform to a multivariate Gaussian distribution, an assumption for which some tests are available.[4] More recent extensions to theory allow this assumption to be relaxed,[5,6] but are computationally demanding.

In the dichotomous case, the proportion affected with a disease is used to estimate the threshold or cut-off z-score on the normal curve between normals and diseased (and its SE). A second trait or disease is then examined in the same way. The concordance of the two diseases in individuals is expressed as a tetrachoric correlation coefficient, that is the maximum likelihood (ML) estimate of the Pearsonian correlation between the two latent variables (under the assumption of bivariate normality). This is estimated by solving:

Correspondence: Dr DL Duffy, Epidemiology Unit, Queensland Institute of Medical Research, 300 Herston Rd, Royal Brisbane Hospital Post Office, Queensland 4029, Australia. Tel: +61 7 3362 0217; Fax: +61 7 3362 0101; E-mail: davidD@qimr.edu.au
Received 12 January 1998; revised and accepted 1 April 1998

$$\int_{-\infty}^{t_2}\int_{-\infty}^{t_1} \Phi(x_1^*,x_2^*,r)\, dx_1^* dx_2^* = \frac{a}{N}, \qquad (2)$$

where ab,c,d are the usual cells of the 2 × 2 table in which a is the number concordant for the endpoints in question, and N = a + b + c + d. One expression for asymptotic standard error of *r* (which is biased below N = 200[7]) is:

$$se = \frac{1}{N.\Phi(z_1,z_2,r)}\left(\tfrac{1}{a}+\tfrac{1}{b}+\tfrac{1}{c}+\tfrac{1}{d}\right)^{\frac{-1}{2}}. \qquad (3)$$

In the case of data where more than two categories are measured, a polychoric correlation (the ML extension of Eq 2) can be estimated,[6,8] while polyserial correlation coefficients can be derived in cases where one variable is ordinal and the other continuous. An advantage of the case of the polychoric correlation is that the goodness of fit of the threshold model to the observed data is easily tested. *The tetrachoric correlation can also be shown to be relatable to other correlation models for the 2x2 table.*[9,10]

The correlations so calculated can then be examined by conventional (Ordinary Least Squares, OLS) factor analysis, or as alluded to earlier, by various Weighted Least Squares (WLS) or ML approaches. The theory behind factor analysis is considerable, and well covered in standard textbooks,[11] and will not be enlarged on here. However, Muthen,[12] notes that the results obtained by the OLS approach to factor analysis for larger numbers of dichotomous variables are usually similar to those of the GLS approach.

Subjects and Methods

In the period from November 1980 to March 1982 a questionnaire was mailed to the 5967 pairs of twins over the age of 18 years registered with the Australian National Health and Medical Research Council Twin Registry. These pairs were volunteers who had been recruited through schools, community groups and by media advertising throughout Australia. Of this group, 3808 twin pairs returned completed questionnaires (a pairwise response rate of 64%). There was an individual response rate of 75%, suggesting little concordance for non-response.

The questionnaire was lengthy and included items on age, sex, zygosity, birth order, tobacco use, psychological traits, and a number of physical and psychiatric symptoms. Zygosity of twins was diagnosed by response to two items, supplemented in ambiguous cases by examination of photographs

Table 1 Estimated tetrachoric/polyserial correlation matrix for individuals of each sex. Asymptotic standard errors of correlations for males range from 0.02 to 0.09 (mean 0.04). For females 0.01 to 0.06 (mean 0.03). Male values above diagonal, female values below diagonal

Males (N=2742)

	Age	*Migraine*	*Other head*	*Sore throat*	*Cold sore*	*Flu*	*EBV*	*Hayfever*	*Sinus*	*Bronchitis*	*Asthma*
Age	–	0.06	–0.22	–0.28	0.11	0.10	–0.15	–0.03	0.04	0.00	–0.03
Migraine	0.00	–	–0.08	0.03	0.08	0.12	0.14	0.11	0.22	0.14	0.09
Other head	–0.16	0.03	–	0.58	0.17	0.39	0.09	0.18	0.18	0.25	0.09
Sore throat	0.19	0.13	0.59	–	0.24	0.56	0.12	0.16	0.19	0.41	0.12
Cold sore	0.11	0.18	0.22	0.32	–	0.19	0.04	0.00	0.07	0.13	0.06
Flu	0.16	0.12	0.39	0.51	0.29	–	0.11	0.19	0.23	0.39	0.15
EBV	–0.14	–0.02	0.14	0.16	0.09	0.09	–	0.14	0.11	0.17	0.09
Hayfever	–0.04	0.05	0.22	0.21	0.06	0.23	0.09	–	0.45	0.25	0.52
Sinus	0.02	0.14	0.23	0.22	0.12	0.23	0.06	0.47	–	0.32	0.27
Bronchitis	0.06	0.17	0.13	0.42	0.15	0.35	0.22	0.29	0.35	–	0.54
Asthma	–0.06	0.06	0.06	0.13	0.06	0.16	0.18	0.52	0.31	0.45	–
Pneumonia	0.22	0.12	0.06	0.16	0.14	0.29	0.15	0.09	0.15	0.32	0.22
Warts	–0.09	0.12	0.29	0.32	0.19	0.26	0.08	0.15	0.15	0.21	0.10
Boils	0.04	0.18	0.29	0.36	0.22	0.32	0.07	0.15	0.14	0.24	0.07
Psoriasis	0.16	–0.06	0.13	0.02	0.11	0.19	0.15	0.00	0.00	0.11	0.13
Eczema	0.06	0.05	0.13	0.11	0.12	0.19	0.12	0.17	0.15	0.l8	0.29
Other skin	0.04	0.13	0.22	0.19	0.20	0.18	0.14	0.09	0.18	0.24	0.10
Dust all	–0.05	0.06	0.22	0.21	0.08	0.18	0.09	0.70	0.42	0.27	0.58
Food all	0.00	0.11	0.22	0.19	0.12	0.25	0.21	0.35	0.21	0.31	0.32
Other all	–0.05	0.11	0.18	0.22	0.10	0.21	0.13	0.41	0.25	0.27	0.32
Stiff joints	0.23	0.20	0.14	0.14	0.22	0.19	–0.02	0.12	0.16	0.21	0.11
Arthritis	0.49	0.16	–0.03	–0.02	0.16	0.18	0.02	0.09	0.15	0.17	0.08
UTI	0.21	0.17	0.14	0.10	0.16	0.21	0.09	0.05	0.18	0.22	0.03
Smoker	0.18	0.07	–0.11	–0.06	0.08	0.05	0.01	–0.07	0.07	0.16	0.07
ALC FREQ	–0.05	0.00	0.00	–0.01	–0.05	–0.04	–0.03	–0.04	–0.09	–0.07	–0.09
	Age	Migraine	Other head	Sore throat	Cold sore	Flu	EBV	Hayfever	Sinus	Bronchitis	Asthma

Females (N=4869)

Table 1 *Continued*

Males (N = 2742)

Pneumonia	*Warts*	*Boils*	*Psoriasis*	*Eczema*	*Other skin*	*Dust all*	*Food all*	*Other all*	*Stiff joints*	*Arthritis*	*UTI*	*Smoker*	*ALC*	*FREQ*
0.14	–0.15	–0.09	0.09	–0.13	–0.06	0.00	0.06	–0.02	0.31	0.53	0.18	–0.11	0.03	
0.14	0.08	0.13	0.00	0.04	0.10	0.12	0.20	0.15	0.23	0.22	0.19	0.06	0.00	
–0.01	0.28	0.31	0.00	0.10	0.25	0.17	0.16	0.24	0.17	0.04	0.16	0.03	–0.05	
0.03	0.41	0.42	0.07	0.13	0.26	0.16	0.19	0.27	0.13	–0.01	0.19	0.02	–0.06	
0.09	0.07	0.13	0.07	0.05	0.09	0.05	0.09	0.06	0.16	0.14	0.14	0.02	0.00	
0.21	0.19	0.29	0.12	0.15	0.19	0.19	0.22	0.21	0.22	0.17	0.22	–0.03	–0.08	
0.15	0.09	0.08	0.18	0.07	0.05	0.10	0.08	0.09	0.00	0.04	0.10	0.04	–0.05	
0.16	0.11	0.12	0.05	0.22	0.21	0.76	0.37	0.45	0.08	0.06	0.11	–0.06	–0.05	
0.15	0.12	0.16	0.08	0.16	0.19	0.43	0.29	0.33	0.19	0.21	0.22	0.05	–0.08	
0.34	0.15	0.24	0.06	0.18	0.18	0.29	0.28	0.34	0.18	0.16	0.22	0.13	–0.09	
0.29	0.07	0.14	0.07	0.26	0.14	0.62	0.36	0.39	0.09	0.09	0.10	0.04	–0.05	
–	0.00	0.06	0.09	0.07	0.05	0.17	0.18	0.18	0.09	0.19	0.19	0.04	–0.03	
0.09	–	0.34	0.00	0.07	0.19	0.11	0.13	0.19	0.11	0.05	0.16	0.04	–0.06	
0.17	0.42	–	0.08	0.10	0.19	0.15	0.13	0.17	0.16	0.07	0.17	0.02	–0.07	
0.23	0.16	0.16	–	0.26	0.07	0.07	0.10	0.00	0.09	0.10	0.13	0.10	–0.04	
0.15	0.23	0.21	0.38	–	0.27	0.27	0.28	0.24	–0.01	0.02	0.07	0.01	–0.01	
0.10	0.22	0.21	0.19	0.28	–	0.20	0.28	0.38	0.16	0.09	0.12	0.06	–0.07	
0.03	0.12	0.09	–0.05	0.26	0.17	–	0.52	0.47	0.12	0.09	0.12	–0.00	–0.03	
0.17	0.16	0.17	0.17	0.31	0.25	0.51	–	0.42	0.19	0.18	0.18	0.01	–0.01	
0.13	0.15	0.25	0.15	0.24	0.29	0.44	0.48	–	0.16	0.17	0.14	0.01	–0.07	
0.19	0.13	0.25	0.09	0.09	0.23	0.19	0.28	0.17	–	0.66	0.29	0.07	–0.04	
0.23	0.04	0.16	0.10	0.14	0.19	0.12	0.18	0.15	0.61	–	0.26	0.01	0.00	
0.16	0.09	0.20	0.13	0.12	0.19	0.04	0.26	0.13	0.26	0.33	–	0.09	–0.05	
0.14	0.04	0.07	0.02	0.03	0.10	0.00	–0.04	–0.07	0.14	0.17	0.09	–	0.35	
0.00	–0.03	–0.06	–0.03	–0.01	–0.08	–0.02	–0.03	–0.00	–0.09	–0.05	–0.05	0.38	–	
Pneumonia	Warts	Boils	Psoriasis	Eczema	Other skin	Dust all	Food all	Other all	Stiff joints	Arthritis	UTI	Smoker	ALC	FREQ

sent in by the twins. This method has been shown to be at least 95% accurate in several other studies.[13,14] For the present analysis of intra-individual phenotypic associations, we have ignored the genetic relatedness of the twins, except as a convenient method to split the sample to examine the reliability of the factors generated.

The physical symptom checklist was prefaced: 'How OFTEN have you had any of the following?' and included 22 diagnoses and symptoms such as 'Hay fever', 'Asthma or wheezing', 'Stiff joints'. Responses fell into four categories: 'Never', 'Only as a child', 'Rarely', 'Quite often'. For the purposes of the present analysis, these last three responses have been combined into 'Ever' versus 'Never'. As a number of subjects had left occasional items from the checklist blank, the overall response to the 22 item physical symptom checklist was examined. This revealed that only 19/7616 individuals had left the table entirely uncompleted and that many respondents had only filled out the positive items that referred to them and left the remainder blank. It was therefore decided to score all blank responses as 'Never'. As a check, prevalences and correlations including the blanks as 'Never' and then as missing values were calculated. Only small differences were found.

Tobacco use is included as a dichotomous variable – ever smoked versus never smoked. Frequency of alcohol consumption is entered as a seven point scale: (1) every day; (2) 3–4 times per week; (3) twice a week; (4) once a week; (5) once or twice a month; (6) less frequently; and (7) never.

In addition to physical symptoms, we also examined scores from the seven item depression and anxiety subscales of the Delusions–Symptoms–States Inventory,[15] and the Neuroticism scale of the Eysenck Personality Questionnaire.[16] Highest educational achievement was also included and was measured on a seven point scale. Surprisingly, these were not significantly associated with reporting of any of the physical symptoms, and have not been included in the analyses presented.

Polychoric/polyserial correlation matrices were generated using the program PRELIS[6] (see Table 1), and then an OLS principal components analysis (PCA) was performed using SPSS Factor[17] to extract the components and perform a varimax rotation. We chose PCA because of its ease of interpretation. Initial analyses were done on the complete sample including sex and age as variables. Subsequently the sexes were analysed separately, and then split samples containing the first born twins or second born twins. The degree of similarity of the loadings

Table 2 Lifetime prevalences (%) reported within age cohorts for females

Symptom or diagnosis	*<25 y*	*26–35*	*36–45*	*46–55*	*56–65*	*>65 y*
Migraine/'sick' headache	40.9	46.8	53.6	54.3	48.4	39.9
Other headaches	84.4	85.8	81.6	74.9	70.5	59.6
Sore throats	90.4	89.4	80.8	75.7	73.9	70.5
Cold sores (eg on lip)	37.5	44.2	48.3	50.2	51.8	45.6
Influenza	53.9	65.1	65.9	64.8	66.3	69.4
Glandular fever	10.8	10.8	9.3	5.2	5.4	1.6
Hay fever	30.9	36.0	38.8	35.0	23.8	24.9
Sinus trouble	29.6	43.1	48.9	43.3	34.0	25.9
Bronchitis (chest cold)	39.1	44.9	46.4	39.0	40.5	43.5
Asthma or wheezing	12.5	12.6	17.8	11.8	9.1	9.3
Pneumonia	8.0	11.1	13.9	15.0	15.6	18.7
Warts on skin	58.3	57.9	50.3	52.6	44.8	32.6
Boils or bad pimples	49.2	53.0	50.8	47.0	40.8	32.6
Psoriasis	1.8	3.2	3.7	2.8	2.3	5.7
Eczema	11.2	11.9	12.3	5.4	4.5	5.2
Other skin trouble	18.7	21.7	20.4	16.7	16.4	9.8
Dust allergy	17.7	20.1	24.9	23.6	15.0	12.4
Food allergy	8.8	9.4	14.3	14.2	9.1	10.9
Other allergies	18.0	19.3	22.2	22.5	15.0	8.3
Arthritis	14.8	18.3	30.8	40.4	43.6	40.9
Stiff joints in morning	7.0	14.0	28.6	42.9	56.7	59.6
Kidney/bladder infection	18.0	39.4	43.9	42.9	38.0	33.2
Ever smoked	43.2	43.5	42.3	33.3	33.0	25.9
Drink alcohol daily	1.5	5.8	14.5	15.6	19.9	11.9
Totals ($n=4869$)	1522	1484	783	534	353	193

Table 3 Lifetime prevalences (%) reported within age cohorts for males

Symptom or diagnosis	*<25 y*	*26–35*	*36–45*	*46–55*	*56–65*	*>65 y*
Migraine/'sick' headache	31.8	31.6	36.5	38.0	30.1	25.7
Other headaches	76.4	78.5	77.1	70.2	61.5	49.6
Sore throats	86.4	88.8	84.5	74.0	72.7	69.9
Cold sores (eg on lip)	39.4	46.6	49.9	52.7	51.7	50.4
Influenza	52.5	64.8	71.1	67.8	76.2	62.8
Glandular fever	11.8	12.3	10.4	5.4	4.2	3.5
Hay fever	29.1	31.6	31.3	29.1	29.4	17.7
Sinus trouble	29.8	36.1	40.1	38.0	32.2	23.0
Bronchitis (chest cold)	39.6	44.9	49.6	41.5	50.3	47.8
Asthma or wheezing	13.5	16.0	13.6	12.4	11.9	6.2
Pneumonia	6.8	10.7	17.4	16.7	21.0	17.7
Warts on skin	60.3	59.8	59.1	55.0	52.4	44.2
Boils or bad pimples	54.7	58.7	64.9	57.4	62.2	58.4
Psoriasis	1.0	2.5	2.2	1.6	6.3	2.7
Eczema	6.1	7.6	9.0	11.6	7.7	5.3
Other skin trouble	16.8	21.8	23.2	19.0	25.2	17.7
Dust allergy	19.5	20.4	23.4	19.8	13.3	8.8
Food allergy	5.8	7.1	5.4	5.8	7.7	3.5
Other allergies	13.6	13.4	11.4	12.8	11.9	8.0
Arthritis	17.1	21.7	25.9	38.8	32.9	42.5
Stiff joints in morning	5.4	10.9	17.7	35.3	37.8	54.0
Kidney/bladder infection	4.7	9.7	9.8	13.2	13.3	16.8
Ever smoked	44.4	51.8	61.2	59.9	69.0	61.9
Drink alcohol daily	4.9	12.0	20.2	22.2	28.9	23.0
Totals ($n=2742$)	1032	829	367	258	143	113

on each factors extracted in the split samples is expressed as a congruency coefficient.[11]

Results

Prevalences

The lifetime prevalences for the physical symptoms and diseases are listed in Table 2 and Table 3. A number of these items exhibit a roughly constant lifetime prevalence across the different age strata under 65 years of age – that is most individuals who report these symptoms have experienced them by the age of 18 years. The drop-off in reporting in the over-65 age group of a number of conditions might represent mortality due to specific conditions, poor recall or secular trend. The 2:1 predominance of females over males is a common feature of volunteer twin studies.[18]

Principal components analysis

The factor loadings for females are presented in Table 4 and for males in Table 5. These are fairly consistent across the sexes, with the most important factor representing atopy or allergy related illnesses such as asthma. Reported bronchitis are associated with this factor but pneumonia only weakly associated.

The second factor (in order of eigenvalue size) seems to represent vulnerability to infection, particularly viral illness. Though orthogonal (uncorrelated) to the 'atopy' component, it is also associated with reported bronchitis. The third factor is associated with increasing age, arthritis, and pneumonia, and seems to represent the diseases of aging. It is interesting to note that the signs and magnitudes of the loadings on this component reflect the trends in age-specific cumulative prevalences of the particular items.

In females, the fourth factor extracted seems to embody a separate susceptibility to respiratory tract infection. The fifth factor represents a variety of skin diseases. The sixth factor represents the association between tobacco and alcohol consumption. It is weakly associated with reporting ever suffering bronchitis.

Looking at the responses from the male subjects, we find the same first six factors, with a slightly changed order. Specific points to note are that cigarette smoking is associated with the diseases of aging, while age now crossloads on to the specific respiratory tract infection factor. Examination of the congruency coefficients derived from the split sample analyses confirms these conclusions about the

first six main groups of physical symptoms within the 22 examined.

Discussion

The principal components derived in this analysis seemed to be consistent across different groups, and are easily interpreted in terms of various clinical entities. The first component represents allergic disease. Its pattern of loadings suggest that reported allergy to dust and hayfever seem to be the most specific single items associated with atopy in this Australian group compared with asthma for example (a finding consistent with genetic studies[19]). Self-reported eczema (not operationally defined in the questionnaire) is only weakly correlated with this component, suggesting that a broad interpretation of this term was made by the respondents. Bronchitis was correlated with the other items on this component, but presumably mainly via asthma as an intervening variable. This conclusion is supported by the fact that asthma and bronchitis loaded together on to a later component that seemed to represent lower respiratory tract disease.

The second major component is interpretable as a vulnerability to a number of minor infections. The loadings on this component imply that individuals reporting ever suffering a sore throat (over 80% of the group) are more likely to report headache, bronchitis, warts and so on. This may represent a group of younger, more educated subjects that better recall such common conditions. But, as mentioned earlier, the reporting of any symptoms was not significantly associated with any of the personality measures examined (loading on this component for anxiety score, –0.02; for depression score, –0.03; and for EPQ Neuroticism, 0.06), or with sex. Even smoking was not strongly correlated with reporting of any these conditions (with the exception of bronchitis). A biological hypothesis for this component might be that individuals at the lower end of the population distribution of immunological variables such as secretory IgA[20,21] or IgG subclass levels[22] will suffer from an excess of all the conditions seen

Table 4 Loadings on first seven rotated principal components for all female twins

	Factor 1	*Factor 2*	*Factor 3*	*Factor 4*	*Factor 5*	*Factor 6*	*Factor 7*
Dust allergy	0.86	0.04	–0.01	0.12	0.04	0.05	0.01
Hayfever	0.81	0.05	–0.05	0.09	0.05	0.00	0.05
Asthma	0.69	0.01	–0.05	0.44**	–0.06	0.04	–0.03
Other allergies	0.68	0.23*	0.09	–0.04	–0.05	0.03	0.04
Food allergies	0.63	0.11	0.17	–0.00	0.04	0.20	0.08
Sinusitis	0.54	0.15	0.17	0.07	–0.09	–0.03	0.22*
Other skin dis	0.37	0.33**	0.11	–0.33**	–0.12	0.29*	0.01
Sore throats	0.12	0.86	–0.08	0.11	–0.01	0.05	0.01
Other headaches	0.15	0.75	–0.02	–0.03	–0.03	0.01	–0.14
Influenza	0.12	0.62	0.21*	0.35**	0.05	0.12	–0.05
Boils	0.10	0.59	0.05	–0.00	–0.02	0.00	0.20*
Warts	0.11	0.55	–0.02	–0.19	–0.06	–0.07	0.28*
Cold sores	–0.06	0.33	0.25*	0.21	0.06	0.15	–0.09
Arthritis	0.12	0.02	0.85	0.04	0.01	0.03	0.09
Morning stiffness	0.12	0.20	0.76	–0.06	–0.09	–0.02	0.11
Age	–0.03	–0.27*	0.75	0.17	0.12	–0.00	–0.17
UTI	0.08	0.24*	0.39	0.15	–0.10	0.09	0.30**
Pneumonia	0.17	–0.02	0.15	0.67	–0.03	0.05	0.17
Bronchitis	0.37**	0.36**	0.07	0.59	–0.17	0.00	0.05
Cigarette use	–0.04	–0.01	0.00	0.03	–0.82	0.08	0.09
Alcohol use	–0.05	–0.05	–0.00	–0.04	0.79	0.02	0.07
Psoriasis	–0.05	–0.00	0.11	0.13	–0.07	0.79	0.10
Eczema	0.36**	0.09	–0.09	–0.07	0.01	0.67	–0.03
Migraine	0.17	–0.01	0.26*	–0.03	0.01	–0.09	0.70
EBV	0.02	0.09	–0.17	0.26*	0.00	0.24*	0.63
Split sample factor congruency	0.95	0.98	0.97	0.92	0.97	0.34	0.86

Boxed data represents the largest loadings on each component (** denotes other loadings explaining > 10% of variance; * denotes loadings explaining > 5% of variance).

associated with this component. Alternatively, it might be a function of socioeconomic status or nutrition.

The third component represents a number of conditions that one normally associates with increasing age, principally arthritis. In the males, the loading for age itself on this component is smaller than for the females, because of the crossloading on to the respiratory disease factor.

The later components vary in (eigenvalue) size across the sexes. Smoking and alcohol use are usually strongly associated in individuals, and appear as a single factor, which as noted earlier, is not strongly correlated with the *presence* of particular symptoms in females (with the exception of bronchitis), though it may of course have some effect on frequency or severity, not measured in this analysis. In males, who smoked more heavily, and increasingly so with age, there is the previously mentioned crossloading of age and smoking. All respiratory diseases tended to be interrelated. The factor representing skin diseases reflects a tendency for individuals to report more than one such problem. This may be due, as suggested above for eczema and allergy, to a lack of specificity in these items.

An extension of these conclusions that is usually made in the psychometric literature is that the associated item factor scores can then be used to form a scale that measures the underlying latent variable, such as 'atopy' or 'immune competency', and this compound score entered into later analyses. These analyses might include validating the scale using other measures of the same variable, whether physiological ones such as serum IgE, or hospital recorded diagnoses; it may involve using the scale score as a ('better') predictor for disease.

The usual (quasi-empirical) criteria for deciding the number of components in PCA are less helpful in the case of tetrachoric correlations. Small tetrachoric correlations equivalent to small odds ratios may nevertheless be clinically significant if the variables are common enough in the population. A notable

Table 5 Loadings on first seven rotated principal components for all male twins

	Factor 1	*Factor 2*	*Factor 3*	*Factor 4*	*Factor 5*	*Factor 6*	*Factor 7*
Dust allergy	0.88	0.06	0.07	0.05	–0.09	0.04	–0.02
Hayfever	0.84	0.14	–0.02	0.00	0.03	0.04	–0.08
Asthma	0.73	–0.03	–0.07	0.15	0.22	–0.15	0.19
Sinusitis	0.58	0.23*	0.13	–0.11	0.12	–0.14	–0.04
Other allergies	0.54	0.13	0.24*	0.28*	–0.14	0.15	0.17
Food allergies	0.51	0.09	0.34**	0.30**	–0.06	0.14	0.28*
Other skin dis	0.39	0.35**	0.16	–0.03	0.39**	–0.17	0.34**
Sore throats	0.17	0.81	–0.01	–0.13	0.10	0.07	0.19
Other headaches	0.16	0.72	–0.02	0.02	–0.00	0.13	0.12
Influenza	0.17	0.62	0.10	0.09	0.42	0.03	–0.04
Boils	0.04	0.59	0.24	0.25	–0.06	–0.08	–0.07
Warts	0.07	0.59	0.04	0.31**	–0.18	–0.12	–0.05
Cold sores	–0.02	0.46	0.27**	0.05	0.10	–0.07	–0.00
Arthritis	0.10	–0.06	0.75	0.11	0.29*	–0.05	–0.22*
Morning stiffness	0.15	0.14	0.71	0.05	0.08	–0.07	–0.15
UTI	0.00	0.12	0.57	0.08	0.15	–0.01	0.17
Migraine	0.04	0.17	0.50	–0.23	–0.12	–0.04	0.17
Other skin disease	0.11	0.24*	0.38	0.37**	–0.21	–0.14	0.22
Psoriasis	–0.07	0.10	–0.02	0.76	0.29*	0.01	0.06
Eczema	0.24*	0.12	0.06	0.71	0.02	–0.03	0.05
Pneumonia	0.08	0.12	0.17	0.13	0.69	–0.03	0.19
Age	–0.08	–0.17	0.39**	0.17	0.52	–0.08	–0.46**
Cigarette use	–0.04	–0.03	0.15	0.00	0.15	–0.81	0.01
Alcohol use	–0.07	–0.03	0.01	–0.03	0.07	0.79	–0.00
EBV	0.06	0.03	0.02	0.14	0.15	–0.01	0.77
Split sample factor congruency	0.99	0.92	0.90	0.86	0.08	0.94	0.26

Boxed data represents the largest loadings on each component (** denotes other loadings explaining > 10% of variance; * denotes loadings explaining > 5% of variance).

example in this dataset is the magnitude of the correlation between smoking, asthma and bronchitis. However, the patterns of loadings are useful in formulating hypotheses that can be examined in more detail using other modelling techniques.

In conclusion, we have uncovered a number of broad structures in the relationships between these common symptoms which seem mostly to correspond to medically recognised entities such as atopy. In addition, the relationship between resistance to a variety of types of reported minor infections in an individual might represent a global factor of immune function that warrants further study.

Acknowledgements

This study was supported by grants from the Australian National Health and Medical Research Council. We thank Dr John Hopper, the staff and members of the Australian NHMRC Twin Registry, and an anonymous reviewer.

References

1 Pearson K. Mathematical contributions to the theory of evolution, VII: On the correlation of characters not quantitatively measurable. *Philos Trans R Soc Lond A* 1901; **195**: 1–47.
2 Curnow RN, Smith C. Multifactorial models for familial diseases in Man. *J R Statis Soc A* 1975; **138**: 131–169.
3 Vaupel JW, Manton KG, Stallard E. The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* 1979; **16**: 439–454.
4 Muthen B, Hofacker C. Testing the assumptions underlying tetrachoric correlations. *Psychometrika* 1988; **53**: 563–578.
5 Mislevy RJ. Recent developments in the factor analysis of categorical variables. *J Educ Stat* 1986; **11**: 3–31.
6 Joreskog KG, Sorbom D. *PRELIS Users' guide*. 1st edn. Moorsville: Scientific Software, 1986.
7 Brown MB, Benedetti JK. On the mean and variance of the tetrachoric correlation coefficient. *Psychometrika* 1977; **42**: 347–355.
8 Olsson U. Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika* 1979; **44**: 443–460.
9 Kraemer HC. Ramifications of a population model for Kappa as a coefficient of reliability. *Psychometrika* 1979; **44**: 461–472.
10 Guilford JP, Fruchter B. *Fundamental Statistics in Psychology and Education* 5th edn. McGraw-Hill: Kogakusha, Tokyo, 1973.
11 Harman HH. *Modern Factor Analysis*. 3rd edn. University of Chicago Press: Chicago, 1976.
12 Muthen B. Contributions to factor analysis of dichotomous variables. *Psychometrika* 1978; **43**: 551–560.
13 Martin NG, Martin PG. The inheritance of scholastic abilities in a sample of twins: I. Ascertainment of the sample and diagnosis of zygosity. *Ann Hum Genet* 1975; **39**: 213–218.
14 Kasriel J, Eaves LJ. The zygosity of twins: further evidence on the agreement between diagnosis by blood group and written questionnaires. *J Biosoc Sci* 1976; **8**: 263–266.
15 Bedford A, Foulds GA, Sheffield BF. A new personal disturbance scale (DDSI/SAD). *Br J Soc Clin Psychiatry* 1976; **15**: 387–394.
16 Eysenck HJ, Eysenck SBG. *Manual of the Eysenck Personality Questionnaire* 1975. Educational and Industrial Testing Service: San Diego.
17 Norusis MJ, SPSS Inc. *SPSS/PC + 2.0 Advanced Manual*. SPSS Inc: Chicago, 1988.
18 Lykken DT, Tellegen A, DeRubeis R. Volunteer bias in twin research: the rule of two-thirds. *Soc Biol* 1978; **25**: 1–9.
19 Cookson WOCM, Faux JA, Sharp PA, Hopkin JM. Linkage between Immunoglobulin E responses underlying asthma and rhinitis and Chromosome 11q. *Lancet* 1989; **i**: 1292–1294.
20 Hobbs JR. Immune imbalance in dysgammaglobulinaemia type IV. *Lancet* 1968; **i**: 110.
21 Clark P *et al*. Directional dominance for low IgM and IgA levels. *Am J Hum Genet* 1981; **33**: 709–721.
22 Morgan G, Levinsky RJ. Clinical significance of IgG subclass deficiency. *Arch Dis Child* 1988; **63**: 771–773.