Heterogeneity in Schizophrenia: An Extended Replication of the Hebephrenic-like and Paranoid-like Subtypes

Julie Williams, Anne E. Farmer, Simon Wessely, David J. Castle, and Peter McGuffin

Received August 17, 1992; revised version received May 27, 1993; accepted June 24, 1993.

Abstract. Hebephrenic-like (H) and paranoid-like (P) subtypes of schizophrenia have previously been described by Farmer et al. (1983, 1984). The stability of this subtypology of schizophrenia was explored using multivariate statistical techniques on a large independent data set. Both a discriminant function analysis and an admixture analysis produced strong evidence for a bimodel distribution of scores consistent with the H-like and P-like subtypes.

Key Words. Diagnosis, paranoia, hebephrenia, discriminant function analysis, admixture analysis.

The diversity of signs and symptoms in schizophrenia has led to a prolific literature delineating schizophrenic subtypes. Since Bleuler's (1911/1950) monograph on dementia praecox, there has been a strong but not universal conviction that schizophrenia could be "dissected into its natural sub-divisions." Recent years have seen the emergence of a wide range of possible subtypes that can be viewed as stemming predominantly from either a theoretical perspective (e.g., Crow, 1980), the interaction of clinical and biological research (e.g., Kendler and Hays, 1982), or the multivariate analysis of schizophrenic symptomatology (e.g., Tsuang and Winokur, 1974; Farmer et al., 1983, 1984; Goldstein et al., 1990; Castle et al., in press). Subtypes arising out of the data-driven approach have the advantage of being less open to bias. Furthermore, if this form of subtypology can be shown to be stable across schizophrenic populations, then the argument that it reflects genuine underlying differences is considerably strengthened. To the authors' knowledge, this study is the first to attempt such a test of a data-driven subtypology.

The delineation of subtypes according to mainly positive (e.g., delusions and hallucinations) and mainly negative (e.g., amotivation and flattened affect) symptoms has achieved a certain popularity (Crow, 1980; Andreasen, 1985; Carpenter et al., 1988), although there seems to be less than complete consensus over the specific components that constitute these syndromes (McGuffin et al., 1987).
Both Crow (1980) and Murray et al. (1988) have proposed different etiologies underlying positive and negative symptom subtypes. Crow (1980) has suggested that a predominantly positive symptom subtype (Type I schizophrenia) is due to dopaminergic dysfunction while a negative symptom subtype (Type II) is caused by structural brain damage. Murray et al. (1988), however, have hypothesized that early (neurodevelopmental) brain damage leads to negative symptoms, which precede the development of positive symptoms by some years.

In contrast, Wing (1988) has expressed the view that the symptom profile is environmentally determined. In an overstimulating environment, a schizophrenic patient tends to exhibit more positive symptoms such as delusions and hallucinations; in an impoverished or understimulated setting, negative symptoms tend to dominate the clinical picture.

The interaction of cognitive efficiency and neurological dysfunction with schizophrenic symptomatology has also been used to differentiate subgroups within schizophrenia. Liddle (1987) and Liddle and Barnes (1990) have examined chronic schizophrenic patients according to the presence of specific factors or syndromes, such as psychomotor poverty (which includes poverty of speech and flatness of affect), disorganization (including thought disorder, inappropriate affect, and reality distortion [which includes delusions and hallucinations]). They also reported a tendency for different syndromes to be associated with length of illness such that reality distortion symptoms were more severe early in the illness and psychomotor symptoms were more pronounced later in the illness.

During the past three decades, a number of attempts have been made to apply multivariate statistical techniques to the analysis of schizophrenic subtypes. Many have used cluster or factor analysis (Lorr et al., 1963; Bartko et al., 1981), producing between four and six possible subtypes that appear to lack stability across data sets. An alternative approach has been to use discriminant function analysis to examine whether clinically derived subtypes of schizophrenia could be successfully distinguished. Tsuang and Winokur (1974) produced operational definitions of hebephrenic and paranoid subtypes, and they proposed that the two subtypes could be reliably separated. They also proposed an undifferentiated category for patients who failed to meet criteria for either subtype.

Farmer et al. (1983, 1984) used cluster analysis to explore the presence of subtypes in schizophrenic populations. In an attempt to improve on the methodology in previous multivariate statistical approaches, two mathematically different methods of cluster analysis were applied and the level of agreement between them for the presence of two clusters was high. To explain whether this good reliability arose by chance, a Monte Carlo approach was taken and 100 replications of the original sample were generated, but with a random distribution of symptoms. Analysis of the large data set of “synthetic patients” with both cluster methods indicated that the reliability of the clusters obtained on the simulation data was very much poorer than that of the real sample. This result strongly suggested that the apparent structure in the real data set was not a chance finding. Farmer and colleagues went on to examine subtype correspondence, or homotypia, in pairs of affected relatives, and again this was significantly better than chance. Finally, discriminant analysis was used (Farmer et al., 1984) to refine the subtype and derive a weighted list of symptoms that best
distinguished between the two main clusters. A paranoid-like (P) type was characterized by poor premorbid social adjustment, well-organized delusions, and late age of onset (after 25), while a hebephrenic-like (H) type had high loadings on poor premorbid work adjustment, early onset, blunted affect, third person auditory hallucinations, thought insertion, and poor insight. A series of twin pairs, ascertained through schizophrenic probands (Gottesman and Shields, 1972) was subtyped, and again a statistically significant trend toward homotypia was found in the pairs who were both schizophrenic. However, the tendency for like to go with like regarding H and P subtypes was incomplete even in monozygotic pairs (Farmer et al., 1984), indicating that subtypes were quantitatively and not qualitatively distinct.

The use of multivariate analysis in the classification of psychiatric illness has a long history (see Kendell, 1975). Although a variety of subgroups has been reported, attempts to distinguish those that reflect a genuine difference are not straightforward. Moran (1966) and Kendell (1975) have both suggested that “a bimodal distribution of scores on a discriminant function, obtained from an unselected population and cross-validated on a second population, should be the accepted criterion of validity for all diagnostic distinctions” (Kendell, 1975, p. 115). The aim of the present study was to attempt the second of these objectives and cross-validate the Farmer subtypology on a large fully independent sample of psychotic patients.

Methods

Subjects. Psychotic patients (n = 360) who fulfilled DSM-III-R criteria (American Psychiatric Association, 1987) for schizophrenia, schizophreniform disorder, atypical psychosis, and delusional disorder were derived from the Camberwell Cumulative Psychiatric Case Register (Wing and Hailey, 1972). The Camberwell Register provides a comprehensive list of all persons from a defined catchment area of South London who had their first contact with psychiatric services between 1965 and 1984. The subjects were selected for an independent research study of the incidence of schizophrenia in Camberwell between 1965 and 1984 (Castle et al., 1991). The detailed case records of all subjects were rated by the London researchers S.W. and D.C. on the Operational Criteria Checklist for Psychotic Illness (OPCRIT).

Procedure. OPCRIT (McGuffin et al., 1991) consists of 74 items relating to the operational definitions of signs and symptoms required to produce diagnoses according to the following major classificatory systems: DSM-III (American Psychiatric Association, 1980), DSM-III-R (American Psychiatric Association, 1987), and five sets of research diagnostic criteria (Schneider, 1959; Feighner et al., 1972; Carpenter et al., 1973; Spitzer et al., 1975; Taylor and Abrams, 1978). The checklist information is entered into the OPCRIT (2.5) computer-scoring program, which produces a computer file of diagnoses and a data file of raw checklist scores. These computer data files were made available to the Cardiff group (A.E.F., J.W., and P.McG.), who carried out further analyses of the data independently. Individual item rating and other diagnostic issues relating to the OPCRIT rating of the Camberwell subjects were not discussed between the London and Cardiff researchers.

Interrater reliability between D.C. and S.W. for their OPCRIT ratings was assessed on a randomly selected subset of 50 case records. A κ score of 0.82 was calculated for diagnoses based on Research Diagnostic Criteria (RDC; Spitzer et al., 1975) and a κ score of 0.72 for DSM-III diagnoses (American Psychiatric Association, 1980).

For the purposes of the study, all ratings were converted to the categories of present (1) or absent (0), with sex scored as male = 0 and female = 1, age of onset < age 25 = 0, and age of onset ≥ 25 = 1. The OPCRIT program calculated a Farmer subtype score based on the discrimination rule (see Table 1 for discriminating variables) produced by the previous
Table 1. Canonical discriminant function coefficients for the hebephrenic-like (H) and paranoid-like (P) subtypes

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poor premorbid work adjustment</td>
<td>5.79</td>
</tr>
<tr>
<td>Previous personality disorder</td>
<td>1.87</td>
</tr>
<tr>
<td>Third person auditory hallucinations</td>
<td>1.82</td>
</tr>
<tr>
<td>Thought insertion</td>
<td>1.58</td>
</tr>
<tr>
<td>Blunted affect</td>
<td>0.97</td>
</tr>
<tr>
<td>Lack of insight</td>
<td>0.81</td>
</tr>
<tr>
<td>Unemployed at first psychiatric illness</td>
<td>0.79</td>
</tr>
<tr>
<td>Age of onset under 25</td>
<td>0.70</td>
</tr>
<tr>
<td>Poor premorbid social adjustment</td>
<td>-2.04</td>
</tr>
<tr>
<td>Well-organized delusions</td>
<td>-1.22</td>
</tr>
</tbody>
</table>

1. Higher scores are indicative of H type and lower scores of P type.

independent discriminant analysis of the original Farmer clusters (Farmer et al., 1984). The Farmer score (S) was calculated from the following formula, where X denoted the variable (i.e., X = 1 if present and X = 0 if absent) and W denoted the weighting given to that variable (Fig. 1 presents a frequency histogram of the Farmer raw score):

\[
S = \left( \sum_{i=1}^{N} X_i \times W_i \right) - 2.07
\]

The program then assigned these scores to their respective subtype according to the following rule: H type = S > 1.3 and P type = S ≤ 1.3. For the purposes of this study, a program was written to extract the raw Farmer scores as well as the subtype category from OPCRIT (copies supplied on request).

**Discriminant Function Analysis.** In line with the previous study (Farmer et al., 1983), the Wilks λ method of stepwise discriminant function analysis as found in the DISCRIMINANT program in the SPSS PC package (Norusis, 1986) was used, and an F to enter of 2 was specified. The data comprised ratings of the first 45 OPCRIT (see Appendix) items, along with the H or P subtype, for each of the 360 subjects. These OPCRIT items included all psychotic symptoms, together with information about premorbid adjustment, personal history, and family history.

**Admixture Analysis.** Admixture analysis is an extremely useful technique for detecting the number of individual distributions that can best describe a data set (Everitt, 1981). It was therefore used to explore the number of distributions that could best describe the distribution of Farmer subtype scores of the 360 subjects in the data set. It uses a likelihood method (Edwards, 1972) in which support for the competing hypotheses of one or more distributions can be evaluated using likelihood ratio tests. The admixture analysis was performed with the FORTRAN program SKUMIX (MacLean et al., 1976; Morton, 1982).

The significance of models postulating different numbers of distributions can be compared since minus twice the difference in log likelihoods asymptotically has a χ² distribution. The degrees of freedom are taken as the difference in the number of parameters between the models.
Results

The stepwise discriminant function analysis of the 360 subjects previously assigned to the H or the P subtype terminated after the addition of the 12th OPCRIT variable. The result of the canonical discriminant function was: canonical correlation = 0.88, Wilks λ = 0.22 and χ² = 535.92 (df = 12, p < 0.00001), Eigenvalue = 3.58. Of the grouped cases, 92.5% were correctly classified (i.e., P type = 100%, n = 184; H type = 84.7%, n = 176). This indicates that the error rate of the discriminatory rule originally derived by Farmer et al. (1984) is 7.5%. Indeed, the method used here and by Farmer et al. (1983, 1984) was recommended by Hand (1985), who stated that one important measure of the quality of a classification rule is the proportion of objects that it will misclassify in the future. On closer examination, the errors were found to range between -0.0736 to -0.0704 with a mean of -0.4282, which translates to an error
range of between 0.2491 to -0.2193, with a mean of -0.0571 on the original Farmer raw discriminatory scores. Table 2 shows the discriminatory coefficients of the 12 variables with an $F$ to enter ratio $> 2$. Also shown is the coefficient association to the H and P subtype. Fig. 2 shows a stacked histogram of discriminant scores based on the weightings as dictated by the 12 discriminator coefficients, where P type is indicated by 1 and H type by 2.

**Table 2. Canonical discriminant function coefficients presented in stepwise order**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>*Poor premorbid work adjustment</td>
<td>3.95</td>
</tr>
<tr>
<td>*Third person auditory hallucinations</td>
<td>0.95</td>
</tr>
<tr>
<td>*Well-organized delusions</td>
<td>-0.84</td>
</tr>
<tr>
<td>*Thought insertion</td>
<td>1.21</td>
</tr>
<tr>
<td>*Unemployed at first psychiatric illness</td>
<td>0.53</td>
</tr>
<tr>
<td>*Poor premorbid social adjustment</td>
<td>-0.33</td>
</tr>
<tr>
<td>Affective symptoms predominate</td>
<td>-0.40</td>
</tr>
<tr>
<td>Bizarre delusions</td>
<td>-0.25</td>
</tr>
<tr>
<td>Negative formal thought disorder</td>
<td>-0.45</td>
</tr>
<tr>
<td>*Lack of insight</td>
<td>0.49</td>
</tr>
<tr>
<td>Incoherence</td>
<td>0.50</td>
</tr>
<tr>
<td>*Passive delusions</td>
<td>-0.23</td>
</tr>
<tr>
<td>Constant</td>
<td>-2.34</td>
</tr>
</tbody>
</table>

*Variable found in Farmer et al. (1984) analysis.

**Fig. 2. Subtypes in a stacked histogram of canonical discriminant function scores**
To determine the number of distributions that best describe these data, an admixture analysis was performed on standardized Farmer raw scores produced where the original discriminant function of Farmer et al. (1984) was applied to the data set (see Table 3). A comparison of the untransformed and transformed single distribution models revealed that the former was significantly skewed ($\chi^2 = 24.72, df = 1, p < 0.001$). Since skewness can bias this statistic, only data that have been power-transformed to remove skew were examined. These results showed that although both the two- and three-distribution models produced a significantly better fit to the data than the single distribution model, the three-distribution model did not significantly differ from the two-distribution model. Therefore, parsimony would dictate that the two-distribution model offered the best explanation of the data.

Finally, the proportion of men and women in the H and P subtype groups was examined. The P type group was found to have equal numbers of both sexes, whereas the H type group contained slightly more men than women (54:46).

**Table 3. Admixture analysis of Farmer subtype data**

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Untransformed</th>
<th>Power transformed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 2 3</td>
<td>1 2 3</td>
</tr>
<tr>
<td><strong>First distribution</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variance</td>
<td>0.996 0.233 0.217</td>
<td>0.976 0.237 0.218</td>
</tr>
<tr>
<td>Mean</td>
<td>0.000 0.000 0.000</td>
<td>-0.143 -0.044 -0.043</td>
</tr>
<tr>
<td>Proportion accounted for by first distribution</td>
<td>(1.000) 0.373 0.670</td>
<td>(1.000) 0.378 0.749</td>
</tr>
<tr>
<td><strong>Second distribution</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>1.806 0.987 0.970</td>
<td>1.760 0.716</td>
</tr>
<tr>
<td>Proportion accounted for by second distribution</td>
<td>0.475 0.380 0.470</td>
<td>0.470 0.376</td>
</tr>
<tr>
<td><strong>Third distribution</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>1.875</td>
<td>1.840</td>
</tr>
<tr>
<td>Proportion accounted for by third distribution</td>
<td>0.255</td>
<td>0.251</td>
</tr>
<tr>
<td>Value of $p$ in the power transformation</td>
<td>(1.000 1.000 1.000)</td>
<td>-0.76 0.47 0.43</td>
</tr>
<tr>
<td>$\chi^2$ log likelihood + constant</td>
<td>1020.6 916.2 915.6</td>
<td>995.8 913.2 912.2</td>
</tr>
<tr>
<td>$\chi^2$ between distributions</td>
<td>1 vs.2 2 vs.3 1 vs.3</td>
<td>1 vs.2 2 vs.3 1 vs.3</td>
</tr>
<tr>
<td></td>
<td>104.4 0.6 105 82.6 1 83.6</td>
<td></td>
</tr>
<tr>
<td>Degrees of freedom</td>
<td>2 1 3</td>
<td>2 1 3</td>
</tr>
</tbody>
</table>

*Note. Parentheses denote fixed parameters.*

1. $p < 0.001$.

**Discussion**

The results provide strong evidence for the presence of two distinct H and P subgroups within the schizophrenic population. Support stems from two findings.
The first is the replication of the original bimodal distribution of scores (Farmer et al., 1984) obtained from a discriminant function analysis performed on a totally independent sample. The second is the further substantiation of the bimodality of psychotic individuals provided by the admixture analysis of the Farmer raw scores, produced by applying the original discriminating rule to these independent data.

Although it is of secondary importance, it was interesting to compare the variables that discriminated between the H and P subtypes both here and in the original analysis (Farmer et al., 1984). A high degree of similarity was found. This stability of subtyping was most pronounced for the variables of poor premorbid work adjustment, thought insertion, third person auditory hallucinations, unemployment, and lack of insight in the H subtype and for poor premorbid social adjustment and well-organized delusions in the P subtype. Three of the discriminating factors from the original Farmer et al. (1984) study failed to discriminate subtypes in this study: previous personality disorder, blunted affect, and early age of onset. Only incoherence, a variable of small effect, was added to those discriminating the H subtype. Four additional variables of minor effect were found to discriminate the P subtype (i.e., predominant affective symptoms, negative formal thought disorder, and passive and bizarre delusions). The degree of similarity in subtyping discriminated was especially interesting in the light of the fact that the Camberwell data set included a broad range of psychotic patients, whereas the original subtyping was based on schizophrenic patients alone, with a minimum of 5 years' duration of illness.

A brief comparison between the H and P subtypes and other subtypologies showed a number of similarities between the variables used to characterize them (e.g., unemployment, age of onset, and well-organized delusions). This seems to be especially true of those subtypes derived using predominantly multivariate statistical techniques (e.g., Tsuang and Winokur, 1972) in contrast to those subtypes based upon mainly theoretical distinctions (e.g., Murray et al., 1985). Although sex has been considered an important factor in distinguishing subgroups (e.g., Goldstein et al., 1990; Castle et al., in press), it played no part in discriminating the H and P subtypes.

The question remains as to whether the Farmer subtypes represent two qualitatively distinct disorders with correspondingly different etiologies or two severity classes assuming a multifactorial, mixed, or polygenic threshold model of schizophrenia (Reich et al., 1972; Farmer et al., 1984). The models assume that a number of genes act additively or interactively (Falconer, 1965) with each other and with the environment to contribute to a liability to the disorder that is normally distributed within the population. As liability increases, it is assumed to pass a number of thresholds, first for the disease itself and then for various degrees of severity of this disease. It could be argued that the H and P subtypes (see Fig. 3) reflect a difference of severity rather than of absolute type between the two groups. In general terms, evidence from twin studies and cluster analysis (Farmer et al., 1983, 1984) suggests this to be the most acceptable perspective from which to view the differences in schizophrenic subtyping. For example, Farmer et al. (1984) found that rates of both the H and P subtypes of schizophrenia were higher in cotwins of monozygotic and dizygotic probands who had H type rather than P type
schizophrenia, indicating that the H and P subtypes are different severity classes of the same disease rather than two separate diseases.

Determining the validity of schizophrenic subtypes is not straightforward. In the absence of any strong validating factors such as knowledge of genetic predispositions to schizophrenia, other less powerful validating variables have been used. The basic strategy has been to examine the relationship between subtypes and abnormalities in or features of such variables as cognitive performance (e.g., Alm et al., 1984), neuropsychology (e.g., Romani et al., 1987), neurology (e.g., Woods et al., 1987), electrophysiology (e.g., Kendler and Hays, 1982; Schwartzkopf et al., 1988), brain morphology (e.g., Alm et al., 1984; Harvey et al., 1990), biochemistry (e.g., Rudduck et al., 1984), homotypian twin and family studies (e.g., Farmer et al., 1984), perinatal complications (e.g., Reveley et al., 1984; Lewis and Murray, 1987; Schwartzkopf et al., 1989), detailed symptomatology (e.g., Nimgaonkar et al., 1988; Goldstein et al., 1990), and demography (e.g., Shur, 1982). The validity of the H and P types needs to be pursued, and to this end, it is an intention to examine homotypia and the patterns of H and P segregation within a large set of multiply affected families.

References


Appendix. Operational criteria checklist for psychiatric illness

Items included in analysis:
1. Sex
2. Age of onset
3. Single
4. Unemployed
5. Duration 2 weeks
6. Duration 6 months
7. Prodromal acute and residual 6 months
8. Poor work adjustment
9. Poor social adjustment
10. Preexisting personality disorder
11. Alcohol/Drug abuse within 12 months of onset
12. Family history of schizophrenia
13. Family history of other disorder

Behavior
14. Bizarre
15. Catatonia

Speech
16. Difficult to understand
17. Incoherent
18. Positive formal disorder
19. Negative formal disorder

Mood
20. Affective symptoms prominent
21. Restricted affect
22. Blunted affect
23. Inappropriate affect
24. Rapport difficult

Beliefs, experiences, perceptions
25. Persecutory delusions
26. Well-organized delusions
27. Grandiose delusions
28. Delusions of influence
29. Bizarre delusions
30. Widespread delusions
31. Delusions of passivity
32. Primary delusional perception
33. Other primary delusions
34. Delusions accompanied by hallucinations for 1 week
35. Persecutory or jealous delusions and hallucinations
36. Thought insertion
37. Thought withdrawal
38. Thought broadcast
39. Thought echo
40. Third person voices
41. Running commentary voices
42. Abusive persecutory accusatory voices
43. Other nonaffective hallucinations
44. Information not credible
45. Lack of insight