

Published in final edited form as:

*Schizophr Res.* 2011 May ; 128(1-3): 143–149. doi:10.1016/j.schres.2011.02.006.

## Paternal age related schizophrenia (PARS): latent subgroups detected by *k*-means clustering analysis

Hyejoo Lee<sup>a</sup>, Dolores Malaspina<sup>b</sup>, Hongshik Ahn<sup>a</sup>, Mary Perrin<sup>b</sup>, Mark G. Opler<sup>b</sup>, Karine Kleinhaus<sup>b</sup>, Susan Harlap<sup>b</sup>, Raymond Goetz<sup>b,c</sup>, and Daniel Antonius<sup>b</sup>

<sup>a</sup> Department of Applied Mathematics and Statistics, Stony Brook University, Stony Brook, NY, USA

<sup>b</sup> Institute for Social and Psychiatric Initiatives (InSPIRES), Department of Psychiatry, New York University School of Medicine, New York, NY, USA

<sup>c</sup> Department of Psychiatry, Columbia University, New York State Psychiatric Institute, New York, NY, USA

### Abstract

**Background**—Paternal age related schizophrenia (PARS) has been proposed as a subgroup of schizophrenia with distinct etiology, pathophysiology and symptoms. This study uses a *k*-means clustering analysis approach to generate hypotheses about differences between PARS and other cases of schizophrenia.

**Methods**—We studied PARS (operationally defined as not having any family history of schizophrenia among first and second-degree relatives and fathers' age at birth  $\geq 35$  years) in a series of schizophrenia cases recruited from a research unit. Data were available on demographic variables, symptoms (Positive and Negative Syndrome Scale; PANSS), cognitive tests (Wechsler Adult Intelligence Scale—Revised; WAIS-R) and olfaction (University of Pennsylvania Smell Identification Test; UPSIT). We conducted a series of *k*-means clustering analyses to identify clusters of cases containing high concentrations of PARS.

**Results**—Two analyses generated clusters with high concentrations of PARS cases. The first analysis (N=136; PARS=34) revealed a cluster containing 83% PARS cases, in which the patients showed a significant discrepancy between verbal and performance intelligence. The mean paternal and maternal ages were 41 and 33, respectively. The second analysis (N=123; PARS=30) revealed a cluster containing 71% PARS cases, of which 93% were females; the mean age of onset of psychosis, at 17.2, was significantly early.

**Conclusions**—These results strengthen the evidence that PARS cases differ from other patients with schizophrenia. Hypothesis-generating findings suggest that features of PARS may include a

---

Correspondence: Daniel Antonius, PhD, New York University School of Medicine, Department of Psychiatry, 550 First Avenue, NBV 22N10, New York, NY 10016, USA, Phone: 212-263-2760, Fax: 212-263-2935, daniel.antonius@nyumc.org.

#### Contributors

Drs. Malaspina and Goetz were involved in the design and writing of the study protocol. Ms. Lee and Drs. Ahn, Harlap, Goetz and Antonius managed literature reviews and statistical analyses pertaining to the study. Ms. Lee and Drs. Malaspina, Ahn, Perrin, Opler, Kleinhaus, Harlap, Goetz and Antonius were involved in the writing of various drafts and the final manuscript. All authors contributed to and have approved the final manuscript.

#### Conflict of Interest

All authors declare that they have no conflicts of interest.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

discrepancy between verbal and performance intelligence, and in females, an early age of onset. These findings provide a rationale for separating these phenotypes from others in future clinical, genetic and pathophysiologic studies of schizophrenia and in considering responses to treatment.

### Keywords

cluster analysis; *k*-means clustering; paternal age related schizophrenia; PARS; schizophrenia

---

### Introduction

The schizophrenias are characterized by significant heterogeneity in symptoms, course of illness, and clinical profiles (Tsuang et al., 1990). This heterogeneity complicates the interpretation of research findings and inhibits the discovery of novel treatments for the disorder. Some of the variability in symptoms and illness features among schizophrenia patients may be explained by the presence of latent subgroups that differ in etiology and key neurobiological underpinnings. Identifying these subgroups is important to set the stage for targeted person-specific pharmacological and/or psychological treatments (Jindal et al., 2005).

Advanced paternal age has been associated with the risk for schizophrenia in cohort studies in Israel (Malaspina et al., 2001; Brown et al., 2002), Denmark, (Byrne et al., 2003), Sweden (Zammit et al., 2003; Sipos et al., 2004), Japan (Tsuchiya et al., 2005), and the United States (Torrey et al., 2009). In the Israeli study, a quarter of the risk for schizophrenia was attributable to paternal age and the risk in offspring of fathers aged 50+ at birth was three-fold that of children whose fathers were younger than 25 at birth (Malaspina et al., 2001). Clinical studies have suggested that paternal age related schizophrenia (PARS) may be a specific variant of the disease, as symptom and cognitive profiles, regional cerebral metabolism, sex effects, and heart rate variability have been shown to differ from those of other cases (Malaspina, 2001; Malaspina et al., 2001; 2002a; 2005; Rosenfield et al., 2010; Antonius et al., 2010). If these studies are confirmed, then PARS may account for a substantial portion of the disease in clinical treatment. The confirmation of PARS as a separate group could lead to better understanding of its etiology and pathogenesis, and allow targeted modes of treatment.

Currently, however, it is not clear whether PARS explains any of the heterogeneity of schizophrenia. To explore this, we have chosen to use an approach based on clustering analysis in order to generate new hypotheses related to PARS. Among various clustering techniques, *k*-means clustering (MacQueen, 1967) has been favored over others (e.g. Yeung et al., 2001; Gibbons & Roth, 2002), and has been used in schizophrenia research (Richards et al., 2008). *k*-means clustering is a partitioning method often used in data mining and machine learning (Huang, 1998; Wagstaff et al., 2001). It aims to partition, or minimize, the average squared distance between *n* observations and a cluster centroid, such that each observation is assigned to the cluster with the nearest mean (Hand & Heard, 2005).

In schizophrenia research, *k*-means clustering has previously been used to examine the heterogeneity of psychosis symptoms (Mohr et al., 2004), antipsychotic responses (Garver et al., 2000), prognostic features (Jonsson & Jonsson, 1992), and cognitive symptoms (Silver & Shmoish, 2008; Bell et al., 2010). This clustering approach, however, has not, to our knowledge, been used to examine the distinctiveness of the PARS subgroup.

## Methods

### Participants

This study relies on cases with schizophrenia or schizoaffective disorder recruited at the New York State Psychiatric Institute (NYSPI) Schizophrenia Research Unit (SRU) in 1992–2007. The study was approved by the Institutional Review Board at NYSPi and all patients provided written informed consent.

For our analyses we were interested in a set of core factors consisting of demographic, clinical and cognitive variables. Thus, we included in our analyses only cases on whom we had the following variables: age of onset of psychosis, sex, family history of schizophrenia, age of the father at the case's birth (paternal age), diagnosis, severity of psychopathological symptoms, and neuropsychological function. The specific measures are described below. We operationally defined PARS as the absence of any family history of schizophrenia among first- and second-degree relatives and for cases whose fathers' age at birth was >35 years; all other cases were considered non-PARS (based on Malaspina et al., 2002b). All cases were taking medication at the time of assessment.

### Measures

Diagnosis was obtained using the Diagnostic Interview for Genetic Studies (DIGS; Nurnberger et al., 1994). Severity of psychopathology was assessed with the Positive and Negative Syndrome Scale (PANSS; Kay et al., 1987), which originally was divided into three subscales: positive, negative and general psychopathology (standard model). More recent research using empirical approaches suggests five factors: positive, negative, activation, dysphoric mood and autistic preoccupation (five-factor model; White et al., 1997). We included both approaches in our analyses.

The Wechsler Adult Intelligence Scale--Revised (WAIS-R; Wechsler, 1981) was used to obtain the following neuropsychological factors: full scale intelligence quotient (FIQ), verbal IQ (VIQ) and performance IQ (PIQ), as well as the verbal subtests (arithmetic, digit span, information, vocabulary, comprehension, similarities) and the performance subtests (object assembly, picture arrangement, picture completion, digit symbol, block design). We also obtained a verbal-performance differential score (VIQ-PIQ).

In addition to these core measures, we also assessed for the presence of trait markers by evaluating the deficit syndrome (enduring and primary negative symptoms of schizophrenia; Carpenter et al., 1988) with the Schedule for the Deficit Syndrome (SDS; Kirkpatrick et al., 1989), as well as odor identification (Kopala et al., 2001) with the University of Pennsylvania Smell Identification Test (UPSIT; Doty et al., 1984). The UPSIT is a standardized, multiple-choice, scratch-and-sniff test with the maximum score of 40 (perfect identification) that is found to be stable and reliably measured in schizophrenia patients (Malaspina et al., 1994).

### Analysis

Following MacQueen's (1967) *k*-means methodology, we used an algorithm in which each item is assigned to the cluster having the nearest centroid (mean). This nonhierarchical method initially takes the number of components of the population equal to the final required number of clusters. The final required number of clusters is chosen such that the points are mutually farthest apart. Next, it examines each component in the population and assigns it to one of the clusters depending on the minimum distance. The centroid's position is recalculated every time a component is added to the cluster and this continues until all the components are grouped into the final required number of clusters. The process is composed

of the following three steps: 1) partition the items into  $k$  initial clusters; 2) proceed through the list of items, assigning an item to the cluster whose centroid is nearest (we used Euclidean distance as the measure of distance). Then, recalculate the centroid for the cluster receiving the new item and for the cluster losing the item; 3) repeat Step 2 until no more reassignment takes place.

The above described  $k$ -means clustering algorithm was run using various combinations of variables in order to identify latent subgroups of PARS. The clusters were generated using the following core variables: age of onset of psychosis, sex (males = 0; females = 1), family history (no family history of schizophrenia = 0; family history of schizophrenia = 1) and paternal age (age of the father at the case's birth). Together with these variables we added one of two sets of variables, or a combination of them. These two sets of variables were: the WAIS-R FIQ, PIQ, VIQ, VIQ-PIQ and the verbal and performance subtests; and the PANSS scores from the standard model (positive, negative, and general psychopathology subscale scores).

The analysis was performed with SAS version 9.2 (SAS Institute, Cary, NC) proc fastclus. In contrast to other clustering methods (e.g. hierarchical clustering) there is no standard way to find the optimal number of clusters; thus we ran numerous analyses with various values of  $k$  (from  $k=2$  to  $k=12$  for the above described two sets of variables), with the goal of finding clusters with high concentrations of PARS subjects.

## Results

Two of our  $k$ -means clustering analyses produced clusters with high PARS concentration. Each of the two clustering analyses generated seven clusters ( $k=7$ ) and yielded some prominent features related to the PARS subjects. The first analysis included the 11 WAIS-R subtests in addition to the four core demographic variables (age of onset of psychosis, sex, paternal age, family history of schizophrenia). The second analysis included the VIQ-PIQ variable and the PANSS factors from the standard model (positive, negative and general psychopathology symptoms) together with the same four core demographic variables used in the first analysis. The descriptive data for the demographic, clinical and neuropsychological variables from these two analyses are presented in tables 1 and 2.

For each variable included in the clustering analysis, we conducted a two-sample  $t$ -test for continuous variables and a chi-square test for categorical variables for a comparison between a specific cluster and the rest of the data. The data are expressed as the mean  $\pm$  standard error of the mean. The means (and standard deviation; SD) are presented in tables 3 and 4. A two-sided  $p$ -value  $\leq 0.05$  is considered significant for all analyses.

Table 3 shows the first analysis, which included 136 cases, 34 (25.0%) of which were PARS. Among the clusters, Cluster 3 ( $N=24$ ) contained 20 PARS cases (83%). This cluster had a higher average differential score between VIQ and PIQ (VIQ-PIQ) than the rest of the sample ( $12.9 \pm 2.3$  vs.  $6.3 \pm 1.0$ ,  $p=0.009$ ). Also, the mean paternal and maternal ages were relatively high at 41 and 33 years, respectively.

Although less relevant to our main focus on PARS, table 3 displays some other significant results<sup>1</sup>. Compared to the rest of the data, Cluster 1 had more familial cases ( $p=0.004$ ), lower average UPSIT score ( $p=0.039$ ), higher proportion of deficit syndrome cases ( $p=0.009$ ), lower neuropsychological scores ( $p<0.0001$ ), and higher PANSS (standard

<sup>1</sup>For additional information on results not pertaining to the PARS clusters (e.g. sample sizes in the individual clusters, group sizes, etc), please contact the corresponding author.

model) subscale scores (all  $p < 0.005$ ). Cluster 2 showed higher scores on block design, object assembly and digit symbol scores ( $p = 0.003$ ,  $0.013$  and  $0.001$ , respectively) compared to the rest of the data. Cluster 4 and Cluster 5 demonstrated higher FIQ, VIQ, PIQ and most of the WAIS-R subtest scores than the rest of the data. The standard model PANSS subscales were lower in Cluster 5 than the rest of the data (all  $p < 0.05$ ), but no difference was observed in Cluster 4. Clusters 5 and 6 consisted of only males with early age onset of psychosis and these cases had higher PIQ scores than VIQ scores.

The second analysis yielded 123 cases of which 30 (24.4%) were PARS. This analysis is shown in table 4. Cluster 2 produced the highest proportion of PARS with 10 out of 14 cases (71%; these were all sporadic patients). Inspection of the data revealed that the majority of the cases in this Cluster were also in Cluster 3 in the first analysis discussed above. Interestingly, compared to the other clusters, the Cluster 2 cases were mostly females (93% vs. 35% females,  $p < 0.0001$ ) with earlier age of onset of psychosis ( $17.2 \pm 1.7$  vs.  $22.4 \pm 0.6$ ,  $p = 0.006$ ). The mean paternal age was higher in the Cluster 2 group ( $39.3$ ;  $SD = 8.8$ ) than the rest of the data. None of the Cluster 2 cases had the deficit syndrome, and the neuropsychological scores and the PANSS scores for this group were not significantly different from those in the other groups.

From the other significant results in table 4<sup>1</sup>, Cluster 1 had a higher proportion of males ( $p = 0.001$ ), more schizophrenia than schizoaffective cases ( $p = 0.018$ ), lower mean UPSIT score ( $p = 0.006$ ), higher proportion of deficit syndrome cases ( $p < 0.0001$ ), lower main WAIS-R scores (FIQ:  $p = 0.025$ ; PIQ:  $p = 0.030$ ; VIQ:  $p = 0.031$ ), lower standard scale PANSS positive symptoms ( $p = 0.0005$ ), and higher standard scale PANSS negative symptoms ( $p < 0.0001$ ) compared to the rest of the data. Cluster 3 contained 100% male cases with a larger difference between VIQ and PIQ scores than the rest of the data ( $p < 0.0001$ ). This is due to the low mean PIQ score, which is the lowest among the 7 clusters ( $p = 0.019$ ). Cluster 4 demonstrated lower paternal age ( $p = 0.003$ ), higher FIQ, VIQ and PIQ scores (all  $p \leq 0.02$ ), and lower standard PANSS negative and general psychopathology scores ( $p < 0.0001$ ) than the rest of the data. Additionally, this is the only group with lower mean VIQ than PIQ. Cluster 5 consisted of 100% females with a later onset of psychosis ( $p < 0.0001$ ), higher UPSIT score ( $p = 0.007$ ), more familial cases ( $p = 0.0003$ ), and lower standard PANSS symptom subscales (all  $p < 0.01$ ) than the rest of the data. Cluster 6 had a higher proportion of females than the rest of the data ( $p = 0.0005$ ). Cluster 7 showed earlier onset of psychosis ( $p = 0.008$ ), higher proportion of males ( $p = 0.018$ ), more familial cases ( $p = 0.003$ ), and lower FIQ ( $p = 0.010$ ) and VIQ ( $p = 0.001$ ) than the rest of the data. Both Cluster 6 and Cluster 7 had no deficit syndrome cases, and they had higher PANSS standard positive and general symptoms scales ( $p < 0.0001$ ) than the rest of the data.

It should be noted that in the first cluster analysis, the 34 PARS cases were distributed as follows: Cluster 1 (N=5), Cluster 2 (3), Cluster 3 (20), Cluster 4 (2), Cluster 5 (2), Cluster 6 (1) and Cluster 7 (1). In the second cluster analysis, the 30 PARS patients were distributed as follows: Cluster 1 (N=1), Cluster 2 (10), Cluster 3 (9), Cluster 4 (3), Cluster 5 (3), Cluster 6 (2) and Cluster 7 (2). Additionally, the 20 PARS cases in the Cluster 3 group of the first analysis were 10 males and 10 females. Among these, three cases were excluded in the second analysis due to missing PANSS scores. Of the remaining 17 PARS cases, most belong to Cluster 2 and Cluster 3, depending on their sex. Cluster 3 contained six male cases, whereas five females and one male were in Cluster 2. The remaining five PARS cases were distributed as follows: Cluster 1 (one male), Cluster 5 (two females) and Cluster 6 (two females).

Due to missing data points on some of the core measures, certain cases were only captured in the first clustering analysis (N=16; 8 males and 8 females), and some cases were only

captured in the second analysis (N=3; 2 males and 1 female). 120 (70 males; 50 females) of the cases were captured in both clustering analyses. See supplementary data for the descriptive data for the cases that overlap (are both in analysis 1 and 2) and for the cases that are either only in analysis 1 or 2.

## Discussion

This study employed *k*-means clustering analyses to examine if specific illness features of schizophrenia are associated with later paternal age. We identified PARS cases that clustered in groups with particular characteristics. One group was characterized by a greater differential between verbal and performance intelligence, and the other group showed a high concentration of female cases and significant early onset of psychosis.

Our first clustering analysis, which considered demographic variables and neuropsychological test score variables, showed a cluster containing 83% PARS cases. It was characterized by relatively high paternal age (mean age = 41 years). The mean maternal age (33 years) was also relatively high in this cluster group. Interestingly, the cases in this group demonstrated a significant difference between the WAIS-R verbal and performance intelligence, with verbal functioning being on average 12.97 points better.

The verbal versus performance IQ decrement is notable. The result is driven by better performance on all the verbal subtasks (arithmetic, digit span, information, vocabulary, comprehension, similarities) compared to the performance subtasks (object assembly, picture arrangement, picture completion, digit symbol, block design). This result supports previous findings in other populations showing a strong relationship between older fathers and human intelligence (Auroux et al., 1989; Malaspina et al., 2005). These data also replicate previous findings from the same population (Malaspina et al., 2002b), despite using a different method of analysis, suggesting reductions in non-verbal intelligence compared to verbal intelligence in cases with older paternal age. The fact that paternal age has a larger effect on performance intelligence than verbal intelligence may be of interest. Future research that examines intelligence together with other cognitive and neurological symptoms would be valuable in further determining the distinctiveness of PARS as a subgroup of schizophrenia.

Our second separate *k*-means clustering analysis, which included the VIQ-PIQ discrepancy score, the PANSS standard subscale scores and the same demographic variables, revealed a cluster consisting of 71% PARS cases. This cluster had a significantly higher concentration of females and demonstrated an earlier age of psychosis onset. These findings show that later paternal age may have a particularly strong influence on the symptoms and clinical characteristics of female PARS cases in comparison with other female cases. A high risk of schizophrenia for females of older fathers was recently reported by Perrin et al (2010). Females of an affected sister born to fathers 35 years and older had a fourfold greater risk of schizophrenia than females with an affected sister born to fathers <35 years at time of birth. By contrast the risk of schizophrenia in males with an affected brother was only doubled for older versus younger fathers. They proposed that paternally expressed genes on the X chromosome could play a role in the risk associated with females of an affected sister born to older fathers.

A non-significant discrepancy (6.21 points) between the verbal and performance intelligence was evident in the high-concentration PARS group in the second analysis. The reduction in discrepancy from the first analysis may not be surprising as we included the VIQ-PIQ discrepancy score as a factor in the second cluster analysis. The reduction may also be partly explained by gender separation in this cluster. As discussed above, 12 of the PARS cases

from Cluster 3 in the first analysis were sorted into Cluster 2 (five females and one male) and Cluster 3 (six males) in the second analysis. The mean VIQ-PIQ of these 12 subjects was 11.8, and it was reduced in Cluster 2 (mean VIQ-PIQ of the six patients was 6.17), while it was increased in Cluster 3 (mean VIQ-PIQ of the six patients was 17.5). The PANSS scores revealed similar results as found in the first cluster analysis.

The strengths of this study include the use of a statistical method that is particularly germane for resolving the heterogeneity of schizophrenia. *k*-means clustering analyses require variability in numerous factors to generate separate independent clusters that share common attributes. Moreover, *k*-means clustering analysis assures minimal variation within the clusters, but maximum variation between the clusters, creating more homogeneous subgroups. Our results show that etiological data and clinical information can both be considered in the procedures.

It should be noted that our total PARS sample size for the *k*-means clustering analysis was relatively small. In the first analysis we had a total of 34 PARS cases, with 20 PARS cases belonging to the high PARS concentration cluster. The second analysis only included 30 PARS cases, with 10 cases being in the high PARS concentration cluster. Future studies that screen for higher inclusion rates of PARS cases may be successful in recruiting larger PARS samples. Such studies may also want to include variables not examined here but found to be of etiological significance.

In summary, our findings demonstrate that *k*-means clustering analysis may be a useful statistical method to explore the heterogeneous nature of schizophrenia symptoms and to examine the relationships between etiological, clinical and cognitive symptoms in PARS subgroups. The results provide further evidence that the genetic and neurobiological underpinnings of schizophrenia associated with the illness risk attributable to later paternal age may be different than that of other cases. This method may be particularly helpful in establishing the phenotype of PARS, and in generating new hypotheses that will help in the development of novel clinical treatment approaches to the variability of symptoms in PARS patients.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

The authors thank Benjamin Barasch for assistance in editing this paper.

### Funding sources

DM (RC1MH088843-02 and 2K24MH00169). The funding sources had no role in study design; in the collection, analysis and interpretation of data; in writing the report; and in the decision to submit the paper for publication.

## References

- Antonius D, Kimhy D, Harkavy-Friedman J, Crystal S, Goetz R, Malaspina D. Paternal Age Related Schizophrenia and Cardiac Autonomic Regulation Profiles. *Schizophr Res.* 2010 Epub ahead of print.
- Auroux MR, Mayaux MJ, Guihard-Moscato ML, Fromantin M, Barthe J, Schwartz D. Paternal age and mental functions of progeny in man. *Human Reproductivity.* 1989; 4:794–797.
- Bell MD, Johannesen JK, Creig TC, Wexler BE. Memory profiles in schizophrenia: categorization validity and stability. *Schizophr Res.* 2010; 118:26–33. [PubMed: 20085855]

- Brown AS, Schaefer CA, Wyatt RJ, Begg MD, Goetz R, Bresnahan MA, Harkavy-Friedman J, Gorman JM, Malaspina D, Susser ES. Paternal age and risk of schizophrenia in adult offspring. *Am J Psychiatry*. 2002; 159:1528–1533. [PubMed: 12202273]
- Byrne M, Agerbo E, Ewald H, Eaton WW, Mortensen PB. Parental age and risk of schizophrenia: a case-control study. *Arch Gen Psychiatry*. 2003; 60:673–678. [PubMed: 12860771]
- Carpenter WT, Heinrichs DW, Wagman AM. Deficit and nondeficit forms of schizophrenia: the concept. *Am J Psychiatry*. 1988; 145:578–583. [PubMed: 3358462]
- Doty RL, Shaman P, Kimmelman CP, Dann MS. University of Pennsylvania Smell Identification Test: A rapid quantitative olfactory function test for the clinic. *Laryngoscope*. 1984; 94:176–178. [PubMed: 6694486]
- Garver DL, Holcomb JA, Christensen JD. Heterogeneity of response to antipsychotics from multiple disorders in the schizophrenia spectrum. *J Clin Psychiatry*. 2000; 61:964–972. [PubMed: 11206608]
- Gibbons FD, Roth FP. Judging the quality of gene expression-based clustering methods using gene annotation. *Genome Res*. 2002; 12:1574–1581. [PubMed: 12368250]
- Hand DJ, Heard NA. Finding groups in gene expression data. *J Biomed Biotechnol*. 2005; 2:215–225. [PubMed: 16046827]
- Huang Z. Extensions to the k-means algorithm for clustering large data sets with categorical values. *Data Mining and Knowledge Discovery*. 1998; 2:283–304.
- Jindal R, MacKenzie EM, Baker GB, Yeragani VK. Cardiac risk and schizophrenia. *J Psychiatry Neurosci*. 2005; 20:293–295.
- Jonsson SA, Jonsson H. Clusters in a cohort of untreated schizophrenia: prognostic importance, atypical cases and the familial versus sporadic distinction. *Acta Psychiatr Scand*. 1992; 86:287–95. [PubMed: 1456072]
- Kay SR, Fiszbein A, Opler LA. The Positive and Negative Syndrome Scale (PANSS) for schizophrenia. *Schizophrenia Bulletin*. 1987; 13:261–276. [PubMed: 3616518]
- Kimhy D, Harlap S, Fennig S, Deutsch L, Draiman BG, Corcoran C, Goetz D, Nahon D, Malaspina D. Maternal household crowding during pregnancy and the offspring's risk of schizophrenia. *Schizophr Res*. 2006; 86:23–29. [PubMed: 16740377]
- Kirkpatrick B, Buchanan RW, McKenney PD, Alphas LD, Carpenter WT. The Schedule for the deficit syndrome: an instrument for research in schizophrenia. *Psychiatry Res*. 1989; 30:119–123. [PubMed: 2616682]
- Kopala LC, Good KP, Morrison K, Bassett AS, Alda M, Honer WG. Impaired olfactory identification in relatives of patients with familial schizophrenia. *Am J Psychiatry*. 2001; 158:1286–1290. [PubMed: 11481164]
- MacQueen, JB. Some Methods for Classification and Analysis of Multivariate Observations. *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*; Berkeley, CA: University of California Press; 1967. p. 281-297.
- Malaspina D. Paternal factors and schizophrenia risk: de novo mutations and imprinting. *Schizophr Bull*. 2001; 27:379–393. [PubMed: 11596842]
- Malaspina D, Brown A, Goetz D, Alia-Klein N, Harkavy-Friedman J, Harlap S, Fennig S. Schizophrenia risk and paternal age: a potential role for de novo mutations in schizophrenia vulnerability genes. *CNS Spectr*. 2002a; 7:26–29. [PubMed: 15254446]
- Malaspina D, Corcoran C, Fahim C, Berman A, Harkavy-Friedman J, Yale S, Goetz D, Goetz R, Harlap S, Gorman J. Paternal age and sporadic schizophrenia: evidence for de novo mutations. *Am J Med Genet*. 2002b; 114:299–303. [PubMed: 11920852]
- Malaspina D, Harlap S, Fennig S, Heiman D, Nahon D, Feldman D, Susser ES. Advancing paternal age and the risk of schizophrenia. *Arch Gen Psychiatry*. 2001; 58:361–367. [PubMed: 11296097]
- Malaspina D, Reichenberg A, Weiser M, Fennig S, Davidson M, Harlap S, Wolitzky R, Rabinowitz J, Susser E, Knobler HY. Paternal age and intelligence: implications for age-related genomic changes in male germ cells. *Psychiatr Genet*. 2005; 15:117–125. [PubMed: 15900226]
- Malaspina D, Wray AD, Friedman JH, Amador X, Yale S, Hasan A, Gorman JM, Kaufman CA. Odor discrimination deficits in schizophrenia: Association with eye movement dysfunction. *The Journal of Neuropsychiatry and Clinical Neurosciences*. 1994; 6:273–278. [PubMed: 7950351]



- Mohr PE, Cheng CM, Claxton K, Conley RR, Feldman JJ, Hargreaves WA, Lehman AF, Lenert LA, Mahmoud R, Marder SR, Neumann PJ. The heterogeneity of schizophrenia in disease states. *Schizophr Res.* 2004; 71:83–95. [PubMed: 15374576]
- Nurnberger JI Jr, Blehar MC, Kaufmann CA, York-Cooler C, Simpson SG, Harkavy-Friedman J, Severe JB, Malaspina D, Reich T. Diagnostic interview for genetic studies. Rationale, unique features, and training. NIMH Genetics Initiative. *Arch Gen Psychiatry.* 1994; 51:849–859. [PubMed: 7944874]
- Perrin M, Harlap S, Kleinhaus K, Lichtenberg P, Manor O, Draiman B, Fennig S, Malaspina D. Older paternal age strongly increases the morbidity for schizophrenia in sisters of affected females. *Am J Genet B Neuropsychiatr Genet.* 2010; 153B:1329–1335.
- Richards AL, Holmans P, O'Donovan MC, Owen MJ, Jones L. A comparison of four clustering methods for brain expression microarray data. *BMC Bioinformatics.* 2008; 25:490. [PubMed: 19032745]
- Rosenfield PJ, Kleinhaus K, Opler M, Perrin M, Learned N, Goetz R, Stanford A, Messinger J, Harkavy-Friedman J, Malaspina D. Later paternal age and sex differences in schizophrenia symptoms. *Schizophr Res.* 2010; 116:191–195. [PubMed: 19926452]
- Silver H, Shmoish M. Analysis of cognitive performance in schizophrenia patients and healthy individuals with unsupervised clustering models. *Psychiatry Res.* 2008; 159:167–179. [PubMed: 18440074]
- Sipos A, Rasmussen F, Harrison G, Tynelius P, Lewis G, Leon DA, Gunnell D. Paternal age and schizophrenia: a population based cohort study. *BMJ.* 2004; 329:1070. [PubMed: 15501901]
- Torrey EF, Buka S, Cannon TD, Goldstein JM, Seidman LJ, Liu T, Hadley T, Rosso IM, Bearden C, Yolken RH. Paternal age as a risk factor for schizophrenia: how important is it? *Schizophr Res.* 2009; 114:1–5. [PubMed: 19683417]
- Tsuang MT, Lyons MJ, Faraone SV. Heterogeneity of schizophrenia. Conceptual models and analytic strategies. *Br J Psychiatry.* 1990; 156:17–26. [PubMed: 2404538]
- Tsuchiya KJ, Takagai S, Kawai M, Matsumoto H, Nakamura K, Minabe Y, Mori N, Takei N. Advanced paternal age associated with an elevated risk for schizophrenia in offspring in a Japanese population. *Schizophr Res.* 2005; 76:337–342. [PubMed: 15949666]
- Wagstaff, K.; Cardie, C.; Rogers, S.; Schroedl, S. Constrained k-means clustering with background knowledge. *Proceedings of the Eighteenth International Conference on Machine Learning*; 2001. p. 577-584.
- Wechsler, D. *Wechsler Adult Intelligence Scale—Revised: Manual.* San Antonio, TX: The Psychological Corporation; 1981.
- White L, Harvey PD, Opler L, Lindenmayer JP. Empirical assessment of the factorial structure of clinical symptoms in schizophrenia. A multisite, multimodel evaluation of the factorial structure of the Positive and Negative Syndrome Scale. The PANSS Study Group. *Psychopathology.* 1997; 30:263–274. [PubMed: 9353855]
- Yeung KY, Haynor DR, Ruzzo WL. Validating clustering for gene expression data. *Bioinformatics.* 2001; 17:309–318. [PubMed: 11301299]
- Zammit S, Allebeck P, Dalman C, Lundberg I, Hemmingson T, Owen MJ, Lewis G. Paternal age and risk for schizophrenia. *Br J Psychiatry.* 2003; 183:405–408. [PubMed: 14594914]

**Table 1**

## Demographic variables

	<b>Cluster 1</b>	<b>Cluster 2</b>
Sex (male/female)	78/58	72/51
Diagnosis (schizophrenia/schizoaffective disorder)	103/33	92/31
Family history of schizophrenia (yes/no)	41/95	40/83
PARS (yes/no)	34/102	30/93
Deficits Syndrome (yes/no)	15/102	13/93

**Table 2**

Demographic, clinical, neuropsychological and olfaction data

	Cluster 1			Cluster 2		
	N	Mean	SD	N	Mean	SD
Age of onset of psychosis	136	21.90	6.62	123	21.77	6.70
Maternal age at the case's birth	129	28.39	5.95	117	28.36	5.86
Paternal age at the case's birth	136	32.49	7.40	123	33.00	7.42
PANSS <sup>1</sup>	122	52.52	13.53	123	52.23	13.26
PANSS standard model	123	12.72	5.58	123	12.55	5.48
Positive symptoms	122	13.94	5.59	123	13.96	5.56
Negative symptoms	123	25.89	6.83	123	25.74	6.72
General psychopathology symptoms	123	9.89	4.52	123	9.72	4.44
PANSS 5-factor model	123	16.74	6.40	123	16.73	6.41
Positive symptoms	123	8.37	3.13	123	8.46	3.22
Negative symptoms	123	8.98	3.57	123	8.96	3.50
Activation	123	10.34	3.48	123	10.23	3.32
Dysphoric mood	134	90.34	13.79	123	90.23	14.96
Autistic preoccupation	134	94.13	14.70	123	93.97	14.72
WAIS-R <sup>2</sup>	134	86.73	15.42	123	86.74	15.68
Full scale intelligence (FIQ)	134	7.40	11.09	123	7.23	11.15
Verbal intelligence (VIQ)	136	8.13	2.87	123	8.11	2.79
Performance intelligence (PIQ)	136	8.88	2.93	123	8.80	2.94
WAIS-R, Verbal subtests	136	9.49	3.29	123	9.47	3.24
Arithmetic	136	9.40	3.39	123	9.30	3.43
Digit span	136	8.26	3.26	123	8.26	3.34
Information	136	8.90	2.90	123	8.98	2.95
Vocabulary	136	7.69	3.31	120	7.72	3.40
Comprehension	136	7.60	2.93	123	7.59	2.96
Similarities	136	7.34	3.20	123	7.43	3.28
WAIS-R, Performance subtests	136	6.74	2.58	122	6.80	2.70
Object assembly	136	8.29	3.30	122	8.29	3.33
Picture Arrangement						
Picture completion						
Digit symbol						
Block design						

	Cluster 1			Cluster 2		
	N	Mean	SD	N	Mean	SD
Smell Identification (UPSIT <sup>3</sup> )						
Total	69	32.01	4.75	66	32.21	4.70

<sup>1</sup> PANSS: Positive and Negative Syndrome Scale (Kay et al., 1989);

<sup>2</sup> WAIS-R: Wechsler Adult Intelligence Scale—Revised (Wechsler, 1981);

<sup>3</sup> UPSIT: University of Pennsylvania Smell Identification Test (Doty et al., 1984)

**Table 3**

First cluster analysis<sup>1</sup>: means (and standard deviations) of the demographic, clinical, neuropsychological and olfaction variables according to cluster.

Cluster	N	Age onset	Sex	Diagnosis	Maternal age	Paternal age	PAKS	Family history	UPST	Deficit syndrome
1	34	20.12 (6.03)	0.35 (0.49)	0.79 (0.41)	28.12 (6.57)	32.53 (7.17)	0.15 (0.36)	0.50 (0.51)	30.11 (5.24)	2.73 (0.45)
2	19	20.11 (5.18)	0.53 (0.51)	0.74 (0.45)	27.08 (3.86)	29.86 (4.43)	0.16 (0.37)	0.05 (0.23)	34.17 (4.55)	2.94 (0.25)
3	24	22.63 (8.02)	0.50 (0.51)	0.71 (0.46)	33.23 (6.36)	41.47 (6.46)	0.83 (0.38)	0.13 (0.34)	32.56 (4.93)	2.89 (0.32)
4	13	24.08 (3.93)	0.62 (0.51)	0.69 (0.48)	29.13 (4.31)	31.87 (5.68)	0.15 (0.38)	0.38 (0.51)	33.25 (3.41)	2.92 (0.28)
5	15	19.80 (5.88)	0.00 (0.00)	0.73 (0.46)	26.60 (4.61)	30.93 (5.98)	0.13 (0.35)	0.20 (0.41)	30.00 (5.83)	2.92 (0.28)
6	2	17.00 (1.41)	0.00 (0.00)	1.00 (0.00)	29.00	32.50 (6.56)	0.50 (0.71)	0.00 (0.00)	31.00	3.00 (0.00)
7	29	25.03 (7.15)	0.55 (0.51)	0.79 (0.41)	25.95 (5.11)	27.82 (5.04)	0.03 (0.19)	0.41 (0.50)	32.67 (3.64)	2.92 (0.28)

  

	PANSS positive symptoms	PANSS negative symptoms	PANSS general symptoms	PANSS-5, negative symptoms	PANSS-5, activation	PANSS-5, dysphoric mood	PANSS-5, autistic preoccup.	PIQ	VIQ	VTQ-PTQ
1	15.16 (6.95)	17.84 (5.13)	28.84 (7.13)	20.56 (6.20)	10.22 (4.05)	9.34 (3.53)	12.63 (3.23)	71.32 (6.13)	76.62 (6.98)	5.29 (7.67)
2	12.56 (4.59)	12.75 (6.89)	25.75 (6.16)	16.44 (9.64)	7.88 (2.16)	9.69 (3.93)	10.00 (2.61)	92.00 (7.23)	98.89 (7.27)	6.89 (10.21)
3	12.36 (5.23)	13.27 (5.29)	25.50 (6.92)	16.05 (5.52)	7.82 (2.15)	8.82 (3.40)	10.18 (3.46)	83.35 (5.97)	96.22 (6.32)	12.87 (10.42)
4	10.50 (3.83)	11.67 (4.23)	23.75 (5.40)	14.17 (4.67)	7.42 (1.68)	9.67 (4.19)	9.17 (2.95)	104.77 (8.02)	119.54 (8.20)	14.77 (11.81)
5	9.85 (2.97)	11.00 (4.28)	20.08 (3.43)	13.77 (4.78)	7.00 (1.68)	6.54 (2.03)	7.54 (1.76)	106.40 (8.20)	102.13 (7.52)	-4.27 (10.49)
6	13.50 (0.71)	12.50 (2.12)	23.00 (2.83)	14.00 (0.00)	7.00 (1.41)	8.50 (0.71)	10.00 (5.66)	142.50 (14.85)	126.50 (12.02)	-16.00 (2.83)
7	12.54 (5.60)	13.04 (4.66)	26.81 (7.08)	15.69 (4.67)	8.12 (3.44)	9.19 (3.75)	9.85 (3.67)	81.75 (5.82)	92.04 (6.22)	10.29 (8.61)

  

Info	Picture comp	Digit span	Picture arrange	Vocab	Block design	Arith assembly	Object	Comp	Digit symbol	Similarities
1	5.91 (2.47)	5.21 (2.28)	5.15 (1.56)	5.62 (1.84)	5.21 (1.45)	5.59 (1.92)	4.91 (2.02)	4.74 (1.78)	4.53 (1.54)	5.68 (1.51)
2	10.74 (2.00)	7.47 (1.87)	10.05 (2.12)	8.68 (1.63)	10.37 (2.39)	8.79 (2.90)	9.42 (2.78)	9.42 (2.36)	8.53 (2.37)	9.63 (1.89)
3	9.67 (2.35)	6.42 (2.15)	9.21 (2.17)	7.25 (2.01)	7.42 (2.06)	9.42 (2.19)	7.17 (2.97)	8.50 (2.13)	6.63 (1.88)	9.13 (2.03)
4	14.15 (1.46)	11.15 (2.97)	10.85 (2.30)	11.23 (2.24)	14.54 (2.26)	11.08 (1.61)	10.00 (2.65)	14.00 (2.24)	9.08 (2.84)	13.69 (1.70)
5	11.40 (1.92)	11.60 (1.50)	11.93 (3.13)	10.47 (1.77)	12.40 (2.38)	9.13 (2.50)	10.87 (2.85)	9.53 (1.85)	8.53 (2.50)	10.07 (1.49)
6	13.00 (2.83)	15.00 (4.24)	12.50 (0.71)	14.50 (4.95)	14.00 (2.83)	13.50 (3.54)	16.50 (0.71)	11.50 (3.54)	9.50 (0.71)	15.00 (2.83)
7	9.41 (2.31)	6.07 (1.75)	8.03 (2.13)	6.45 (2.31)	7.34 (1.84)	7.38 (2.13)	6.97 (1.74)	8.00 (1.71)	6.07 (1.67)	8.86 (1.55)

<sup>1</sup> Variables used in generating the clusters: age of onset of psychosis, sex, paternal age, family history of schizophrenia, WAIS-R verbal subtests (arithmetic, digit span, information, vocabulary, comprehension, similarities) and WAIS-R performance subtests (object assembly, picture arrangement, picture completion, digit symbol, block design).

**Table 4**

Second cluster analysis<sup>1</sup>: means (and standard deviations) of the demographic, clinical, neuropsychological and olfaction variables according to cluster.

Cluster	N	Age onset	Sex	Diagnosis	Maternal age	Paternal age	PAMS	Family history	UPST	Deficit
1	21	21.24 (5.84)	0.10 (0.30)	0.95 (0.22)	27.28 (5.25)	30.41 (4.74)	0.05 (0.22)	0.52 (0.51)	29.44 (3.85)	2.47 (0.51)
2	14	17.21 (4.68)	0.93 (0.27)	0.57 (0.51)	29.18 (6.78)	39.32 (8.83)	0.71 (0.47)	0.00 (0.00)	35.40 (2.70)	3.00 (0.00)
3	22	19.82 (4.68)	0.00 (0.00)	0.73 (0.46)	29.74 (6.33)	34.05 (7.24)	0.41 (0.50)	0.05 (0.21)	31.00 (7.95)	2.90 (0.30)
4	21	21.33 (5.16)	0.24 (0.44)	0.81 (0.40)	26.75 (4.65)	28.70 (5.21)	0.14 (0.36)	0.19 (0.40)	32.90 (2.88)	3.00 (0.00)
5	19	29.95 (7.41)	1.00 (0.00)	0.74 (0.45)	29.16 (5.86)	35.18 (7.80)	0.16 (0.37)	0.68 (0.48)	35.31 (2.93)	2.88 (0.33)
6	11	25.00 (5.71)	0.91 (0.30)	0.55 (0.52)	29.97 (4.04)	31.95 (4.08)	0.18 (0.40)	0.09 (0.30)	31.67 (4.23)	3.00 (0.00)
7	15	17.53 (3.38)	0.13 (0.35)	0.73 (0.46)	26.83 (7.32)	33.25 (9.02)	0.13 (0.35)	0.67 (0.49)	31.50 (4.66)	3.00 (0.00)

  

	PANSS positive symptoms	PANSS negative symptoms	PANSS general symptoms	PANSS-5, negative symptoms	PANSS-5, positive symptoms	PANSS-5, activation	PANSS-5, dysphoric mood	PANSS-5, autistic preoccup.	PIQ	VIQ	VIQ-PIQ
1	8.86 (1.85)	23.19 (4.03)	27.19 (4.73)	26.29 (6.34)	6.33 (1.74)	9.52 (2.11)	7.62 (3.46)	10.24 (2.62)	80.00 (12.84)	87.67 (15.13)	7.67 (7.36)
2	13.43 (4.16)	13.29 (4.30)	25.21 (4.17)	16.43 (5.09)	10.29 (3.71)	8.14 (2.63)	9.14 (3.03)	10.29 (2.70)	84.00 (8.73)	90.21 (13.24)	6.21 (8.59)
3	10.50 (3.38)	12.00 (2.81)	23.55 (4.87)	14.55 (3.46)	8.18 (2.94)	6.95 (1.46)	8.68 (3.30)	9.50 (2.54)	79.68 (9.62)	97.00 (11.84)	17.32 (9.55)
4	10.90 (2.98)	9.71 (1.98)	20.57 (3.67)	11.90 (2.17)	8.71 (3.00)	7.19 (1.54)	7.33 (2.80)	8.43 (2.46)	105.33 (17.79)	100.76 (13.55)	-4.57 (8.92)
5	9.42 (3.13)	10.53 (3.08)	21.95 (4.78)	13.11 (3.35)	7.95 (3.15)	6.74 (1.45)	8.37 (3.44)	8.58 (2.48)	88.74 (14.37)	99.32 (15.43)	10.58 (9.25)
6	20.45 (5.47)	13.18 (2.48)	35.91 (6.59)	16.91 (3.83)	16.18 (4.05)	10.55 (4.93)	13.55 (3.27)	13.27 (4.17)	85.18 (12.25)	98.27 (15.34)	13.09 (7.06)
7	20.40 (4.10)	15.40 (3.81)	32.00 (6.07)	18.07 (5.24)	15.07 (3.77)	11.93 (4.80)	10.73 (1.87)	13.60 (3.42)	81.67 (12.81)	82.40 (9.93)	0.73 (9.16)

  

Info	Picture comp	Digit span	Picture arrange	Vocab	Block design	Arith	Object assembly	Comp	Digit symbol	Similarities
1	7.90 (3.92)	6.52 (2.86)	6.95 (2.69)	7.86 (3.20)	6.76 (2.51)	7.62 (3.17)	6.65 (3.70)	6.90 (3.21)	5.90 (2.21)	7.29 (2.72)
2	8.29 (3.29)	6.79 (2.04)	7.79 (2.19)	9.07 (2.87)	7.57 (2.38)	7.79 (2.69)	7.93 (3.36)	7.93 (2.50)	7.71 (3.02)	8.36 (2.65)
3	10.14 (2.53)	6.18 (2.56)	6.32 (2.28)	9.59 (3.14)	7.68 (3.08)	8.86 (2.64)	6.59 (2.56)	8.59 (3.49)	5.18 (1.56)	9.50 (2.24)
4	10.48 (2.48)	10.76 (3.36)	10.90 (2.74)	10.10 (2.76)	11.95 (3.46)	9.33 (2.56)	10.45 (3.55)	9.33 (2.78)	8.86 (2.90)	10.38 (2.75)
5	10.21 (3.29)	7.74 (3.45)	8.47 (2.20)	11.32 (3.23)	8.05 (3.27)	8.53 (2.52)	7.37 (3.15)	10.32 (3.84)	7.63 (2.54)	10.05 (3.01)
6	11.00 (3.16)	6.45 (2.73)	8.73 (3.26)	11.09 (4.21)	8.00 (2.79)	7.36 (3.11)	7.91 (3.24)	9.00 (2.32)	6.36 (2.01)	10.18 (3.16)
7	8.33 (2.99)	6.80 (3.08)	7.07 (2.02)	6.13 (2.63)	7.29 (2.40)	6.33 (2.02)	7.21 (2.78)	5.33 (2.13)	5.86 (2.60)	7.00 (2.56)

<sup>1</sup>Variables used in generating the clusters: age of onset of psychosis, sex, paternal age, family history of schizophrenia, VIQ-PIQ, PANSS subscales (standard model: positive, negative and general psychopathology subscale scores).