Psychological Medicine, 1994, **24**, 41–51.

The subtyping of schizophrenia in men and women: a latent class analysis

D. J. CASTLE,[1] P. C. SHAM, S. WESSELY AND R. M. MURRAY

From the Genetics Section, Institute of Psychiatry, London

SYNOPSIS Latent class analysis on an epidemiologically based series of 447 first contact patients with a broad diagnosis of schizophrenia revealed evidence for two subtypes: a 'neurodevelopmental' type characterized by early onset, poor pre-morbid social adjustment, restricted affect and a male:female ratio of 7:3; and a 'paranoid' type characterized by later onset, persecutory delusions and an almost equal sex ratio. A third 'schizoaffective' subtype, whose existence was less clear cut, was almost entirely confined to females and characterized by dysphoria and persecutory delusions, and had negligible familial risk of schizophrenia. The aetiological, biological and clinical significance of this typology remains to be tested.

INTRODUCTION

In attempting to enhance our understanding of those illnesses commonly subsumed under the label 'schizophrenia', attention has recently been directed at gender differences. That there are gender differences in schizophrenia is beyond doubt: males have an earlier age at onset (Lewine, 1981, 1988; Lewine *et al.* 1981; Loranger, 1984; Angermeyer & Kuhn, 1988; Goldstein *et al.* 1989; Riecher *et al.* 1989), respond less favourably to neuroleptic medication (Seeman, 1986), relapse more often (Angermeyer *et al.* 1989, 1990), and generally have a worse outcome than females (reviewed by Seeman, 1986; Shepherd *et al.* 1989). Males are also more likely to exhibit 'typical' schizophrenic symptomatology (see Lewine, 1981; Goldstein & Link, 1988), and some studies suggest that males are particularly likely to manifest 'negative' symptoms (Kay *et al.* 1986; Goldstein & Link, 1988; Haas *et al.* 1990).

Despite some detractors (Done *et al.* 1991), a history of obstetric complications has been reported with increased frequency among schizophrenic patients, particularly those with ventricular enlargement (Murray *et al.* 1985; Lewis & Murray, 1987). Furthermore, more male than female schizophrenics have a history of obstetric complications (Pearlson *et al.* 1985; Wilcox & Nasrallah, 1987; Owen *et al.* 1988; Foerster *et al.* 1991*a*; O'Callaghan *et al.* 1992), and a number of studies have revealed that male schizophrenics are more likely to have structural brain abnormalities than are their female counterparts (reviewed by Castle & Murray, 1991). These findings, together with evidence that male schizophrenics are more likely than females to have a history of pre-morbid social and occupational dysfunction (Zigler & Levine, 1973; Klorman *et al.* 1977; Zigler *et al.* 1977; Foerster *et al.* 1991*b*), and to have low pre-morbid IQ (Offord, 1974; Aylward *et al.* 1984), have led to the suggestion that males are particularly prone to a severe early-onset form of illness consequent upon neurodevelopmental damage (see Castle & Murray, 1991).

In contrast, later-onset schizophrenia occurs predominantly in women, and generally shows less pre-morbid dysfunction (reviewed by Castle & Murray, 1991). Female schizophrenics are more likely than males to exhibit 'atypical' and affective symptoms (Goldstein & Link, 1988). Several recent studies suggest that schizophrenia in women has a higher familial loading than in men (Bellodi *et al.* 1986; Goldstein *et al.* 1990; Pulver *et al.* 1992), but other authors have shown that late-onset schizophrenia has a lower

[1] Address for correspondence: Dr David J. Castle, Genetics Section, Institute of Psychiatry, De Crespigny Park, Denmark Hill, London SE5 8AF.

familial risk for the disorder (Harris & Jeste, 1988; Shimizu *et al.* 1988). In addition, 'atypical' cases generally show higher rates of affective disorder in their relatives (e.g. McCabe *et al.* 1971; Pope & Lipinsky, 1978). A history of obstetric complications is less likely in later-onset cases (Foerster *et al.* 1991*a*), and Owen *et al.* (1989) claimed that schizophrenics with a family history of affective disorder show fewer structural brain changes than those with a family history of non-affective functional psychosis. It has been suggested that later-onset female cases may have an illness which is distinct from the early-onset 'neurodevelopmental' type (Castle & Murray, 1991; Murray *et al.* 1992).

Recently, Goldstein *et al.* (1990) applied latent class analysis to the Iowa 500 schizophrenia sample, and delineated two subtypes of illness with very different sex prevalence. Males were over-represented in the subtype characterized by illness-onset before 25 years, restricted affect, poor pre-morbid adjustment, low familial risk of schizophrenia, and winter birth (thought to be a marker of 'environmental' causal factors). The second subtype, with a female preponderance, showed later onset, more dysphoria, more persecutory delusions, higher familial risk of schizophrenia, and non-winter birth.

The current study examines a series of first-contact patients with a broad diagnosis of schizophrenia using latent class analysis. We were particularly interested to see if a 'neuro-developmental' subtype, characterized by early onset, deficit features, poor pre-morbid adjustment and male preponderance, would emerge from a latent class analysis of these patients.

METHOD

The Camberwell Register First-Contact Sample

Subjects were drawn from the Camberwell Register First-Contact Sample, which has been described fully elsewhere (Castle *et al.* 1991). The Register recorded all patients from the defined catchment area of Camberwell, in South London, who had first contact with the psychiatric services between the years 1965 and 1984. The demography of the area has changed considerably over this period. The total population declined from 171000 in 1965 to 118000 in 1984. However, the male:female ratio remained stable across the years (52% of the population were female in both 1965 and 1984). Furthermore, there was remarkably little change in the age-structure of the population. Thus, in 1965, 38% of males and 35% of females were under 25, and 66% of males and 61% of females under 45. The comparable figures for 1984 were: 37% of males and 35% of females under 25, and 65% of males and 61% of females under 45. National figures for England and Wales are very similar; for example, the 1981 census shows that 38% of males and 34% of females were under the age of 25, and 65% and 60%, respectively, under the age of 45.

The current sample comprised 91% of all patients on the Camberwell Register, who on their first contact with the psychiatric services had received a Register diagnosis of 'schizophrenia' (ICD-9 codes 295.0–.9), 'schizoaffective disorder' (ICD 295.7), 'paraphrenia' (ICD 297.2), or 'other non-organic psychosis' (ICD 298.1–.9). Thus, we selected all patients with a functional psychotic illness which was not primarily affective. The residual 9% of patients were those for whom case-records were unobtainable. There is no reason to suspect that they differed in any meaningful way from those included in the study; specifically, they did not differ significantly in terms of gender or age at first contact.

Fig. 1 shows the age-at-onset distribution of males ($N = 245$) and females ($N = 236$) for all the ICD-9 groups. Mean age-at-onset was 31·2 years for males and 41·1 years for females. Males exceeded females in those patients with an onset of less than 35 years, whereafter a female preponderance was seen. According to DSM-III-R diagnostic criteria, 73% of the males and 69% of the females had schizophrenia or related disorders (i.e. atypical psychosis, schizophreniform psychosis, schizophrenia); mean age-at-onset for this group was 35·6 years (S.D. 19·0). In contrast, 9% of males and 20% of females fulfilled criteria for schizoaffective disorder (mean age-at-onset 28·8 years; S.D. 10·8), while the proportions for delusional disorder were 13% of the males and 8% of the females (mean age-at-onset 39·4 years; S.D. 18·8).

Variables

All patients were rediagnosed according to a range of operational definitions of schizophrenia, using the Operational Criteria

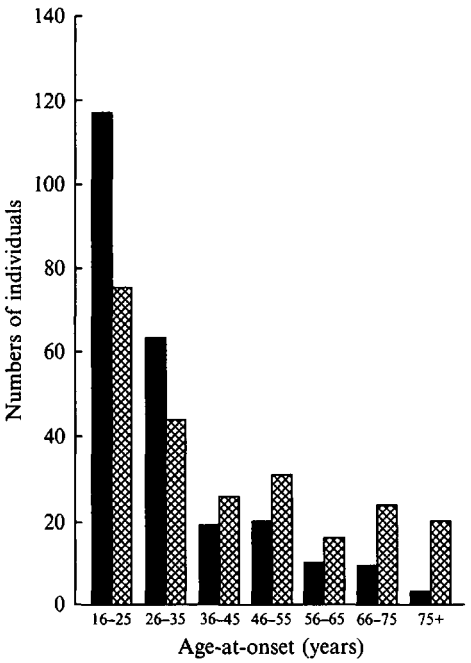

FIG. 1. Number of males and females, by age-at-onset. (■, Males; ⊠, females.)

(OPCRIT) Checklist (McGuffin *et al.* 1991). The checklist provides a simple, reliable method of applying multiple operational diagnostic criteria in studies of psychotic illness. Two independent workers (D.J.C. and S.W.) rated the case records using OPCRIT; S.W. was 'blind' to the fact that the study was addressing gender issues in schizophrenia, and inter-rater reliability for diagnoses, established on a random set of 50 case-records, was good (see Castle *et al.* 1991).

The variables for the analysis were chosen to approximate those used by Goldstein *et al.* (1990). These were: (1) family history of schizophrenia in first- or second-degree relatives (FH); (2) restricted affect (RA); (3) persecutory delusions (PD); (4) poor pre-morbid social adjustment (SA); (5) dysphoria (DS); (6) early onset, i.e. 25 years or younger (EO); (7) winter birth, i.e. December to April (WB); and (8) male sex (MS). The variables are defined in OPCRIT.

Latent class analysis

It is well known that the observed association between several variables can sometimes be explained by 'confounding' variables. Thus, in a sample of children, an overall association between height and reading ability would largely disappear after age is controlled for. In other words, one would expect height and reading ability to be roughly independent at any fixed age. However, if one knew nothing about the possibility that age was a confounder, then one might be led to postulate the existence of an unobserved, or latent, variable to account for the association between height and reading ability. Many statistical methods involve the construction of latent variables to explain the associations between observed or manifest variables (Everitt, 1984). Latent class analysis is a method which deals with categorical manifest variables and assumes a categorical latent variable (Green, 1951; Lazarsfeld & Henry, 1968).

To describe the latent class model, we consider the case of three manifest variables. Let A, B and C be variables with I, J and K categories respectively, and let X be the postulated latent variable with L categories (latent classes). The model assumes that, within each class of the latent variable, the manifest variables are independent. This leads to the equation

$$\pi_{ijk} = \sum_{m=1}^{L} \pi_m \pi_{i|m} \pi_{j|m} \pi_{k|m}$$

which states that, for any individual, the probability that $A = i$, $B = j$ and $C = k$ (π_{ijk}) is equal to the sum of the products of the probability of each latent class (π_m) and the conditional probabilities of $A = i$, $B = j$, and $C = k$ given $X = m$ ($\pi_{i|m}$, $\pi_{j|m}$ and $\pi_{k|m}$) over all the latent classes ($m = 1, \ldots, L$). The parameters of the model are therefore the L latent class probabilities π_m, and the $IJKL$ within class conditional probabilities ($\pi_{i|m}$, $\pi_{j|m}$ and $\pi_{k|m}$). Given any set of parameter values, one can calculate the probability of the observed data. The parameter values which maximize this probability are known as maximum likelihood estimates. They can be found by an iterative scheme such as the EM algorithm (Dempster *et al.* 1977).

It is generally possible to fit several different latent class models to the same set of data. Models can differ from each other in two respects: (1) the number of latent classes; and (2) restrictions among the parameters. The latter

involves equating certain parameters to constants or to other parameters. The choice between the possible models is made using three criteria: (1) the goodness-of-fit between model and data, measured globally by a χ^2 statistic, and locally by residuals; (2) the number of estimated parameters, or parsimony; and (3) whether the model has a 'sensible' interpretation. A 'good' model is sensible, parsimonious (i.e. has few estimated parameters) and fits the data (i.e. has a small χ^2 statistic and small residuals). However, there is no general formal statistical test for choosing between two models, unless one of them is nested on the other, that is, one of them can be obtained by restricting some of the parameters of the other. In this case, twice the natural logarithm of the likelihood ratio of the more general model to the more restricted model has asymptotically a χ^2 distribution with degrees of freedom equal to the number of restricted parameters.

Since schizophrenia shows considerable gender differences, we followed the example of Goldstein *et al.* (1990) in using simultaneous latent class models (Clogg & Goodman, 1985), with gender-specific latent classes. These models are specified by restricting the conditional probability of being male at unity in some latent classes, and at zero in others. Interest is focused on the number of latent classes in each gender, and the extent to which the latent classes of the two genders correspond to each other. The latent classes of the two genders may bear no relation to each other, but this is unparsimonious and biologically implausible, since the subtypes in men and women would then represent different disorders. More interesting is the existence of two latent classes, one in each gender, which closely resemble each other. This can be interpreted as the occurrence of the same disorder in men and women. Models of this type are called 'homogeneous' (Clogg & Goodman, 1985), and in practice can be specified by imposing equality restrictions on the within class conditional probabilities of the corresponding latent classes in the two genders. If all the parameters of the latent classes in one gender are constrained to be equal to the corresponding parameters of the latent classes in the other gender, then the model is said to be 'totally homogeneous'. In the current study, a sequence of latent class models was fitted, ranging from one to three latent classes per gender, with and without the restrictions of total homogeneity on the within class conditional probabilities. The program MLLSA (maximum likelihood latent structure analysis) developed by Clogg (1977) was used throughout.

RESULTS

The number of patients with non-affective functional psychosis, for whom all the manifest variables had been recorded, was 447. The distributions of the eight chosen manifest variables are given in Table 1. Summary statistics for pairwise relationships are given in Table 2; with the size of association being measured by product moment (Pearson's) correlation coefficients, and the statistical significance assessed by χ^2 tests for 2×2 tables. Product moment correlations are easy to interpret, because they range from -1 (maximal negative correlation), through 0 (no correlation), to 1 (maximal positive correlation). Although the bounds do not reach -1 and 1 for binary data, the correlation still provides an indication of the sign and relative magnitude of the associations between the variables (Shil & Huang, 1992). Moreover, for binary variables the usual normality assumption is clearly violated, so that the χ^2 test gives a more accurate indication of statistical significance. The most significant positive associations were between male sex and poor pre-morbid social adjustment, early onset and poor pre-morbid social adjustment, and the most significant negative association was between early onset and paranoid delusions.

Considering all eight dichotomous manifest variables simultaneously, the total number of cells in the 8 dimensional contingency table is 2^8

Table 1. *The distributions of the manifest variables*

	Present (1)	Absent (0)
Family history (FH)	35	412
Restricted affect (RA)	48	399
Persecutory delusion (PD)	346	101
Poor social adjustment (SA)	165	282
Dysphoria (DS)	213	234
Early onset (EO)	166	281
Winter birth (WB)	196	251
Male sex (MS)	227	220

Table 2. *Pairwise product moment correlations of the manifest variables*

FH	1	—	—	—	—	—	—	—
RA	0·060	1	—	—	—	—	—	—
PD	0·038	−0·020	1	—	—	—	—	—
SA	0·053	0·169***	−0·086*	1	—	—	—	—
DS	−0·095*	0·045	0·044	0·069	1	—	—	—
EO	0·086	0·197***	−0·282****	0·237****	0·027	1	—	—
WB	0·078	−0·030	−0·008	−0·069	0·015	−0·017	1	—
MS	−0·046	0·183***	−0·050	0·215****	−0·145**	0·164***	0·022	1
	FH	RA	PD	SA	DS	EO	WB	MS

* $P < 0{\cdot}1$, ** $P < 0{\cdot}01$, *** $P < 0{\cdot}001$, **** $P < 0{\cdot}0001$ (χ^2 test of independence with Yate's correction).

Table 3. *Global goodness-of-fit test statistics for models* 1–6

Model	X^2	L^2	df
M1	465·18	325·85	247
M2	419·39	265·42	240
M3	264·18	222·99	238
M4	250·41	176·96	225
M5	245·34	204·51	230
M6	260·68	157·67	214

Table 4. *Parameter estimates for M4*

	Males		Females	
Model 4	$X = 1$	$X = 2$	$X = 3$	$X = 4$
$P(X)$	0·22	0·29	0·11	0·38
$P(\text{FH} = 1\|X)$	0·11	0·03	0·14	0·08
$P(\text{RA} = 1\|X)$	0·30	0·06	0·13	0·03
$P(\text{PD} = 1\|X)$	0·65	0·83	0·43	0·90
$P(\text{SA} = 1\|X)$	0·72	0·29	0·44	0·21
$P(\text{DS} = 1\|X)$	0·50	0·34	0·56	0·55
$P(\text{EO} = 1\|X)$	0·78	0·20	1·0	0·08
$P(\text{WB} = 1\|X)$	0·36	0·51	0·53	0·40
$P(\text{MS} = 1\|X)$	(1)	(1)	(0)	(0)

The latent categorical variable, X, takes values 1, 2, 3, 4, each representing a latent class. The parameters $P(X)$ are the probabilities of the different values of X, i.e. the probabilities of the latent classes. The parameters $P(\text{FH} = 1|X)$, $P(\text{RA} = 1|X)$, etc., are the conditional probabilities of the manifest variable taking a value of 1, given the value of the latent categorical variable, i.e. the within-class conditional probabilities of the manifest variables. The parameters in brackets, i.e. the conditional probabilities of being male within the latent classes, are fixed constants.

= 256. Each cell in the contingency table is described fully by a vector of 9 numbers, the first 8 indicating the categories of the manifest variables (henceforth called response vector), and the 9th being the number of individuals who fall into these categories, i.e. the cell counts. To these data, the following six models were fitted.

M1. One latent class per gender, with total homogeneity.

M2. One latent class per gender, totally unconstrained.

M3. Two latent classes per gender, with total homogeneity.

M4. Two latent classes per gender, totally unconstrained.

M5. Three latent classes per gender, with total homogeneity.

M6. Three latent classes per gender, totally unconstrained.

Two global measures of goodness-of-fit, the Pearson (X^2) and likelihood ratio (L^2) χ^2 statistics, and their associated degrees of freedom (df), were obtained for each model using MLLSA (Table 3). The definitions of these statistics are given in Appendix A. Both X^2 and L^2 were very large for M1, which was therefore excluded from further consideration. If M1 had fitted the data adequately, then that would imply the absence of any associations between the manifest variables, so that any further analyses would be futile. The clinical interpretation of M1, that there is one disorder identical in the two genders, is in any case contrary to the numerous previous demonstrations of gender differences (Castle & Murray, 1991).

The interpretation of M2 is that there are two subtypes, one of which occurs solely in men, and the other in women. In contrast, M3 suggests the existence of two subtypes which occur in men and women in different proportions. These two models differ only slightly in degrees of freedom (df) but enormously in both X^2 and L^2, with M 3 providing a much better fit to the data. We therefore excluded M2 from further consideration.

Table 5. *Parameter estimates for M5*

	Males			Females		
Model 5	$X=1$	$X=2$	$X=3$	$X=4$	$X=5$	$X=6$
$P(X)$	0·29	0·22	0·00	0·13	0·23	0·13
$P(\text{FH}=1\|X)$	0·10	0·07	0·01	0·10	0·07	0·01
$P(\text{RA}=1\|X)$	0·22	0·03	0·00	0·22	0·03	0·00
$P(\text{PD}=1\|X)$	0·61	0·88	0·93	0·61	0·88	0·93
$P(\text{SA}=1\|X)$	0·60	0·20	0·23	0·60	0·20	0·23
$P(\text{DS}=1\|X)$	0·50	0·31	0·98	0·50	0·31	0·98
$P(\text{EO}=1\|X)$	0·74	0·09	0·16	0·74	0·09	0·16
$P(\text{WB}=1\|X)$	0·41	0·51	0·28	0·41	0·51	0·28
$P(\text{MS}=1\|X)$	(1)	(1)	(1)	(0)	(0)	(0)

The latent categorical variable, X, takes values 1, 2, 3, 4, 5, 6 each representing a latent class. The parameters $P(X)$ are the probabilities of the different values of X, i.e. the probabilities of the latent classes. The parameters $P(\text{FH}=1|X)$, $P(\text{RA}=1|X)$, etc., are the conditional probabilities of the manifest variable taking a value of 1, given the value of the latent categorical variable, i.e. the within-class conditional probabilities of the manifest variables. The parameters in brackets, i.e. the conditional probabilities of being male within the latent classes, are fixed constants.

The next comparison is between M3 and M4. The interpretation of M4, like that of M2, is unattractive because it suggests that there is no relationship between the subtypes in men and women. Thus, if the fit of M4 is not significantly better than that of M3, one will be inclined to accept M3 or a model similar to M3. Here the evidence from X^2 and L^2 diverges; whereas the difference in X^2 (13·77) is not significant on 13 df, the difference in L^2 (46·03) is significant.

M5 postulates the existence of three subtypes which occur in both genders in different proportions. The parameter restrictions in M5 make it more parsimonious than M4 (by 5 df), so that a slightly worse goodness-of-fit is acceptable for M5. However, the evidence from X^2 and L^2 point to opposite directions; X^2 favours M5 while L^2 favours M4.

Finally, we consider M6, a model with three subtypes in men and three unrelated subtypes in women. Comparing M5 and M6, which differ by 16 df, the difference in X^2 is 15·34 in favour of M5, but the difference in L^2 is 46·84 in favour of M6. Clearly, these global goodness-of-fit statistics are not behaving appropriately.

The reason for the anomalous behaviour of X^2 and L^2 is that the contingency table is sparse. In the 256 cells of the contingency table, the average count is 1·7, and 145 cells are empty. Under these circumstances X^2 and L^2 cannot be expected to approximate the χ^2 distribution. The inaccuracy of the likelihood ratio χ^2 test in latent class analysis has been demonstrated by simulation (Everitt, 1988*a*). Thus, instead of relying solely on global test statistics, it is also desirable to examine expected cell counts and residuals. The five largest absolute residuals are 10·0, 6·7, 5·6, 5·4, and 4·1 for M3; 4·1, 3·9, 3·9, 3·8, and 3·5 for M4; 4·4, 4·3, 4·2, 4·0, and 3·8 for M5; and 4·2, 3·4, 3·2, 3·0, and 2·8 for M6. Clearly, the fit is poor for M3, but otherwise similar for M4, M5, and M6. A more accurate impression can perhaps be obtained by examining transformed residuals with approximate standard normality, such as the Freeman–Tukey residuals (defined in the Appendix). The five largest absolute Freeman–Tukey residuals are 2·97, 2·14, 2·13, 2·08, and 2·08 for M3; 2·44, 2·14, 2·06, 2·00, and 1·91 for M4; 2·24, 2·05, 1·93, 1·89, and 1·89 for M5; and 2·28, 2·09, 1·95, 1·82, and 1·74 for M6. Again the fit of M3 appears poorer than that of the three models. Considering that the degrees of freedom for M4, M5, and M6, which fit the data almost equally well, are 225, 230, and 214 respectively, one is inclined to take M6 no further because it is far less parsimonious than the other two models. From a statistical point of view there is little to choose between M4 and M5.

The parameter estimates of M4 and M5 are given in Tables 4 and 5 respectively. We should not interpret the exact values of these estimates too literally, since they are subject to errors. The parameter estimates of M4 (Table 4) suggest the existence of a subtype (classes 1 and 3) with a high frequency of positive family history, early onset, restricted affect and poor pre-morbid adjustment, and a male to female ratio of 2 to 1. Interestingly, within this type, more males (class 1) than females (class 3) have restricted affect

and poor pre-morbid adjustment, but more females than males have positive family history. The second possible subtype (classes 2 and 4) has a higher frequency of persecutory delusions, and a female excess. The two subtypes do not differ much with regard to dysphoria. Confusingly, winter birth appears to be more frequent in class 2 and 3 than in class 1 and 4, i.e. it appears to be more frequent in the males of the second subtype and the females of the first subtype.

M5 is easier to interpret than M4 because of the restrictions on its parameters (Table 5). First, there is a subtype (classes 1 and 4) with a high frequency of positive family history, early onset, restricted affect, and poor pre-morbid adjustment, and a male to female ratio of just over 2 to 1. This subtype is very similar to the first subtype in M4. The second subtype (classes 2 and 5) has a high frequency of persecutory delusions and winter birth, a low frequency of early onset, and a male to female ratio of about one. This subtype is similar to the second subtype in M4. The third subtype (classes 3 and 6) has a very high frequency of dysphoria and persecutory delusions, a very low frequency of family history of schizophrenia and restricted affect, and very few men.

DISCUSSION

The scope and limitations of latent class analysis

Before we interpret these results, let us examine the meaning of a class in the context of the method. Clearly, the definition of a class will be based on certain categorical variables, chosen according to the purpose of the classification. One possible criterion of a class is that all members of the class must be identical with regard to all the chosen variables. However, this often results in too many classes. A less stringent criterion is that all members of the class must share the same tendency to exhibit the chosen features, so that any differences between individuals are due to chance. Members of a class are characterized by the probabilities that they will exhibit the chosen features; probabilities which are identical for all members of the class. It follows that whether a member exhibits any feature does not affect the probability that he will exhibit any other feature. We have therefore, within any class, statistical independence of the variables. A sample of individuals in whom there are associations between variables does not constitute a class by our definition. Latent class analysis attempts to extract classes which satisfy the criterion of within-class independence, and which explain the associations between the variables in the entire sample.

The criterion of within-class independence in latent class analysis is more an axiom than an assumption (Bartholomew, 1987). What latent class analysis does assume is that the observed associations in a sample are caused by the stratification of the sample due to a latent categorical variable. This assumption is violated, for example, when the observed associations are caused by a latent continuous variable, or by direct causal relationships between the manifest variables. Even if a latent categorical variable exists and is responsible for the associations between the observed variables, it may have only limited biological significance.

A second limitation of latent class analysis is its exclusive use of discrete data. It is possible to generalize the method to include continuous variables of known distribution (Everitt, 1988*b*). However, in its basic form, as implemented by standard programs such as MLLSA, latent class analysis can deal with discrete variables only, and often this causes loss of information when some continuous variables (e.g. age) have to be converted to discrete variables for the analysis. A closely related problem is that the number of cells in a multidimensional contingency table increases very rapidly as a function of the number of variables. Thus, very large samples are often required to prevent the table from becoming too sparse, with the accompanying problems of imprecise parameter estimates and unreliable test statistics.

Strengths and weaknesses of the current analysis

The use of a first contact sample from a defined population means that our subjects were representative of the entire spectrum of clinical non-affective functional psychotic cases in the population. Furthermore, this precluded bias consequent upon hospitalization or chronicity. We included a very broad range of psychotic disorders, including schizoaffective disorder and

'atypical' psychoses. The failure to include manic or depressed patients with psychotic features to their illness could be criticized, and an analysis including such individuals would be of interest with respect to gender and genetic liability. However, the extension of the analyses to include such individuals was not feasible in this study.

Only eight variables were used in the analysis, which is the maximum number for MLLSA. However, we believe that these variables measure important aspects of the illness, including some which may have aetiological significance (Murray *et al.* 1985, 1992). Furthermore, they were reliably recorded, and few subjects had to be excluded because of missing data. Nevertheless, the sample size of 447 was too small to allow the reliable discrimination between different models, and the accurate estimation of parameters. This was the reason why no attempt was made to fit further models.

The interpretation of the latent classes

Latent class analysis helps us to find structure in multivariate categorical data. While structure certainly exists in our data, we should be cautious in claiming that the derived latent classes have any underlying aetiological, biological or clinical significance. A different choice of variables might have resulted in a different latent structure. Different models could not be clearly discriminated.

Some consistency did emerge, however, in that M4 and M5 were quite similar in their parameter estimates, except that M5 had an extra group of females with dysphoria and persecutory delusions. We therefore regard M4 as a special case of M5, and proceed to a tentative typological interpretation of M5.

(1) Classes 1 and 4 of M5 are characterized by early onset, poor pre-morbid adjustment, restricted affect, and male preponderance. We propose that these classes represent the clinical correlates of a 'neurodevelopmental' illness (Castle & Murray, 1991), which we denote as Type A. We would emphasize that none of the variables included in this analysis is a direct measure of structural brain abnormalities, and thus the 'neurodevelopmental' label is presumptive. However, some support for our use of the label comes from studies showing an association between poor pre-morbid functioning (and negative symptomatology) and structural brain abnormalities in schizophrenia (see Weinberger *et al.* 1980; Pearlson *et al.* 1985, 1989; Orel *et al.* 1991), as well as the fact that schizophrenic males (and expressly young males) appear most likely to show such abnormalities on neuroimaging investigations (see Andreasen, 1990; reviewed by Castle & Murray, 1991). To address this question definitively, a prospective study including measures of structural brain abnormalities would be required.

In contrast to the current findings, Goldstein *et al.* (1990) found that their early-onset group had a lower familial risk for schizophrenia and more winter births. However, other investigators who defined a similar severe early-onset form of illness showed such patients to be at greater genetic risk for schizophrenia (Tsuang & Winokur, 1974; Farmer *et al.* 1983). A genetic aetiology is obviously not inconsistent with a neurodevelopmental hypothesis; indeed, Jones & Murray (1991) have recently argued that 'the genetics of schizophrenia is the genetics of neurodevelopment'.

(2) Classes 2 and 5 of M5 are characterized by late onset, paranoid delusions, and an almost equal sex ratio. We propose that these classes represent a 'paranoid' illness, which we denote as Type B. This subtype is similar to the paranoid subtype of Tsuang & Winokur (1974), and to P cluster of Farmer *et al.* (1983). In line with these previous studies, our 'paranoid group' showed a lower familial risk for schizophrenia. The season of birth effect in this group is intriguing. In an extensive review of the literature, Bradbury & Miller (1985) found no consistent schizophrenic subtype to be more prone to the seasonality effect. However, Opler *et al.* (1984) and Takei *et al.* (1992) reported more winter birth effect in patients with later onset of illness.

(3) Classes 3 and 6 of M5 are characterized by dysphoria, paranoid delusions, a negligible familial risk of schizophrenia, and an absence of men. We propose that these classes represent a 'schizoaffective' subtype, which we denote as Type C. Previous attempts at subtyping schizophrenia have not identified such a group of patients. This may be due to patient selection: while we used a broad 'ICD-9' conception of schizophrenia, others (e.g. Tsuang & Winokur, 1974) have used restrictive criteria (e.g. those of Feighner *et al.* 1972) which could have excluded

such patients. The delineation of this subtype supports suggestions (e.g. Pope & Lipinsky, 1978; Castle & Murray, 1991) that some late-onset females who are diagnosed as schizophrenic have an illness closely related to affective disorder in aetiology.

The second subtype in the study of Goldstein *et al.* (1990), had a female preponderance, and showed later onset, more dysphoria and more persecutory delusions; thus, there are features of both our Type B (later onset, persecutory delusions) and C (later onset, female excess, persecutory delusions, dysphoria). However, Goldstein *et al.* (1990) found familial risk of schizophrenia, and non-winter birth more frequently in their second subtype. This is at odds with our findings, which are in turn more in line with previous studies in this area, as outlined above.

Since Types A and B are present in both M4 and M5, we can be more certain of their existence than that of Type C. If we remove Type C, then the current typology is similar to those of Tsuang & Winokur (1974) and Farmer *et al.* (1983). We have recently proposed that schizophrenia should be subdivided into 'neurodevelopmental' and 'adult onset' types (Murray *et al.* 1992). The results of this analysis are readily interpretable in that framework. However, it is also possible that Type A and Type B represent more and less severe forms of the same disorder.

We would like to stress that the proposed typology needs to be refined further by additional analyses on larger sets of data, ideally collected prospectively. Whether the typology has aetiological, biological and clinical significance needs to be established by studies examining the correlates of the subtypes.

We are grateful to Dr Jill Goldstein, Dr Anne Farmer and Professor Brian Everitt for advice, and the Wellcome Trust for financial support.

APPENDIX

Consider a contingency table whose cells are labelled by the subscript i. Let O_i be the observed count in cell i, and E_i be the count expected for that cell under some model. The Pearson chi-squared statistic, X^2, is defined by

$$X^2 = \sum_i \frac{(O_i - E_i)^2}{E_i},$$

and the likelihood ratio chi-squared statistic, L^2, is defined by

$$L^2 = 2\sum_i O_i \log \frac{O_i}{E_i}.$$

The behaviour of these statistics is erratic for sparse tables because of the problem of taking ratios of small numbers. Thus, small values of E_i can lead to a large contribution to the overall X^2 even if the residual $(O_i - E_i)$ is really quite small. For L^2 the problem is similar, but in addition if O_i is 0, then strictly speaking L^2 is undefined since the expression involves the logarithm of 0 (Bishop *et al.* 1975).

The Freeman–Tukey residuals, Z_i, are obtained by the transformation

$$Z_i = \sqrt{O_i} + \sqrt{(O_i + 1)} - \sqrt{(4E_i + 1)},$$

so that they are approximately normal with zero mean and unit variance (Freeman & Tukey, 1950).

REFERENCES

Andreasen, N. C., Swayze, V. W. II, Flaum, M., Yates, W. R., Arndt, S. & McChesney, C. (1990). Ventricular enlargement in schizophrenia evaluated with computed tomographic scanning. *Archives of General Psychiatry* **47**, 1008–1015.

Angermeyer, M. C. & Kuhn, L. (1988). Gender differences in age at onset of schizophrenia: an overview. *European Archives of Psychiatry and Neurological Sciences* **237**, 351–364.

Angermeyer, M. C., Goldstein, J. M. & Kuhn, L. (1989). Gender differences in schizophrenia: rehospitalization and community survival. *Psychological Medicine* **19**, 365–382.

Angermeyer, M. C., Kuhn, L. & Goldstein, J. M. (1990). Gender and the course of schizophrenia: differences in treated outcomes. *Schizophrenia Bulletin* **16**, 293–307.

Aylward, E., Walker, E. & Bettes, B. (1984). Intelligence and schizophrenia: meta-analysis of the research. *Schizophrenia Bulletin* **10**, 430–459.

Bartholomew, D. J. (1987). *Latent Variable Models and Factor Analysis*. Griffin: London.

Bellodi, L., Bussoleni, C., Scorza-Smeraldi, R., Grassi, G., Zacchetti, G. & Smeraldi, E. (1986). Family study of schizophrenia: exploratory analysis for relevant factors. *Schizophrenia Bulletin* **12**, 120–128.

Bishop, Y. V. V., Fienberg, S. E. & Holland, P. W. (1975). *Discrete Multivariate Analysis*, pp. 123–175. MIT Press: Cambridge, Mass.

Bradbury, T. N. & Miller, G. A. (1985). Season of birth in schizophrenia: a review of evidence, methodology, and aetiology. *Psychological Bulletin* **98**, 569–594.

Castle, D. J. & Murray, R. M. (1991). The neurodevelopmental basis of sex differences in schizophrenia. *Psychological Medicine* **21**, 565–575.

Castle, D. J., Wessely, S., Der, G. & Murray, R. M. (1991). The incidence of operationally defined schizophrenia in Camberwell, 1965 to 1984. *British Journal of Psychiatry* **159**, 790–794.

Clogg, C. C. (1977). *Unrestricted and Restricted Maximum Likelihood Latent Structure Analysis: A Manual for Users*. Population Issues Research Office: University Park, PA.

Clogg, C. C. & Goodman, L. A. (1985). Simultaneous latent structure analysis in several groups. In *Sociological Methodology* (ed. N. B. Tuma), pp. 81–110. Josey-Bass: San Francisco.

Dempster, A. P., Laird, N. M. & Rubin, D. N. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B* **39**, 1–38.

Done, D. J., Johnstone, E. C., Frith, C. D., Golding, J., Shepherd, P. M. & Crow, T. J. (1991). Complications of pregnancy and delivery in relation to psychosis in adult life: data from the British perinatal mortality survey sample. *British Medical Journal* **302**, 1576–1580.

Everitt, B. S. (1984). *An Introduction to Latent Variable Models.* Chapman & Hall: London.

Everitt, B. S. (1988*a*). A Monte Carlo investigation of the likelihood ratio test for number of classes in latent class analysis. *Multivariate Behavioural Research* **23**, 531–538.

Everitt, B. S. (1988*b*). A finite mixture model for the clustering of mixed-mode data. *Statistics and Probability Letters* **6**, 305–309.

Farmer, A. E., McGuffin, P. & Spitznagel, E. L. (1983). Heterogeneity in schizophrenia: a cluster-analytic approach. *Psychiatry Research* **8**, 1–12.

Feighner, J. P., Robins, E., Guze, S. B., Woodruff, R., Winokur, G. & Munoz, R. (1972). Diagnostic criteria for use in psychiatric research. *Archives of General Psychiatry* **26**, 57–000.

Foerster, A., Lewis, S. W., Owen, M. & Murray, R. M. (1991*a*). Low birth weight and a family history of schizophrenia predict poor premorbid functioning in psychosis. *Schizophrenia Research* **5**, 13–20.

Foerster, A., Lewis, S. W., Owen, M. & Murray, R. M. (1991*b*). Premorbid personality in psychosis: effects of sex and diagnosis. *British Journal of Psychiatry* **158**, 103–105.

Freeman, M. F. & Tukey, J. W. (1950). Transformations related to the angular and square root. *Annals of Mathematics and Statistics* **21**, 607–611.

Goldstein, J. M. & Link, B. G. (1988). Gender and the expression of schizophrenia. *Journal of Psychiatric Research* **22**, 141–155.

Goldstein, J. M., Tsuang, M. T. & Faraone, S. V. (1989). Gender and schizophrenia: implications for understanding the heterogeneity of the illness. *Psychiatry Research* **28**, 243–253.

Goldstein, J. M., Santangelo, S. L., Simpson, J. C. & Tsuang, M. T. (1990). The role of gender in identifying subtypes of schizophrenia: a latent class analytic approach. *Schizophrenia Bulletin* **16**, 263–275.

Green, B. F. (1951). A general solution of the latent class model of latent structure analysis and latent profile analysis. *Psychometrika* **16**, 151–166.

Haas, G. L., Sweeney, J. A., Keilp, J. & Frances, A. J. (1990). Sex differences of neurocognition of schizophrenia. Paper presented at American Psychiatric Association Annual Meeting, San Francisco.

Harris, M. J. & Jeste, D. V. (1988). Late-onset schizophrenia: an overview. *Schizophrenia Bulletin* **14**, 39–55.

Jones, P. B. & Murray, R. M. (1991). The genetics of schizophrenia is the genetics of neurodevelopment. *British Journal of Psychiatry* **158**, 615–623.

Kay, S. R., Opler, L. A. & Fiszbein, A. (1986). Significance of positive and negative syndromes in chronic schizophrenia. *British Journal of Psychiatry* **149**, 439–448.

Klorman, R., Strauss, J. S. & Kokes, R. F. (1977). Pre-morbid adjustment in schizophrenia. III. The relationship of demographic and diagnostic factors to measures of pre-morbid adjustment in schizophrenia. *Schizophrenia Bulletin* **3**, 214–225.

Lazarsfeld, P. L. & Henry, N. W. (1968). *Latent Structure Analysis.* Houghton Mifflin: Boston, Mass.

Lewine, R. J. (1981). Sex differences in schizophrenia. Timing or subtypes? *Psychological Bulletin* **90**, 432–444.

Lewine, R. J. (1988). Gender and schizophrenia. In *Handbook of Schizophrenia, Vol 3* (ed. H. A. Nasrallah), pp. 379–397. Elsevier: Amsterdam.

Lewine, R. J., Strauss, J. S. & Gift, T. E. (1981). Sex differences in age at first hospital admission for schizophrenia: fact or artifact? *American Journal of Psychiatry* **138**, 440–444.

Lewis, S. W. & Murray, R. M. (1987). Obstetric complications, neurodevelopmental deviance, and risk of schizophrenia. *Journal of Psychiatric Research* **21**, 413–421.

Loranger, A. W. (1984). Sex differences in age at onset of schizophrenia. *Archives of General Psychiatry* **41**, 157–161.

McCabe, M. S., Fowler, R. C., Cadoret, R. J. & Winokur, G. (1971). Familial differences in schizophrenia with good and poor prognosis. *Psychological Medicine* **1**, 326–332.

McGuffin, P., Farmer, A. E. & Harvey, I. (1991). A polydiagnostic application of operational criteria in psychotic illness: development and reliability of the OPCRIT system. *Archives of General Psychiatry* **48**, 764–770.

Murray, R. M., Lewis, S. & Reveley, A. M. (1985). Towards an aetiological classification of schizophrenia. *Lancet* **i**, 1023–1026.

Murray, R. M., O'Callaghan, E., Castle, D. J. & Lewis, S. W. (1992). A neurodevelopmental approach to the classification of schizophrenia. *Schizophrenia Bulletin* **18**, 319–332.

O'Callaghan, E., Gibson, T., Colohan, H. A., Buckley, P., Walshe, D. G., Larkin, C. & Waddington, J. L. (1992). Risk of schizophrenia in adults born after obstetric complications and their association with early onset of illness: a controlled study. *British Medical Journal* **305**, 1256–1259.

Offord, D. R. (1974). School performance of adult schizophrenics, their siblings and age mates. *British Journal of Psychiatry* **125**, 12–19.

Opler, L. A., Kay, S. R., Rosado, V. & Lindenmayer, J.-P. (1984). Positive and negative syndromes in chronic schizophrenic patients. *Journal of Nervous and Mental Disease* **172**, 317–325.

Orel, O., Cannon, T. D., Hollister, J. M., Mednick, S. A. & Parnas, J. (1991). Ventricular enlargement and premorbid deficits in school-occupational attainment. *Schizophrenia Research* **4**, 49–52.

Owen, M. J., Lewis, S. W. & Murray, R. M. (1988). Obstetric complications and schizophrenia: a computed tomographic study. *Psychological Medicine* **18**, 331–339.

Owen, M. J., Lewis, S. W. & Murray, R. M. (1989). Family history and cerebral ventricular enlargement in schizophrenia: a case-control study. *British Journal of Psychiatry* **154**, 629–634.

Pearlson, G. D., Garbacz, D. J., Moberg, P. J., Ahn, H. S. & de Paulo, J. R. (1985). Symptomatic, familial, perinatal, and social correlates of computer axial tomography (CAT) changes in schizophrenia and bipolars. *Journal of Nervous and Mental Disease* **173**, 42–50.

Pearlson, G. D., Kim, W. S., Kubos, K. L., Moberg, P. J., Jayaram, G., Bascom, M. J., Chase, G. A., Goldfinger, A. D. & Tune, L. E. (1989). Ventricle–brain ratio, computed tomographic density, and brain area in 50 schizophrenics. *Archives of General Psychiatry* **46**, 690–697.

Pope, H. G. & Lipinsky, J. F. (1978). Diagnosis in schizophrenia and manic depressive illness. *Archives of General Psychiatry* **35**, 811–828.

Pulver, A. E., Liang, K.-Y., Brown, C. H., Wolynieck, P., McGrath, J., Adler, L., Tam, D., Carpenter, W. T. & Childs, B. (1992). Risk factors in schizophrenia: season of birth, gender and familial risk. *British Journal of Psychiatry* **160**, 65–71.

Riecher, A., Maurer, K., Loffler, W., Fatkenheuer, B., van der Heiden, W. & Hafner, H. (1989). Schizophrenia – a disease of young single males? *European Archives of Psychiatry and Neurological Sciences* **239**, 210–212.

Seeman, M. V. (1986). Current outcome in schizophrenia: women vs men. *Acta Psychiatrica Scandinavica* **73**, 609–617.

Shepherd, M., Watt, D., Falloon, I. & Smeeton, N. (1989). *The Natural History of Schizophrenia: A 5-year Follow-up Study of Outcome and Prediction in a Representative Sample of Schizophrenics. Psychological Medicine Monograph 15.* Cambridge University Press: Cambridge.

Shil, W. J. & Huang, W. M. (1992). Evaluating correlation with proper bounds. *Biometrics* **48**, 1207–1213.

Shimizu, A., Kurachi, M., Noda, M., Yamaguchi, N., Torri, H. & Isaki, K. (1988). Influence of sex on age at onset of schizophrenia. *Japanese Journal of Psychiatry and Neurology* **42**, 35–40.

Takei, N., O'Callaghan, E., Sham, P. & Murray, R. M. (1992).

Winter birth excess in schizophrenia: its relationship to place of birth. *Schizophrenia Research* **6**, 102.

Tsuang, M. T. & Winoker, G. (1974). Criteria for subtyping schizophrenia: clinical differentiation of hebephrenic and paranoid schizophrenia. *Archives of General Psychiatry* **31**, 43–47.

Weinberger, D. R., Cannon-Spoor, E., Potkin, S. G. & Wyatt, R. J. (1980). Poor premorbid adjustment and CT scan abnormalities in chronic schizophrenia. *American Journal of Psychiatry* **137**, 1410–1413.

Wilcox, J. A. & Nasrallah, H. A. (1987). Perinatal distress and prognosis of psychotic illness. *Neuropsychobiology* **17**, 173–175.

Zigler, E. & Levine, J. (1973). Pre-morbid adjustment and paranoid–nonparanoid status in schizophrenia: a further investigation. *Journal of Abnormal Psychology* **82**, 189–199.

Zigler, E., Levine, J. & Zigler, B. (1977). Pre-morbid social competence and paranoid–non-paranoid status in female schizophrenic patients. *Journal of Nervous and Mental Disease* **164**, 333–339.