Subtyping Schizophrenia In An African Population: A Mixture Model Approach

by

Daniel Jan Hendrik Niehaus

Dissertation presented for the degree of Doctor of Psychiatry

in the Faculty of Medicine and Health Sciences

March, 2014

Promoters: Dr. Esme Jordaan

Biostatistics Unit

Medical Research Council

Bellville

Prof. W. Pienaar

Department of Psychiatry

Faculty of Health Sciences

University of Stellenbosch

Cr tkn 2014

DECLARATION

By submitting this dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: ....

ABSTRACT

Schizophrenia is a phenotypically heterogeneous disorder believed to have a strong genetic component. Limiting its clinical heterogeneity by means of subtyping may help to shed light on some of the genetic underpinnings of the disease. This study describes the application of factor analysis (FA), latent class analysis (LCA) and factor mixture modeling in a sample of 734 Xhosa-speaking schizophrenic subjects using factor analytically derived variables previously identified in an independent sample of this population. LCA was performed on the following 8 SANS and SAPS items identified by preliminary exploration of the data: eye contact, auditory hallucinations, global hallucinations score, global delusions score, grooming, affective non-responsiveness, spontaneous movement, and commenting voices. A four class model provided the best fit. Classes 1 and 2 were characterized by predominantly positive and predominantly negative symptoms, respectively, class 3 by both positive and negative symptoms and class 4 by few or absent symptoms. A history of cannabis use or abuse increased the probability of a subject being allocated to class 1, while being male made a person more likely to be included in class 2. Factor mixture modelling was performed by first using latent class analysis, then factor analysis and then the factor mixture analysis were done. The fit among these three types were then investigated. The results show that factor mixture modelling uncovered a heterogeneous latent variable structure that fits the data well with the latent classes capturing distinct positive symptom/behaviours and factors capturing severity variations. This study, the first to report on the latent class structure of schizophrenia in a sample of patients from a sub-Saharan African population, supports the universality of specific latent classes across ethnic boundaries. The results further support reports that gender, sibpair status and cannabis use may influence the phenomenology of schizophrenia. The identification of subgroups may represent an intermediate step in the search for endophenotypes of schizophrenia.

ABSTRAK

Skisofrenie is 'n psigiatriese steuring met 'n heterogene fenotipe en 'n vermoedelik sterk genetiese vatbaarheid. Ten einde die lig te werp op die genetiese onderbou van skisofrenie word gepoog om die kliniese heterogenisiteit te beperk deur middel van subgroepering. Hierdie studie beskryf die gebruik van latente klas analise (LKA) in 'n groep van 734 Xhosa-sprekendes met skisofrenie. Die LKA word baseer op die gebruik van veranderlikes wat deur middel van faktor analise op simptome in 'n onafhanklike studiegroep van Xhosa-sprekendes met skisofrenie verkry is. Die LKA is gedoen op die volgende 8 "SAPS" en "SANS" veranderlikes wat deur voorlopige ondersoek van die data ge-indentifiseer is: oogkontak, gehoorshallusinasies, globale hallusinasie telling, globale waantelling, selfversorging, affektiewe nie-responsiwiteit, spontane beweging en stemme wat kommentaar lewer. 'n Vierklas oplossing het die beste passing getoon. Klas 1 en 2 is gekenmerk deur oorwegend positiewe en negatiewe simptome onderskeidelik, klas 3 het beide positiewe en negatiewe simptome gehad en klas 4 het baie min of geen simptome getoon nie. 'n Geskiedenis van kannabis gebruik of misbruik het die kans verhoog dat die individue in klas 1 gevind sou word, terwyl manlike geslag as veranderlike die kanse verhoog het vir allokasie in klas 2. Faktor mengsel modelering is gedoen deur eers 'n latent klas analise te voltooi, gevolg deur 'n faktor analise, en laastens 'n factor mengsel analise. Die passing tussen die drie analises is daarna evalueer. Faktor mengsel modelering toon 'n heterogene latente klas struktuur wat voldoen aan die passingsvereistes. Die latente klasse blyk spesifieke positiewe simptome/gedrag te verteenwoordig, terwyl die factor grad van erns variasie aandui. Hierdie studie is die eerste om die latente klas struktuur van skisofrenie in 'n subsahara-Afrika populasie, die Xhosa, te beskryf. Die resultate onderstreep die universialiteit van die latente struktuur van skisofrenie se simptome oor etniese grense heen. Verder ondersteun die resultate die moontlike rol van geslag, aangetaste sibstatus en kannabis gebruik in skisofrenie se fenomenologie. Die identifisering van die subgroepe mag 'n intermediêre stap in die soektog vir endofenotipes van skisofrenie verteenwoordig.

TABLE OF CONTENTS

CHAPTER 2 METHODS

CHAPTER 3 RESULTS

CHAPTER 4 DISCUSSION AND CONCLUSIONS

APPENDIX

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ASP   Affected sibling pairs.

BIC   Bayesian information criterion

CFA   Confirmatory factor analysis

CFI   Comparative fit index

Df   Degrees of freedom

DIGS   Diagnostic interview for genetic studies

EM   Expectation maximization

FIML   Full information maximum likelihood

FMA   Factor mixture analysis

LL   Log likelihood

MAR   Missing at random

MCAR   Missing completely at random

MI   Multiple imputation

ML   Maximum likelihood

MLM   Maximum likelihood estimation, mean adjusted

RMSEA   Root mean square error of approximation

SANS   Schedule for the assessment of negative symptoms scale

SAPS   Schedule for the assessment of positive symptoms scale

SEM   Structural equation modeling

SRMR   Standardized root mean residual

WLSMV   Weighted least squares estimation, mean and variance adjusted

ACKNOWLEDGEMENTS

Om nooit op te gee is 'n gawe wat jou ouers jou gee. Om in jou te glo is die gawe wat jou gesin

jou gee. Vir beide gawes is ek oneindig dankbaar.

CHAPTER 1: INTRODUCTION

1.1 Background

This chapter provides an overview of the literature related to the factors purported to be involved in the pathogenesis of schizophrenia, focusing mainly on the complex genetic underpinnings of the disorder. The phenotypic heterogeneity of schizophrenia, the reasons for performing the study in a Xhosa population, and the statistical methods for reducing the heterogeneity of schizophrenia as a prelude to further genetic studies is also discussed.

1.1.1 What causes schizophrenia?

Schizophrenia is considered to be one of the most disabling of the psychotic disorders in terms of its impact on families, the health care system and the patients themselves. Indeed the World Health Report (2001) lists schizophrenia as one of the eight leading causes of disability-adjusted life years (age 15-44 years), probands have a reduced life-expectancy by approximately 10 years and the direct and indirect cost of the management of schizophrenia accounts for between 1.3% and 2.5% of annual health expenditure in Northern hemisphere countries (Knapp 1997, Rouillon et al. 1994, Statistishes Bundesamt 2004, Rössler et al. 2005).

Psychotic phases of the illness frequently require intensive care and hospitalization, most patients being left with a greater or lesser degree of residual impairment in their personal, social and occupational functioning. The majority of patients need regular, lifelong follow up and care by

psychiatric services and the community. The core clinical features and course of the illness have been well described and appear to be invariant across ethnic groups worldwide (Emsley et al. 2001, Niehaus et al.2005).

What appears more elusive, is finding the cause of schizophrenia. Research supports both environmental and genetic contributions. While the strongest risk factor for developing schizophrenia is a family history (heritability estimates ranging from 64 to 81%), environmental risk factors such as urbanization, substance abuse, childhood psychological trauma and adverse life events, have also been shown to be associated with the disorder (Stilo et al. 2010).

Indeed, the interplay between these seems critical for an understanding of the complex nature of schizophrenia. This interaction between environment and genetic make-up can be referred to as epigenetics and the modifications to the genes can be through several mechanisms. I refer the reader to Dudley et al. (2011) for a detailed description of the epigenetic mechanisms that may confer risk for the development of schizophrenia in an individual. In brief, environmental factors, including both prenatal and postnatal environments, can influence an individual's vulnerability for or resilience to developing schizophrenia. Khashan et al. (2008) in a study of 1.38 million Danish births between 1973-1995, supported the notion that perinatal factors could heighten an individual's risk for the disorder by demonstrating a relative risk of 1.67 for schizophrenia in offspring of mothers exposed to increased stress (i.e. death of a relative) in the first trimester. This is not surprising as a number of animal studies indicate that the prenatal and postnatal environments can affect adult behavioural outcome through interference with developmental processes (Dudley et al. 2011).

During the development of the human brain across its growth period, different environmental insults can occur and the resilience to these insults seems to differ from individual to individual. It

is assumed that periods of sensitivity exist and that insults during these periods may have a more detrimental effect on the behavioral outcome (i.e. no psychosis versus sub-clinical psychosis versus clinically significant psychosis) (Van Os 2010). It is important to understand each of these concepts and the interplay between them if we aim to best understand the vulnerability for schizophrenia.

The most widely accepted environmental insults of interest, with a relative risk in the range of 2, include urbanicity, developmental trauma or childhood adversity, minority status and cannabis use (Van Os 2010). Childhood adversity or developmental trauma studies suggest that the impact and timing of insults, such as severe physical abuse by the mother before the age of 12 years, significantly increases the risk for adult psychosis up to 3 times that of non-exposed children (Fisher et al. 2010). The findings on the impact of childhood adversity or developmental trauma seem independent of genetic confounding and reverse causality (Schreier et al. 2009, Arseneault et al. 2011). Minority group membership is an even more complex environmental risk factor. Research suggests that the risk is linked not to immigration or pre-immigration factors per se, but rather to own ethnic density in a geographical area, to the extent that the individual stands out in the wider social environment and that the increased risk for schizophrenia may be mediated by ongoing social adversity and discrimination (Bresnahan 2007, Schofield 2011). Indeed Schofield et al. (2011) provides an excellent example of this phenomenon in their study of 1500 of 6000 people per ward in South East London. When they compared psychosis incidence over a ten year period between Afro-Caribbean and Caucasian inhabitants of this area, they found that ethnic density of less than 25% was associated with an odds ratio of 2.88 (CI 1.89-4.39) for the development of psychosis. Urbanicity (independent of the definition of urban environment) shows a direct dose-response association with the development of psychosis and this holds true even when various confounders are taken into consideration and can apparently be modified by moving between the two settings during childhood and may be sensitive to social maladjustment

(including single parent families and residential instability) (March et al. 2008, Pedersen and Mortensen 2001, Kelly et al. 2010). Kelly et al. (2010) serves as an example of this when they reviewed the role of urbanicity in Ireland and found an age adjusted incidence ratio of 1.92 (95% confidence interval 1.52-2.44) for males and 1.34 (1.00-1.80) for females living in urban environments versus rural environments. Another well described environmental risk factor, statistically robust in the face of various confounding factors, is the use of cannabis (Van Os 2011, Minozzi et al. 2010). A meta-analysis by Arseneault et al. (2004) evaluating the data from 5 studies (n= 101497) reported an adjusted OR of 2.34 (95% confidence interval 1.69, 2.95) for the development of schizophrenia if an individual ever used or was dependent on cannabis. Although this work has been criticized for the fact that the same cohort of Swedish conscripts were used twice in the analysis, the data from several other analyses and studies involving more than 160 000 participants support Odds ratios in the region of 2.1 – 2.93 (95% confidence interval range 1.2 - 3.64) (Henquet et al. 2005, Macleod et al. 2004, Semple et al. 2005, Moore et al. 2007).

From these odds ratios and the prevalence of 0.5-1.0% in the population it is clear that the risks conferred by these exposures do not necessarily equate with the development of the disease, and as Van Os (2010) makes clear in a perspective on the interplay between environment and schizophrenia, "[schizophrenia]…may be considered the poor outcome fraction of a truly complex multidimensional psychotic syndrome". This implies that the population may be exposed to these insults (i.e., cannabis use, urban upbringing) and have varying degrees of symptoms present. An individual may present with hallucinations in the presence of an urban upbringing, but may maintain normal functioning and thus "avoid" the diagnosis of schizophrenia. Indeed, Howes et al. (2004) reports that up to 20% of the "normal" population may present with psychotic symptoms in the absence of a diagnosis within the psychosis spectrum. This concept of differential expression of liability is discussed in Van Os (2010) and illustrated in diagrammatic fashion in figures 1.1 to 1.4.

Figure 1.1: A diagrammatical representation of the concept of differential liability
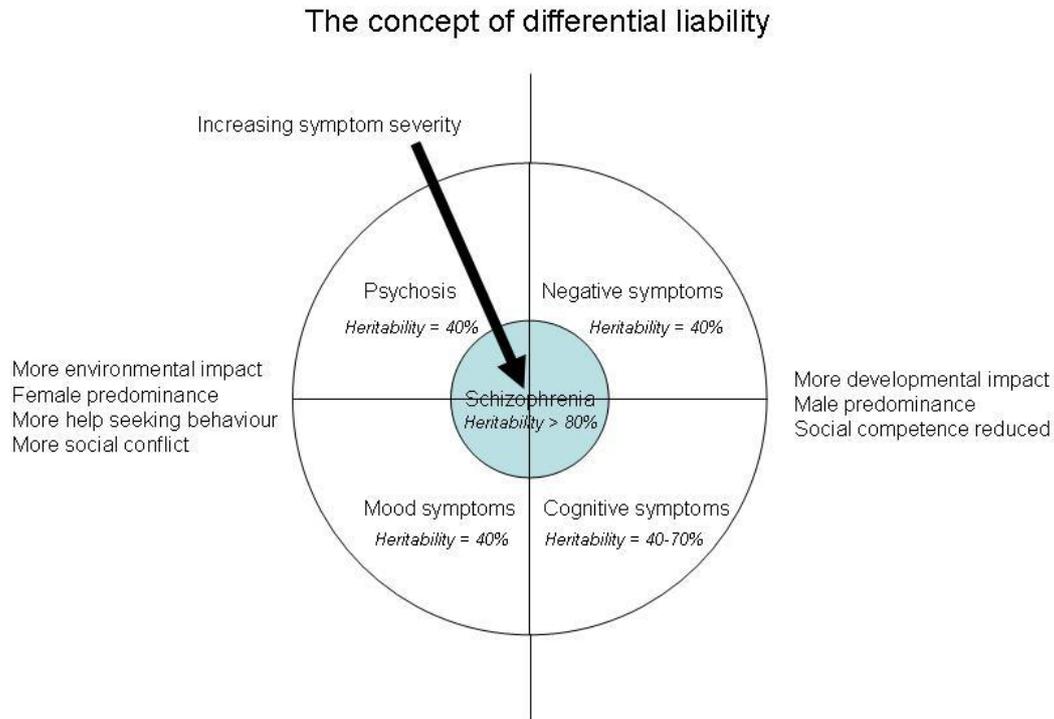
In these visual illustrations, the "zero" level of the clinical features (i.e., symptoms not present) is represented by the outside circumference of the circle. The closer these features are to the inner, blue circle, the more typical they become of schizophrenia, as defined by the various criteria developed as a result of research into the disorder. The diagnosis of schizophrenia is made when the threshold (blue circle) for symptomatology is met and this is clearly related to level of functioning. Each of the domains seem to represent a separate degree of heritability and thus it is understandable that the normal population may have cognitive symptoms reminiscent of schizophrenia (this is often seen in family members of schizophrenia patients) without fulfilling the diagnostic criteria for schizophrenia. It is thus possible for the following scenarios to occur within the normal and ill population:

Figure 1.2: A diagrammatical representation of the concept of differential liability: individual without a diagnosis of schizophrenia

Individual A without a diagnosis of schizophrenia
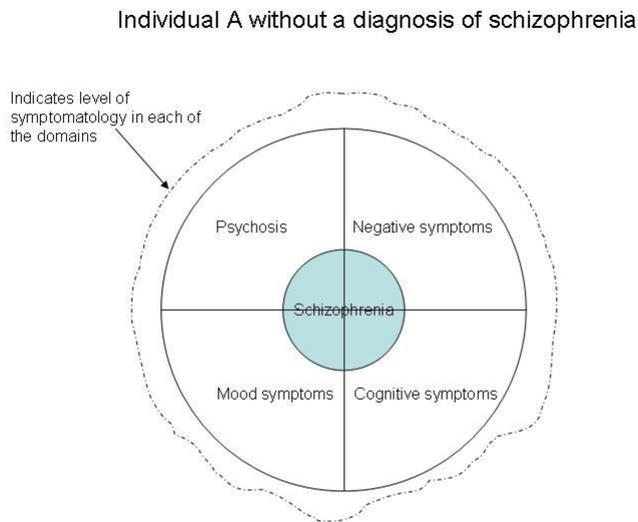
Figure 1.3: A diagrammatical representation of the concept of differential liability: individual without a diagnosis of schizophrenia but with some symptoms of psychosis

Individual B without a diagnosis of schizophrenia or any of the psychosis spectrum diagnoses but with some psychotic symptoms
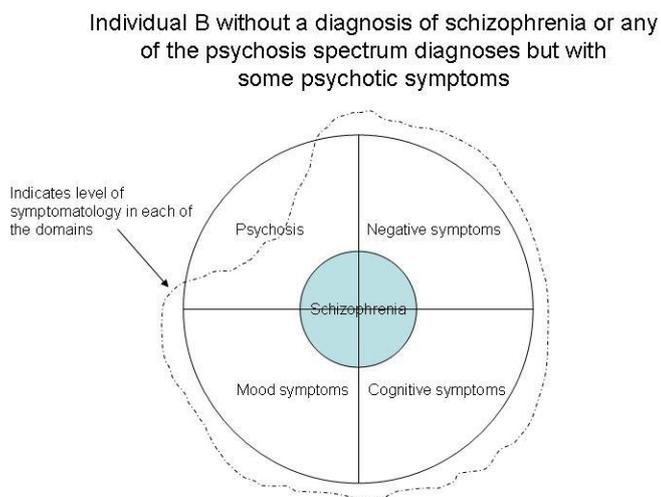
Figure 1.4: A diagrammatical representation of the concept of differential liability: individual with a diagnosis of schizophrenia
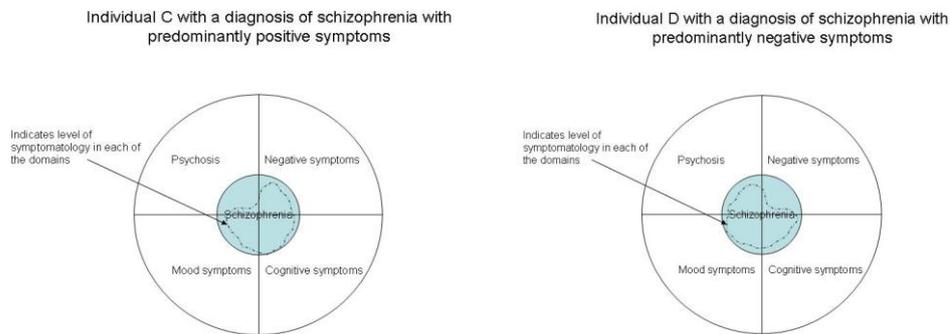
In summary it thus seems that although there seems to be clear evidence of both an environmental and inherited component to schizophrenia, symptoms reminiscent of schizophrenia can occur in the normal population, and that the diagnosis of schizophrenia does not imply homogeneity in the four symptom domains, therefore creating a complex background to the investigation of the genetic underpinnings of schizophrenia.

1.1.2 Genetic Influences in Schizophrenia

As predicted in the previous section, the details regarding influences and interactions occurring at the biochemical and molecular levels are far from clear despite decades of studies designed to determine genetic influences, including classical genetic studies that attempted to investigate patterns of inheritance in relatives of affected individuals such as pedigree analysis, twin studies, adoption studies, and newer advanced laboratory methods used in conjunction with specialized statistical techniques to test for susceptibility loci and alleles (linkage studies, association studies).

In part, the difficulties in identifying the causative heritable factors are linked to the multitude of possible pathological processes at the genetic level. Genetic abnormalities that can give rise to complex diseases range from chromosomal defects (e.g., trisomy 21) which are relatively easy to identify, to differences in portions of the DNA base pair sequences such as insertions, deletions and translocations (copy number variants, or CNVs), and even changes in a single base in the DNA sequence. Where such variations in nucleotide sequences occur with a frequency of 1-2% of the population, they are known as polymorphisms. Where a polymorphism involves a single base pair, it is known as a single nucleotide polymorphism (SNP). Other mechanisms that can also play a role include epigenetic factors (in which the DNA sequences are unaffected, but abnormalities such as methylation of DNA can affect the phenotype) as discussed in the previous section. Furthermore, the genes themselves may not be affected, but problems with RNA transcription, whereby RNA is involved in the formation of proteins according to the DNA template, might result in abnormal proteins and therefore disease. Accurate phenotyping is thus a prerequisite for the accurate identification of candidate genes in affected individuals.

The failure to translate the predicted 80% heritability suggested by a meta-analysis of twin studies (Sullivan et al. 2003) to specific causative genes or variants has been the impetus for the use of "newer" techniques such as the identification of copy number variants, genome-wide linkage studies and exome sequencing (Drögemöller et al. 2011, Wright et al. 2011).

Results from meta-analyses of genome-wide linkage studies clearly illustrate the difficulties.   In two of the studies, 2 very large areas on chromosome 2 (152 Mb and 168Mb, respectively, possibly containing a few hundred genes each), were identified. More detailed studies of these areas revealed no major causative genes, suggesting a complex genetic make-up for schizophrenia (See Ngi 2009 for a full review of findings), where even if areas are identified in genome-wide linkage studies, finding single genes in those areas that make a significant contribution to illness

risk will be difficult. Indeed, more than 7000 genes have been implicated via different candidate paradigms to contribute to the risk for the development of schizophrenia. Smaller genome wide associations studies (GWAS), despite their relatively low power (< 2000 probands per study), also identified a few candidate genes, including plexin A2 (Mah et al. 2006), reelin (Shifman et al. 2008; females only) and zinc finger protein 804 A (Lencz et al. 2010). The latter two findings were replicated, and thus seem to be relatively strong candidates for further research (Girard et al. 2011). Large GWAS studies by the International Schizophrenia Consortium (2009) and Stefansson et al. (2009) supported a link with the major histocompatibility complex (MHC), especially around the *NOTCH4* gene. The ISC data also supported the zinc finger protein 804 A gene as a possible candidate, strengthening support for this gene.

So, although GWAS studies found some overlapping evidence in support of a number of genes that may play a role in schizophrenia, these studies face the same problems as found in association studies: diverse findings, in different cohorts, that cannot be explained only on the basis of differences in the sizes or power of the studies. This suggests that we are still missing a crucial part of the model needed to explain schizophrenia risk in terms of heritability.

 Another approach that has received significant attention given the lack of a full explanatory model provided by the above-mentioned approaches, are copy number repeats (CNVs). Walsh et al. (2008) showed that CNV's larger than 100kb were 3 times (and 4 times for childhood onset) more common in schizophrenia cases than in controls.   In a non-familial group of affected individuals the de novo CNVs were significantly more common (Xu et al. 2008). The CNVs were more commonly found in certain chromosomal areas, such as chromosome 22q11.2, 15q13.3 and 1q21.1.

Taken together, the data to date point to a few chromosomal areas that survive close scrutiny of data and replication studies, and also that SNPs in these areas are likely to have a small effect (OR < 1.2) (Girard et al. 2011). Thus, researchers have also looked at the possibility of large-scale candidate gene sequencing, looking for rare variants that can explain illness in selected families or individuals. One such an example is that of the monoamine oxidase B *(MAOB)* gene where Piton et al. (2010) identified a novel stop codon mutation and three missense mutations in schizophrenia patients and none in the control group. This however accounts only for a very small number of cases and schizophrenia genetics remains a largely unsolved conundrum. It is possible that multiple independent de novo mutations in several different genomic regions lead to schizophrenia (Lee et al. 2012).

The picture is further complicated by the fact that there seems to be a clinical and genetic overlap between schizophrenia and several other disorders, including autism, bipolar disorder and mental retardation (Knight et al. 2009). Jablensky (2010) predicts that the underlying genetic influences in schizophrenia will probably be found to overlap with other disorders, including schizotypal personality disorder and bipolar disorder.

The data from GWAS studies support this possible overlap of genetic factors (ISC 2009, Wang et al. 2010) with similar loci identified across diagnostic boundaries. One example of this is a locus on chromosome 16p13.1, which is considered a shared candidate locus between schizophrenia and mental retardation (Ullmann et al. 2007, Ingason et al. 2011), and the *PLXNA2* gene that is implicated in both schizophrenia and bipolar disorder (Detera-Wadleigh et al. 2009, Gurling et al. 2001). GWASs appear to support the concept that schizophrenia and bipolar disorder represent not two distinct diseases, but different expressions of shared underlying genetic susceptibilities (Tiwari et al. 2010).

Based on the currently available fragmentary data from linkage studies, GWAS and CNVs, it seems as if the risk for schizophrenia is linked to a combination of rare alleles of variable effect size, common alleles with small effect sizes, epistatic interactions and epigenetic mechanisms (see 1.1.2). Derks et al. (2009) point out that the enormous investments at the genotype level are not balanced by significant investments in improved phenotyping, and they propose that latent class factor analysis may provide an avenue to explore improved phenotyping. Indeed in a study on 1888 controls and 2290 patients and with mixed diagnosis of schizophrenia, schizo-affective disorder, bipolar disorder and depression, exploratory analysis suggested five latent dimensions namely positive, negative, mania, disorganization and depression. A latent class analysis of estimated factor scores for each individual yielded eight clusters of subjects and IQ scores were associated with class membership in both controls and patients. These findings further support the notion that careful phenotyping is essential to optimize the power of genetic studies.

 In order to clearly delineate the risk factors for the development of schizophrenia, researchers will need to carefully consider a model that includes the genetic factors, clinical phenotype and the interplay between them. Having shown that schizophrenia's core features are similar across ethnic boundaries (Niehaus et al. 2006, Emsley et al. 2005), the next important question is whether within the Xhosa schizophrenia group, we can identify a subgroup of patients with a distinct phenotype that can lead to the identification of risk factors unique to that phenotype. Lee et al. (2012) strongly propose that future GWAS and CNV studies need large cohorts of patients with "distinct" and "better defined" phenotypes.

1.1.3 Phenotypic heterogeneity of schizophrenia

Attempts to identify specific genetic or environmental aetiologies have been hampered by the considerable heterogeneity of the phenotypic features of the disorder (as illustrated in section 1.1.1). It is important to understand that the current diagnostic system relies on assumptions about diagnostic hierarchy, rules, chronicity and symptom specificity (Peralta et al. 2007). The current diagnostic system that the DSM is based upon implies adherence to decision-making rules that relates a specific group of clinical features to a specific diagnosis. For example, for a diagnosis of schizophrenia to be made a decision rule requires that two or more Criterion A symptoms (delusions, hallucinations, disorganized speech, grossly disorganized or catatonic behavior and negative symptoms) be present for a significant portion of time during a 1-month period (or less if successfully treated). Furthermore, only one Criterion A symptom is needed if the delusions or hallucinations meet further specified requirements (bizarre delusions, commenting voices, and conversing voices). Thus the diagnosis of schizophrenia rests on the subject experiencing one or more specific symptoms over a certain minimum time period. It thus becomes clear that not everyone with schizophrenia will have or need the same set of symptoms to obtain a diagnosis of schizophrenia. Indeed, although one of the lists of symptoms is necessary to make the diagnosis, the presence of the symptom is not sufficient for the diagnosis, but merely indicative of a specific diagnosis of schizophrenia. A specific diagnosis like schizophrenia can thus be seen to be inferred (i.e., the criteria for diagnosing the condition are indicative of an underlying unobserved construct) rather than directly observed, as only the symptoms are observed (Young 1983). In a study by Peralta et al. (2007) the authors compare 23 different diagnostic systems to evaluate whether these systems could predict familial loading and found significant variation between the different systems and even differences in familial loading within systems depending on the broadness of the definition of schizophrenia. Our current systems therefore suggest poor correlation between familial liability and our diagnostic systems.

Several strategies have been employed to reduce the heterogeneity of the phenotypic features of schizophrenia, including the development of rating scales based on psychiatric symptomatology (e.g., the Positive and Negative Symptom Scale for Schizophrenia), the restriction of research samples to homogeneous subgroups (such as those recruited from geographically isolated or culturally distinct racial groups) and the use of endophenotypes (See Niehaus 2005 for a detailed description).

1.1.3.1 Problems with a categorical approach:

Although there is abundant evidence that dopamine dysregulation is the common underlying abnormality in schizophrenia (Stilo 2010), the underlying cause of this is not known. Schizophrenia is thought to be a complex disease with both environmental and genetic components. No reliable diagnostic marker of the disease has been identified and no single genetic abnormality has been linked to all cases of the illness (Jablensky 2010). Concordance rates for schizophrenia between monozygotic twins approach, but do not reach 100%. Furthermore, psychotic phenomena have been shown to occur in so-called 'normal' individuals, without a formal psychiatric disorder as illustrated below (see section 1.1.1):

Figure 1.5: A diagrammatical representation of the concept of differential liability: comparison of individuals with and without psychotic symptoms
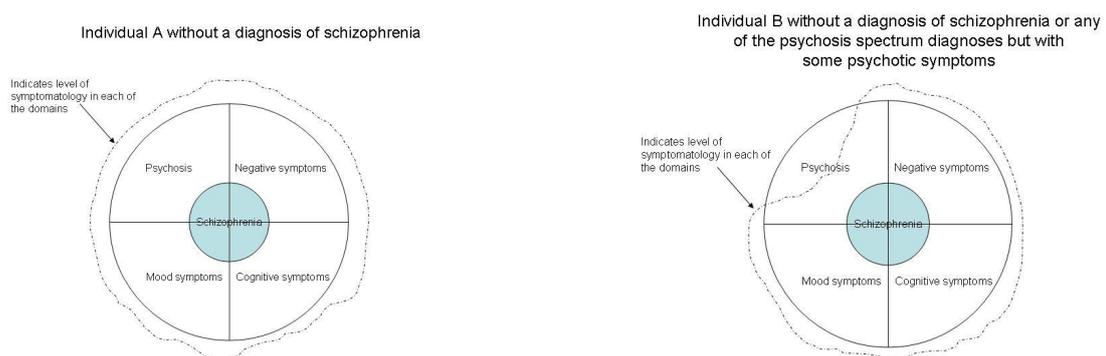
In addition, patients diagnosed with schizophrenia exhibit clinical features that constitute a whole spectrum of symptoms that are shared with other conditions such as bipolar disorder, and might represent one end of a spectrum in which subclinical forms occur (e.g., schizotypal personality disorder, autism, Asperger's syndrome). It therefore seems plausible that several genetic variants, interacting with each other and/or environmental factors, underlie the clinical syndrome known as schizophrenia. Therefore, when searching for associations between schizophrenia and genetic markers, a categorical approach, in which patients are diagnosed as being either schizophrenic or not, might lose valuable information if individual symptoms or symptom clusters are not examined.

One approach to sorting out the dilemma of having to categorize subjects into discrete "types" of schizophrenia has been to develop a theory of dimensionality. Several dimensional approaches have been favored throughout the last four decades, with Strauss et al. (1974) supporting the idea of three dimensions, namely positive symptoms, negative symptoms, and disordered relating. Crow 1980 and 1981 on the other hand suggested only two dimensions namely positive and negative. Indeed a number of scales commonly used today were developed in response to the dimensional theories of schizophrenia. These scales include the Scale for the Assessment of Negative Symptoms (SANS) and the Scale for the Assessment of Positive Symptoms (SAPS) (Andreasen & Olsen 1982, Andreasen et al. 1995b), the Positive and Negative Syndrome Scale (PANSS) (Kay et al. 1987, Bell et al. 1994, Lindenmayer et al. 1994, White et al. 1997), and the Schedule for the Deficit Syndrome (Kirkpatrick et al. 1989, Kirkpatrick et al. 2001).

In order to explore whether the symptoms and signs of schizophrenia can be "dimensionalised", principal component analysis and factor analysis were commonly used (Klimidis et al. 1993,

Brekke et al. 1994, Buchanan & Carpenter 1994, Murphy et al. 1994, Burke et al. 1996, Ratakonda et al. 1998, Mojtabai 1999). Klimidis et al. (1993) very specifically addresses the validity of the Crow model of two subtypes, positive and negative. Following on from a previous study by Minas et al. (1992) where they found that three dimensions (negative symptoms, positive thought disorder and hallucinations/ delusions) best represented the SANS/SAPS items in psychosis, they analyzed the global SANS/SAPS ratings of a diagnostically mixed group of 114 psychotic patients and the data derived from nine published studies that had correlation matrixes or factor analysis for the SANS/SAPS data available (Bilder et al. 1985, Kulhara et al. 1986, Walker et al. 1988, Andreasen & Olsen 1982, Andreasen & Grove 1986, Lezenweger et al. 1989, Fenton & Mcglashan 1991, Arndt et al. 1991). The analysis of this data did not support the two dimensional theory of psychosis. Indeed, even the data of Fenton and McGlashan (1991), which contains patients from the pre-antipsychotic treatment era supports more than two dimensions and thus makes it less likely that treatment effect can be responsible for the dimensionality seen in psychosis.

The interpretation of this type of study is dependent on several components including study design (e.g., sampling method), observation period, which symptoms or signs were reported on, the availability of data sources, and the statistical method used. The use of cross-sectional versus lifetime assessment is intuitively felt to be important in interpreting the results of studies as cross-sectional assessments may be vulnerable to stage of illness (Mojtabai 1999). Indeed, the use of the Operational Criteria Checklist for Psychotic Illness (OCCPI) in subjects with schizophrenia and schizo-affective illness has yielded more complex factor-analytic results (Cardno et al. 1996, Cardno et al. 1999, Van Os 1999). Although the more complex results must be interpreted against the background of the possible impact of a broader range of symptoms as used in this analysis versus using the PANSS/SANS/SAPS in other studies. In these studies the positive dimension separated into three factors namely paranoid delusions, first rank delusions and first rank

hallucinations. Van Os et al. (1999) in a mixed sample of 706 (54% had schizophrenia or schizo-affective disorder) found four factors, namely positive, negative, depressive and manic when using cross-sectional items from the Comprehensive Psychopathological Rating Scale (CPRS), but had an additional factor when they included life-time symptoms (using the OCCPI and SANS and medical record data).

Studies using the PANSS mostly revealed 5 factors (positive, negative, dysphoric, excited and cognitive or autistic pre-occupation (Bell et al. 1994, Lindenmayer et al. 1994, Lindenmayer et al. 1995, White et al. 1997, Lancon et al. 1998)). The use of the individual SANS and SAPS scales in mixed samples (only 35% were schizophrenia) yielded even more factors (5 to 7) including (1) negative, (2) disorganized, (3) disordered relating, (4) bizarre delusions, (5) auditory hallucinations, (6) non-bizarre (paranoid) delusions and (7) other hallucinations (Toomey et al. 1997).

The addition of other indicators such as "social adjustment" has increased the number of factors identified by factor analysis from the usual 3 (positive, negative, disorganized) to 4 (original three plus social relations) (Peralta et al. 1994, Lenzenweger & Dworkin 1996).

Statistical methodology further complicates the picture. McGrath et al. (2004) used latent class factor analysis (LCFA) that differs from traditional factor analysis, as discussed in chapter 2. In this study they assessed 1043 stable in and outpatients (877 with schizophrenia and 166 with schizo-affective disorder) by direct assessment with the Diagnostic Interview Schedule (Robins et al. 1981) or the Diagnostic Interview for Genetic Studies (Nurnberger et al. 1994) and used consensus diagnosis and a diagnostic checklist for DSM-IV or DSM-III-R. They used 44 indicators from these assessment sources for the LCFA. They found five factors namely positive, affective, disorganized, negative and early onset/developmental. There was a weak positive

correlation between the positive and early onset developmental factor scores (r=0.12, p<0.0001) and between negative symptom scores and affective scores (r=0.08, p<0.009) as well between the disorganized and negative symptom scores (r=0.10, p< 0.0007). The use of this method therefore led to the identification of a new factor called early onset/developmental. This factor was characterized by early onset, school deterioration, and attendance at a therapeutic school, seizures and several positive symptoms.

1.1.3.2 Data-driven studies attempting to identify schizophrenia subtypes

1.1.3.2.1 Studies including a wider phenotype

Kendler et al. (1989) have argued for a data-driven approach to the classification of psychiatric disorders as the symptoms and genetic underpinnings of psychotic disorders seem to overlap. There have been two approaches, namely a wider phenotype and a narrow phenotype approach. The wider phenotype approach looked at the diagnosis of schizophrenia based on DSM or ICD criteria and often included other psychotic disorders with or without mood symptoms. The narrow phenotype approach focused on specific symptoms or clinical subgroups within the wider phenotype. We start by looking at the data-driven studies in the wider phenotype approach.

Castle et al. (1994) and Sham et al. (1996) applied LCA in a two step approach on 447 "non-affective functional psychosis" patients from the Camberwell Register First-Contact Sample. They identified three subtypes, namely neurodevelopmental (early onset, poor premorbid functioning, restricted or inappropriate affect, negative features in general, thought disorder, and male gender), paranoid (later onset, milder severity, persecutory delusions, no gender predominance) and schizo-affective (mood symptoms and a family history of affective disorders and a female predominance). Premorbid variables such as personality disorder, obstetric complications and developmental

problems were more commonly associated with the neurodevelopmental subtypes, while family history was more common in the schizo-affective subtype. However, the subtypes did not differ significantly in first rank symptoms. The neurodevelopmental subtype demonstrated the highest frequency of negative symptoms.

The similarities between these findings and that of Farmer et al. (1983) are intriguing and are discussed in more detail in the paper by Sham et al. (1996). Using CATEGO to diagnose patients, Farmer et al. (1983), used cluster analysis in a group of 64 chronically ill (>5 years) schizophrenia patients. The first group, also the largest, was characterized by later onset, good premorbid functioning and well-organised delusions, while the second group had poor premorbid functioning, early onset, a family history of schizophrenia, bizarre behavior, incoherent speech and blunted affect.

Kendler et al. (1997) did latent class analysis on data extracted from the Irish Study of High Density Schizophrenia Families. The sample of 713 participants (256 sibling pairs concordant for schizophrenia (DSM-III-R) and 457 sibling pairs concordant for non-affective psychosis) were assessed with modified sections of the structured clinical interview for DSM-III-R (Spitzer et al. 1987) and the structured interview for schizotypy (Kendler et al. 1989). They identified five classes in the schizophrenia concordant group namely:

Class 1: characterized by high levels of depressive and especially manic symptoms, moderate positive symptoms and relatively few negative symptoms. This class was associated with a later age of onset and better outcome.

Class 2: characterized by high levels of negative symptoms including flat affect, negative thought disorder and catatonic symptoms. The class had low levels of positive and affective symptoms and was associated with poor outcome and course.

Class 3: the class shows high levels of positive and negative symptoms and was associated with an early age of onset, a chronic course and poor outcome.

Class 4: this class shows very little evidence of symptoms in general and was associated with a late onset age and had a relatively good outcome.

Class 5: this class mostly contains clients with catatonic symptoms, negative thought disorder and affect flattening. The illness course was characterized by a remitting and relapsing course.

Of note is the finding that sibling resemblance within classes was most pronounced for class 1 and 5 (p<0.0001, phi=0.46, df=16, X^2=53.22 overall, class 5 X^2=9.1 for individual cells, class 1 X^2=8.3 for individual cells) and this remained significant even after removing the non-independent sibpairs.

Kendler et al. (1998) also analyzed a group of 374 probands from the Roscommon County Case Register with broadly defined schizophrenia and affective illness (using the structured clinical interview for DSM-III-R (Spitzer et al 1987). The latent class analysis revealed six classes:

Class 1(26.2%; classic schizophrenia class): this class was characterized by significant positive and negative symptoms with a relatively low presence of mood symptoms. The class was associated with a chronic course with poor insight, poor premorbid functioning and prognosis (outcome) and low rates of employment (and poor occupational adjustment) and marriage.

Class 2 (20.8%; major depression class): depressive symptoms were common with very little if any positive, negative or manic symptoms. The outcome and course of illness was good with high rates of marriage and employment. Low rates of premorbid personality abnormalities were reported.

Class 3 (18.0%; schizophreniform class) this class showed significant positive symptoms, fewer negative symptoms than class 1, very few mood symptoms, a better premorbid level of

functioning, a good outcome and brief episode duration. In addition, more marriages and a higher employment rate were reported.

Class 4 (17.6%; Bipolar-schizomania class) this class was made up of probands with prominent psychotic symptoms and mood symptoms (both mania and depression). Negative symptoms were present and the course and outcome were benign. Compared to class 1 there was more marriages and a better employment rate and less premorbid personality abnormalities.

Class 5 (14.5%; Schizodepression class) this class had more psychotic symptoms and similar negative symptom load as class 1 and depression symptoms similar to class 2. The course, outcome, marital and employment status and presence of premorbid personality abnormalities were somewhere in-between class 1 and 2.

Class 6 (3.6%; Hebephrenia class) this class resembled class 1 with the difference being that individuals in this class were likely to have positive thought disorder, inappropriate affect and other symptoms suggestive of mania.

This study was followed by a LCA design study in 660 psychotic inpatients (DSM-IV and ICD 10 diagnostic systems used) by Peralta and Cuesta (2003). The LCA was applied to 16 index episode and 16 lifetime symptom ratings.

Class 1 (schizophrenia – present in both index and lifetime LCA): this class had high rates of positive symptoms (psychosis and disorganization), negative symptoms, poor insight and residual symptoms.

Class 2 (psychosis- present in both index and lifetime LCA): this class dominated by hallucinations, delusions and poor insight.

Class 3 (schizomania – in index LCA and schizobipolar in lifetime LCA) showed psychosis, disorganized behaviour, manic symptoms, poor insight and an acute onset. The schizobipolar class also had depressive symptoms.

Class 4 (schizodepression in index LCA and lifetime LCA): this class was characterized by significant psychotic symptoms, negative symptoms, depressive symptoms and moderate residual symptoms.

Class 5 (cycloid psychosis in index LCA and atypical schizophrenia in the lifetime LCA): this class had a mixture of most symptoms with the index episode LCA class showing a more acute onset, lots of confusion symptoms and few residual symptoms.

Of these classes, the schizophrenia class showed the best nosological stability across the classification systems used in this study.

Using The Comprehensive Assessment of Psychiatric History (Andreassen et al. 1992), Boks et al. (2007) interviewed 1056 patients referred to a hospital in Utrecht (Netherlands) with psychosis. The sample included all patients referred with psychosis, whether further assessment confirmed a lifetime experience of psychotic symptoms or not. Schizophrenia was diagnosed in only 597 (57%) of the total sample, other diagnoses included bipolar disorder, major depression and other psychosis spectrum disorders.

Boks et al. (2007) generated a 6 class best fit model based on the symptom complexes extracted from a factor analysis performed on the sample:

Cluster 1 (33.1%): highest number of negative symptoms, highest number of mania symptoms, 2nd highest number of disorganization symptoms and second highest number of depression symptoms. Sixty-one percent of the bipolar disorder I patients and 52% of the schizo-affective disorder patients were represented in this group. This cluster seems similar to the bipolar-schizomania (Kendler et al. 1998) or schizobipolar subtypes (Peralta & Cuesta 2003) identified in other studies.

Cluster 2 (26%): low numbers of disorganization symptoms, substantial negative and the highest number of depression symptoms and some positive and mania symptoms. Sixty-seven percent of the 'psychotic depression' patients fell into this group. It had a significantly higher percentage of females (Chi square=36.8, df=5, p=0.001) and the second highest number of married people. The second cluster, with the highest depression score, shows considerable overlap with the schizodepression groups of previous LCAs (Peralta & Cuesta 2003, Kendler et al. 1997, Kendler et al. 1998).

Cluster 3 (14%): highest number of disorganization symptoms, positive symptoms, second highest number of negative symptoms and almost no depression or mania symptoms. Forty percent of the patients with 'schizophrenia-disorganized type' were classified in this group. This subgroup had a significantly lower level of education and suffered the highest number of episodes, although the latter may be due to their higher age. The third cluster, which included patients with the highest number of disorganization and positive symptoms and a limited number of mood symptoms, resembles the "hebephrenia" group of Kendler (Kendler et al. 1998).

Cluster 4 (14%): low numbers of disorganization symptoms, substantial negative and positive symptoms, and a low number of mood symptoms. This group included 37% of the catatonic schizophrenia patients. The fourth cluster shows many similarities with the classic "schizophrenia" groups of previous LCAs (Peralta & Cuesta 2003, Kendler et al. 1998).

Cluster 5 (7%) had very few if any, psychotic symptoms. Only 2% of the schizophrenia patients and about 50% of the miscellaneous patients were found in this group. This group had significantly fewer episodes compared to all other groups and seems to represent a combination of individuals with no or very few psychotic symptoms.

Cluster 6 (6%): This group represented 50% of the depression patients and 30% of the delusional patients and cluster members mostly had depressive symptoms. Membership of this group was associated with a significantly later age of onset and a greater probability of being married compared to the other groups. Cluster 6 seems to mirror the "major depression" subtype of psychosis as described by Kendler et al. (1998).

Boks et al. (2007) then followed up the LCA with a meta-analysis of the relative importance of the symptom dimensions in the identified LCAs in delineating the psychosis subgroups (See Table 1) and showed that mania was the most important symptom in discriminating between subgroups of psychosis. The McGrath et al. (2004) (N=1043) data was not included, most likely because it did not find a mood/mania class but rather positive, negative and disorganization and an early onset/developmental factor. The depression and mania subgroups have relatively large standard deviations and this suggests sensitivity to the inclusion criteria of studies.

Table 1.1: Meta-analysis of relative importance of the symptom dimension in the current and previous LCAs to the delineation of psychosis subgroups

	Disorganization	Positive	Negative	Depression	Mania
Boks 2007 (N=1054)	0.336	0.160	0.337	0.348	0.569
Kendler 1998 (N=343)	0.290	0.239	0.313	0.678	0.759
Kendler 1997 (N=580)	N/A	0.080	0.341	0.088	0.320
Peralta 2003 (N=660)	0.263	0.015	0.467	0.297	0.192
Mean (SD)	0.296 (0.03)	0.123 (0.08)	0.364 (0.05)	0.353 (0.16)	0.460 (0.20)
McGrath 2004 (N=1043)	Excluded from meta-analysis.				

Presented numbers are R^2.
Kendler '98 (Kendler et al. 1998).
Kendler '97 (Kendler et al. 1997).
Peralta (Peralta & Cuesta, 2003).
Data for this table extracted from Boks et al (2007).

Negative symptoms were one of the three most discriminating symptoms in Boks et al. (2007) and Kendler et al.'s (1998) studies, and in the meta-analysis by Boks et al. (2007) proved to be very important. As the two classes with absence of negative symptoms (Boks et al. 2007) were also the classes with the lowest number of schizophrenia subjects, it is conceivable that negative symptoms can be seen as one of the core symptoms of schizophrenia that can help to delineate it from bipolar or depression in terms of underlying pathophysiology. Negative symptoms have been the focus of renewed efforts to clearly define their role within schizophrenia and serves as a good introduction to the narrower phenotype approach.

1.1.3.2.2 Studies focusing on a narrower phenotype

It is important to understand that negative symptoms have been defined and classified differently over the past few decades and one such classification is that of deficit syndrome. Deficit syndrome is described as enduring and persistent negative symptoms (Carpenter et al. 1988) and occurs in approximately 15% of first-episode patients and 25-32% of patients with chronic schizophrenia (Kirkpatrick et al. 2001; Peralta & Cuesta 2004). However, diagnosing deficit syndrome is complicated, as it requires the researcher to distinguish between primary (part of core pathology) and secondary (due to illness process) negative symptoms (Kirkpatrick et al. 1993). Although Boks et al. (2007) suggests a role for negative symptoms in subtyping schizophrenia, a study by Peralta et al. (2007) in 358 schizophrenia probands showed that the deficit/negative subtype diagnosis did not contribute to a better delineation between more and less familial cases. Negative symptoms (including deficit symptoms) have been viewed in two ways: a categorical approach or

a dimensional approach and support exist for both of these (Carpenter et al. 1988, Waller et al. 1998). Blanchard et al. (2005) thus used a latent class model and compared that to a taxonic technique in 238 schizophrenia patients in order to see which model best described negative symptoms. Blanchard et al. (2005) found that patients belonging in the negative symptom taxon-based group were more likely to be male and show poorer function than those in the non-taxon based group. As they found no difference in psychotic and affective symptoms, they concluded that a categorical approach to negative symptoms was more likely to be a true representation. These findings were supported in part by a study on the latent structure of negative symptoms in 238 schizophrenia subjects by Blanchard et al. (2005). They reported that male gender and poorer social functioning were more likely in the negative symptom latent class. It is not clear whether these findings can be generalised to other population groups and independent replication of these results in ethnically homogenous samples is indicated.

Catatonic schizophrenia has been considered part of the classification of schizophrenia from the first descriptions of Kraepelin (1899) of dementia praecox, despite ongoing controversy as to its existence as a separate clinical or biological entity (Caroff et al. 2007, Northoff 2002). Ungvari et al. (2009) applied LCA on a sample of 178 Chinese inpatients with schizophrenia (DSM-IV) in long-stay psychiatric wards that had at least one sign of catatonia (rated with a modified version of the Bush-Francis Catatonia Rating Scale). The LCA found a four-class solution that included:

Class 1 (18%): "Automatic" phenomena – automatic obedience, Mitgehen, waxy flexibility.

Class 2 (39.4%): "Repetitive/echo" phenomena – perseveration, stereotypy, verbigeration, mannerisms, grimacing.

Class 3 (20.1%): "withdrawal" phenomenon – immobility, mutism, posturing, staring, withdrawal.

Class 4 (22.5%): "agitated/resistive" patients – excitement, impulsivity, negativism, combativeness.

This study enabled the researchers to classify chronic schizophrenia patients with catatonia into four distinct syndromal groups that more or less correspond to our clinical experience. Of interest to our study is the finding of an association between the 'withdrawn' catatonia class and the covariates higher SANS scores and Parkinsonian symptoms (specifically akinesia) and poor social adaptation. It raises the question of whether negative symptoms (as reflected by the higher SANS score) and the withdrawn catatonia class are a similar disease entity, looked at from different angles or whether they are independent syndromes with a common endpoint (Liddle 1991, Ungvari 2009).

Despite the promising findings in these studies, the reader should be aware that even in narrow phenotype approaches, the question still remains whether the "cut-off" point for pathology is correct. Although this question does not form an integral part of the study, the following section will provide some background to this issue and the results from any of the studies in phenomenology should reflect on this.

1.1.3.2.3 The problem of normality versus pathology in data-driven approaches

It is important to note that although the studies discussed thus far focus on the identification of subgroups of psychosis, it is also noteworthy that a continuum hypothesis has considerable support. Large nationally representative samples reveal that significant percentages of the population experience symptoms similar to clinically defined psychosis, such as hallucinations. Seventy-five percent of study participants in England and Wales reported hallucinatory experiences but did not meet the diagnostic criteria for psychotic disorder (Johns et al. 2004), while 17.5% of Dutch participants endorsed at least one psychosis screening item although the

prevalence of psychosis in the Netherlands is only 2.1% (Bijl et al. 1998). The argument behind the continuum model does not fall within the ambit of this discussion and the reader is referred to Krabbendam et al. (2004) for additional background information. However, the use of latent class analysis in a sample drawn from the general population based on the assumption of continuity of psychosis or psychotic symptoms (as opposed to discrete categories) is of interest as one could argue that the continuum model of psychosis will extend throughout the spectrum of pathology as well, and that psychosis in terms of subtyping might be more than just the absence or presence of a symptom, but also a representation of the degree and chronicity of the specific symptom.

Shevlin et al. (2007), using The National Comorbidity Survey (NCS) data on 5893 (full data available) of 8000 participants across 48 coterminous states of the USA, found that a four-class solution was considered to be the best model. The four classes were labelled a psychosis class, a hallucinatory class, an intermediate class, and a normative class. The psychosis class was the smallest (1%) and members endorsed all the psychosis items but had a very low chance of endorsing hypomania. Class two (hallucinatory class) had a lower propensity for endorsing thought disorder items, but hypomania was endorsed more commonly and paranoia had a very high chance of endorsement in this class. Class 3 differed from class two mainly in the very low endorsement of paranoia. The fourth class was the biggest with almost no probability of endorsing the items of interest. Within latent classes one to three, there was a declining gradient in the odds ratios for associations between class membership and various demographic risk factors, clinical variables and a history of traumatic events. This might suggest a continuum of psychosis from frank psychotic symptomatology to near normalcy. In addition, the lower age categories (16-24 years) were more likely to be in the psychosis group.

In summary, LCA and other techniques used to subtype schizophrenia have revealed that there is most likely a continuum of psychosis, from very mild in the "normal" population to more severe

in "psychiatrically ill" populations. Looking at the published studies using LCA it becomes evident that despite differing methodologies, several dimensions of psychosis are evident. However, it also becomes apparent that samples used in these studies were mostly of "mixed" diagnosis and thus a "pure" schizophrenia sample is rare. It seems from the literature that authors preferred investigating psychosis as a general concept and not schizophrenia as a diagnostic entity. The question that remains is whether studies focusing on a broader definition of psychosis versus a narrow definition of schizophrenia will show the same dimensional outcome. This question is of particular interest as most of the LCA studies had mood symptoms as a separate dimension or part of a dimension.

1.2 Schizophrenia in African subjects

1.2.1 Studies on the phenomenology of schizophrenia in African subjects

The Xhosa, an African population forming part of the larger Nguni language group, have been the focus of ongoing efforts to identify intermediate endophenotypes of schizophrenia (Niehaus et al. 2005, Koen et al 2012 – in press). Clinical intermediate endophenotypes representing "simpler" bio-behavioral characteristics may be helpful in identifying a more homogenous subgroup of Xhosa schizophrenic patients, thus reducing the complexity of the genetic analysis (Gottesman & Gould 2003).

Two studies reported on clinical variables that might be helpful in subtyping schizophrenia in Xhosa patients (Emsley et al. 2001, Niehaus et al. 2005). Emsley et al. (2001) performed a principal component analysis on a heterogeneous sample of 422 Xhosa subjects with schizophrenia, who were recruited from in- and outpatient hospital services and community clinics throughout the Western, Southern and Eastern Cape Provinces of South Africa. The primary

variables of interest were the clinical ratings of the SAPS (Schedule for the Assessment of Positive Symptoms) and SANS (Schedule for the Assessment of Negative Symptoms) (Andreassen et al. 1983, Andreassen et al. 1984). This study yielded a five-factor solution for both global and individual items, respectively, accounting for 55% of the variance. The five factors (extracted from the global items of the SAPS and SANS) were negative symptoms, psychotic symptoms, disorganization, impaired attention, and alogia. The individual items identified were diminished expression, disordered relating, psychosis, thought disorder and bizarre behaviour. The findings were similar to those reported in Caucasian samples.

A subsequent investigation of the subjects studied by Emsley et al. (2001) was conducted by Niehaus et al. (2005). The sample was larger (n=514), additional subjects having been recruited in the interim.   All the SANS and most SAPS items (alogia and concentration difficulties having been excluded based on findings from the Emsley [2001] study) as well as clinical variables assessed by the Diagnostic Interview for Genetic Studies (DIGS) (e.g., age of onset, duration of illness) were first subjected to FA to identify possible differences, if any, from the original (Emsley et al. 2001) schizophrenic subjects. They obtained virtually identical results, but without separation between the negative symptom and disorganization domains. This difference may reflect the expansion and stratification of the sample.

The sample was divided into two groups, a nonsib-pair group (i.e., subjects who had no other siblings affected with schizophrenia) and a group consisting of affected sib pairs selected from 104 sibships (100 pairs, 2 trios, 2 fours), only the first two individuals assessed being included as a sib pair.

Concordance analysis of the dichotomously defined SANS and SAPS items revealed higher than expected concordance for forty individual items, mostly from the SAPS. The global items

hallucinations (p=0.002), delusions (p=0.01) and anhedonia (p=0.037) had higher than expected concordance.   To further explore the group, items that were concordant were analyzed to determine whether spurious differences between concordance results in the two subgroups could have been caused by confounding factors (such as current age, duration of illness and age of onset) or by differences in the prevalence of certain items, such as gender.

Gender differences in several aspects of schizophrenia have been extensively reported (Häfner et al. 1989). Since the sample was predominantly male, to control for the effects of gender further analyses were restricted to the men only. Only four items remained concordant between sib pairs: eye contact (p=0.027), auditory hallucinations (p=0.010), global hallucinations (p=0.017) and delusions of control (p=0.001).   Even after controlling for 9 possible confounding variables, (including, among other factors, gender, substance abuse/dependence and the presence of stressors prior to illness onset) these four items still showed higher than expected concordance between the sib pairs. These four items (eye contact, auditory hallucinations, global hallucinations and delusions of control) therefore appear to be likely candidates for closer scrutiny in genetic analyses of the Xhosa sample. Niehaus et al. (2005) provide evidence in support of each of these items.

The core question to this study is whether these items can serve as endophenotypes for schizophrenia in the Xhosa population. These endophenotypes can assist us in re-examining genetic data on this sample. The following section explains the rationale for endophenotypes and illustrates the value of these in genetic studies.

1.3 Endophenotypes

1.3.1 Endophenotypes as markers of disease risk

Endophenotypes of a disorder are intermediate markers of disease risk that fall somewhere between an individual's genotype and the final, phenomenological expression of the disease. They are heritable factors that confer vulnerability to a disease, but are easier to study than the actual genetic makeup of patients and might be simpler to characterize than the many, often nonspecific, symptoms that have been designated as diagnostic criteria for psychiatric disorders. They may be biochemical, neuroanatomical, genetic, psychobiological or even neurophysiological. Allen et al. (2009) in their review identified several possible endophenotypes of psychiatric disorders including structural and functional brain abnormalities (e.g., ventricular, planum temporal and superior temporal gyrus volume abnormalities and differences in fMRI activation), sensory processing and event-related potential measures (pre-pulse inhibition, P50-, P300 and N400 event-related potentials), neuromotor abnormalities (smooth pursuit eye movement, saccadic eye movement, handedness and neuromotor deviations), neuropsychological measures (including performance on the Wisconsin card sorting, continuous performance task and visuospatial delayed response), physiological abnormalities such as niacin flushing and minor physical anomalies (including dysmorphology). Indeed, they report that these abnormalities seem to be present in schizophrenia much more frequently than would be expected by chance alone. The authors conclude that further studies of endophenotypes and more specifically the prevalence of the endophenotypes will contribute significantly to better understanding the aetiology of schizophrenia.

During the past few decades there has been an upsurge of interest in determining the prevalence of schizophrenic symptoms or symptom clusters, in the identification of subgroups based on symptomatology, and also in the possibility of utilizing symptoms or combinations of symptoms of schizophrenia as endophenotypes. Studies using variable reduction techniques such as factor analysis (FA) generally agree that the symptoms of schizophrenia fall into between 3 and 5

categories, the strongest support being given to a five-factor model, the most commonly cited domains being: positive symptoms, negative symptoms, a disorganized domain, cognitive changes and mood components (albeit in more broadly defined "psychosis" samples). The factors derived from these studies have not yielded reliable endophenotypes in terms of genetic perspectives. Factors identified by FA cannot be directly measured in patient samples and can therefore not be subjected to validation of their ability to allocate individuals to homogeneous subgroups which may be used as endophenotypes of the disorder.

1.3.2 The search for endophenotypes of schizophrenia.

 In order to be called an endophenotype, a characteristic should be associated with a specific illness, should be trait-linked and not state dependent (in other words – stable over time, present during remission and independent of stage of illness). The characteristic should also identify individuals at risk and family studies should support genetic transmission. However, if this characteristic is linked to latent genes, the characteristic must also be present in unaffected family members (Calkins et al. 2008). The value of finding universal subtypes that span ethnic boundaries lies in their application to genetic studies. Indeed, subtypes such as deficit schizophrenia have formed the basis for small initial analyses of genetic variants in the GNAS1, COMT, DTNBP1, G72/G30, RGS4 and PIP5K2A genes (see Galderisi et al. 2009 for a review of these findings). The use of LCA based phenotypes to investigate subtypes of schizophrenia is illustrated by Leask et al. (2009) who took the concept of the use of latent class analysis and tested the idea that previously identified risk factors of small effect will have a more significant effect in more homogenous phenotypes based on symptom dimensions. Two hundred and fifty six subjects from the National Child Development Study in the United Kingdom were grouped based on data-derived subtypes and effect size of neurodevelopmental risk factors compared between DSM-IV

derived and data-derived subgroups. They found that the data-derived subgroup lacking affective symptoms were less heterogeneous in terms of neurodevelopmental risk factors such as birth weight, cognition, childhood behavioural problems, and neurological soft signs, including handedness.

Identifying reliable subtypes or more preferably, endophenotypes, remains the main aim of researchers in phenomenology of schizophrenia.   In this regard the use of homogeneous populations have the advantage that the genetic variability is lower and the cultural and environmental factors more limited than that of more cosmopolitan populations.   Arajarvi et al. (2004) used this argument to study the clinical phenotype of schizophrenia in a Finnish isolate. The Finnish population has been relatively isolated for the past 2000 years and the researchers focused on an area in the North of Finland founded by 40 families at the end of the 17th century. In this isolate they identified 588 patients of which 112 were singleton patients with schizophrenia and 78 from multiplex families and compared these to 466 multiplex families from other parts of the country. A factor analysis revealed a four factor solution with factors mainly characterized as:

Factor 1: delusions and hallucinations (20.3% of variance explained).

Factor 2: "manic" symptoms and behavior (13.3% of variance explained) – included pressured speech, thoughts racing, elevated mood, irritable mood, increased self-esteem and grandiose delusions.

Factor 3: "negative" symptoms (explained 11% of variance) – included catatonia, speech difficult to understand, negative formal thought disorder and inappropriate affect.

Factor 4: "depressive" symptoms (explained 10.2% of variance) – included slowed activity, loss of energy and tiredness, loss of pleasure, excessive self-reproach, and suicidal ideation.

This factor analysis showed no gender differences or age of onset correlations. Interestingly, the patients from multiplex families were more likely to have slowed activity and less widespread

delusions, while the singleton affected had more delusions of influence and less negative formal thought disorder and restricted affect. However the classification tree analysis (looking at the pattern of symptoms) showed no significant differences between the groups.

This finding, based on the use of familiality of schizophrenia, to a certain extent underscores Kraepelin's (1921) notion that "the occurrence of mania or melancholia in parents or brothers and sisters will point in this direction [i.e. discriminating between dementia praecox and manic-depressive insanity], though certainly by no means absolutely". Kendler et al. (1987) simulated diagnostic uncertainty and the impact of thereof on familial aggregation of illness and came to the sobering conclusion that misclassification can lead to significant error in the interpretation of family studies. Thus careful phenotyping may reduce the signal-to-noise ratio in linkage and association studies (Fanous et al. 2008) but will not necessarily provide a clear-cut distinction between subtypes of schizophrenia or even the boundary between schizophrenia and other psychotic disorders. Indeed, Kendler et al. (2000) and Chiu et al. (2002) reports that clinical features in families with specific genomic risk areas differ from other families linked to other areas. In addition, a higher level of negative symptoms has also been linked to high-risk alleles in the dysbindin 1 gene (Fanous et al. 2005, DeRosse et al. 2006). Fanous et al. (2008) used latent class analysis in 755 Irish individuals with psychosis based on variables derived from the OPCRIT (Operational Criteria Checklist for Psychotic illness) (McGuffin et al. 1991). They arrived at a six class solution labeled bipolar, schizoaffective, mania, schizomania, deficit syndrome and core schizophrenia. The latter four classes had a prevalence of more than 0.08 and were used to test individually for linkage in a genome-wide scan. The class called "mania" was characterized by the presence of mania symptoms and early age of onset. The "schizomania" class was characterized by delusions, hallucinations and mania symptoms. The "deficit syndrome" class was characterized by negative symptoms and an early age of onset while the "core schizophrenia" class had delusions, hallucinations, negative symptoms, schneiderian symptoms and

premorbid/social dysfunction. Of significance is that the mania and schizomania class achieved minimum criteria for a specific area 20p13-12.2 while the deficit syndrome class was linked to an area 28cM centromeric to 20p13-12.2. This is important for future studies as these areas previously did not show significant linkage when data-driven phenotyping was not utilized. However, there is support for this area in bipolar mood linkage studies (Detera-Wadleigh et al. 1997). In addition to this finding, a close inspection of the data indicates that a majority of patients in the deficit (83.3%) and core schizophrenia (94.08%) classes had a DSM-III-R diagnosis of schizophrenia versus 67.98% in the schizomania class and 35.25% in the mania class.

In another candidate gene study, on Finnish schizophrenia families, Wessman et al. (2009) used a clustering technique to see whether this could reveal more consistent results in terms of the risks for schizophrenia associated with the genes DTNBP1, DISC1, and NRG1.   The three class clustering in the 904 individuals from 288 families revealed a cluster described as psychotic disorder with mood symptoms (n=172), a core schizophrenia class (n=223; characterized by prominent negative symptoms, general cognitive impairment and less mood symptoms) and a class with absence of psychotic disorder (n=509). Of note is that the core schizophrenia class was associated with the DTNBP1 gene.

Now that we understand the potential value of endophenotypes, it is important to understand the basis of the statistical method used to explore clinical symptoms and signs for the presence of an endophenotype. In the next section we briefly explore the statistical methods available for use in this regard.

1.4 Statistical methods for subtyping psychiatric conditions.

Reducing the phenotypic heterogeneity found in schizophrenia is an important step in the search for genetic determinants of the disease. Factor analysis and latent class analysis are two methods commonly employed. In both methods, responses (also known as observed, manifest or "indicator" variables) to items such as symptoms elicited using structured rating instruments are subjected to statistical analyses. Both methods assume that "hidden" constructs underlie the manifest data; in the former these are known as factors and in the latter, latent classes. In factor analysis the degree of correlation between variables allows the researcher to estimate a smaller number of factors (new composite variables) that group together variables in a meaningful way. It is assumed that underlying the various observations there are unobserved or "latent" factors that explain the associations between the observed variables. FA was the method employed by Niehaus et al. (2005) in a prior study of schizophrenic subjects.

LCA, on the other hand, is a statistical modelling technique used to categorise similar individuals into subtypes (latent classes) using observed categorical indicator variables evaluated in the individuals under study. Typically, measurement instruments are used to rate the presence or absence of a number of symptoms (e.g., in the case of psychiatric rating instruments, items such as auditory hallucinations, affective non-responsiveness, etc.). Each subject has a particular response pattern to the items and LCA seeks to group individuals with similar response patterns into homogeneous classes, which can be interpreted as subtypes. The model allocates a probability to an individual of belonging to a class. Latent class analysis can include continuous variables as covariates. The influence of various grouping variables (e.g., gender, substance abuse) on latent class allocation can also be examined.

It is important to understand the differences between factor analysis and latent class analysis and the specific role of each (See Table 2). Factor analysis can be seen as an effort to identify a factor structure that accounts for linear relationships among a set of variables and can thus be seen as a variable-oriented approach. In contrast, LCA is often seen as a person-centred approach, meaning that the focus is the person's unique pattern of characteristics and therefore focuses largely on identifying subtypes of individuals with similar patterns.

The nature and the distribution of the latent variable are the core differences between these two models (see Table 2 for a comparison between the models, Hagenaars 2002, Schumacker 2004).

Table 1.2. Comparison of latent class analysis and factor analysis

Latent Variable	LCA	FA
Nature*	Categorical	Continuous (dimensional)
Distribution	Multinomial	Normal
Basis for assigning names to subgroups	Estimated Conditional probabilities (Ref: www.statisticalinnovations.com/ articles/sage11.pdf)	Factor loadings
Latent variable	Categorical	Continuous
Function of latent variables	Yes	Yes
Function of error (latent variable)	Yes	Yes
Starting point	Contingency table	Correlation matrix

*Categorical data defined by Ruscio and Ruscio (2008) as data where "qualitative differences exist between groups of people or objects" and Continuous data where "people or objects differ quantitatively along one or more continua"

Having considered factor and latent class analysis, the question now remains: Is the underlying structure of schizophrenia categorical or dimensional? To address this important issue, we need to revisit some of the basic principles of latent class and factor analysis.

If we propose a categorical underlying structure for schizophrenia, as supported by latent class analysis, individuals can be categorised into their most likely group based on the observed symptoms (i.e. response pattern of SAPS or SANS items). As latent classes can be interpreted as diagnostic subtypes, persons would fit into diagnostic categories that indicate whether a person is affected or falls within a specific subtype. A categorical structure would of course meet clinical needs as it will allow for tailored treatment or genetic subgroup approach to a specific category or subtype. Indeed, it implies that all individuals in a class have the same item probability and that the symptoms are independent within a class (The diagram below shows a circle indicating the latent variable, measured by four items).

Figure 1.6: Model diagram of LCA
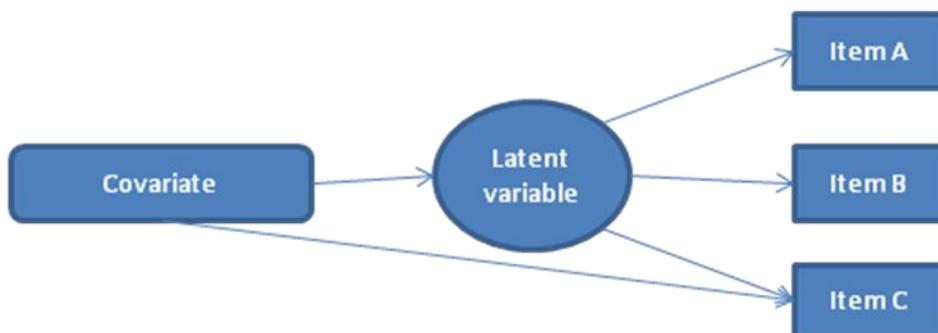
However, categories cannot give us an indication of the range in severity and impairment within and across diagnostic classes. The latter problem can be addressed by the use of factor analysis.

If we propose a dimensional structure, the illness is represented as a continuous distribution, with each individual having different degrees of severity of the disorder. Although this allows for

greater statistical power and a more precise measure of severity of illness, which can be represented as a quantitative score, the clinical needs value is lower in terms of tailored treatment or genetic subtypes. In factor analysis, factors are used to model the correlations among the symptoms and each factor represents an underlying dimension of the disorder. The problem is however that there is generally no obvious way to classify individuals into groups and thus the clinical utility is lowered.

Ideally, one would like a solution that takes into account both these approaches: namely, an approach that allows the underlying structure to be both categorical and dimensional, giving rise to the scenario where individuals can be classified into diagnostic groups, but that also allows for variation in the severity of the disorder/symptoms within a group. One such method is called factor mixture modeling (FMM) (See chapter two for detailed discussion on the principles of FMM). In factor mixture modeling, individuals can have different factor values within a class, which allows people to vary in their conditional item probabilities. The latent class and factor analytic models can thus be seen as special cases of the FMM, with constrained parameters.

Given the properties of FMM it may be an appropriate method to investigate disorder subtypes in mental illness, can assist in exploring diagnostic boundaries (is schizophrenia one multi-faceted illness or several distinct illnesses) and help resolve the continuity issue in psychiatry (the "taxon" question).

Indeed, this method has been used in the validation of the DSM-5 criteria for autism spectrum disorder (Frazier et al. 2012). In this study the primary aim was to evaluate the validity of proposed *DSM-5* criteria for autism spectrum disorder (ASD). The authors used data generated from symptom analysis on 14,744 siblings (8,911 ASD and 5,863 non-ASD) included in the Interactive Autism Network. The structure of autism symptoms (measured using the Social

Responsiveness Scale and the Social Communication Questionnaire) was examined using latent variable models. They illustrated that a hybrid (mixture) model that included both a category (ASD versus non-ASD) and two symptom dimensions (social communication/interaction and restricted/repetitive behaviours) was the most parsimonious of the tested models and replicated across measures and subsamples. They then used empirical classifications from this hybrid model to illustrate that the model closely mirrors clinical ASD diagnoses (90% overlap). This implied that a broad ASD category exists that is distinct from non-ASD. This approach supported the validity of the proposed *DSM-5* criteria for ASD as provided in Phase I Field Trials criteria and suggests a categorical ASD diagnosis with two dimensions within it.

The use of FMM in anxiety has yielded similarly revealing results. It has long been believed that health anxiety as a construct varies in degree along a continuum. In other words there is no clear distinction between nonpathological versus pathological classes or taxons (Ferguson 2009; Longley et al. 2010). Asmundson et al. (2012) evaluated the latent structure of health anxiety using factor mixture modelling (FMM) on anxiety symptom data (used indicator symptoms) obtained from the Illness Attitude Scales (IAS) administered to 1768 university undergraduate students. FMM of these indicator symptoms proposed two classes: (a) an "anxious" class (81.4% of the sample; primarily a somatic focus and symptom interference) and (b) a "non-anxious" class (18.6% of the sample; low scores on all indicator symptoms). These results contradict directly the previous continuous outcome of other more restrictive statistical models and suggested taxonic or categorical approach to be more suitable for health anxiety.

The FMM approach has also been extended to other areas of psychiatric illness including substance use disorders and depression. Sunderland et al. (2013) evaluated the latent structure of DSM-IV major depression in a large 1165 patient strong treatment-seeking population undergoing

online treatment for depression. A two-factor model (represented a psychological and somatic factor) fit the data best when compared to a one-factor model, latent class models, and factor mixture models. The psychological factor (include the following items lack of interest/pleasure, B) depressed mood, C) worthlessness/guilt, D) suicidal thoughts. The latent dimension represents the 'psychological depression' factor) and the somatic factor (include the following items sleep disturbance, fatigue/lack of energy, eating disturbance, trouble concentrating, psychomotor disturbance) suggests that the structure of depression consists of two underlying dimensions.

However, a note of caution: The interpretation of FMM is tempered by uncertainty as FMM can actually describe data better in the absence of true groups if the factors is non-normal in distribution, the data non-linear or the measurement model is mis-specified. In order to support groups such as those described in the above-mentioned studies these uncertainties must first be ruled out. The reader should be aware that papers often neglect to mention that the classes might reflect non-normality or other model violations. This is a problem especially in psychiatric research where measurement tools were often designed as screening tools and the items used in the measurement tools were not selected for psychometric properties. Given the complexity of the mixture models a significant effort is needed for researchers to develop a real understanding of the outcomes. Indeed, FMM might not be the magic bullet to identify true latent structure of data and alternative explanation must always be explored.

Nevertheless, it is surprising not to find published FMM studies as one of the methods used to explore the validity of the DSM-IV or DSM-V criteria for schizophrenia. There has been concern about the way the DSM-IV deals with schizophrenia and related disorders (Tandon & Maj 2008) and the need for the incorporation of new pathophysiological and treatment data has necessitated the introduction of DSM-V. Some of the concerns included the imprecise boundary between schizophrenia and schizoaffective disorder, distinctions between delusional disorder and the

psychotic subtypes or variants of obsessive-compulsive disorder and body dysmorphic disorder needed clarification. More specific to schizophrenia was the poor performance (in terms of longitudinal stability, heterogeneity of subtypes and the limited use subtypes other than paranoid and undifferentiated) of the classic subtypes of schizophrenia (Carpenter et al. 1976, McGlashan & Fenton 1994, Peralta & Cuesta 2003, Tandon et al. 2009). It is in light of this that the DSM-V eliminated the classic subtypes of schizophrenia and the special significance placed on Schneiderian 'first-rank symptoms. In addition, the delineation between schizoaffective disorder, schizophrenia and psychotic mood disorders was redefined. A new addition to the DSM classification system of psychotic disorders is ''attenuated psychosis syndrome'' as a condition for further study and addition of unique psychopathological dimensions.

Looking in more at the changes to the schizophrenia section reveals only modest changes from DSM-IV. The most obvious changes are to criterion A where the special focus on bizarre delusions and other Schneiderian first-rank symptoms were eliminated based on poor reliability (Carpenter et al. 1973, Carpenter & Strauss 1974, Cermolacce et al. 2010, Mojtabai & Nicholson 1995, Peralta & Cuesta 1999, Mullen 2003, Bell et al. 2006, Nordgaard et al. 2008, Tandon 2012). These symptoms are now treated as any other positive symptom. The core positive symptoms (one of these required for criteria A) are delusions, hallucinations, and disorganized thinking. The DSM-IV subtypes of schizophrenia were eliminated. Schizoaffective disorder also underwent significant changes to improve the nosological problems reflected its poor reliability and low diagnostic stability (Cheniaux et al. 2008, Malhi et al. 2008, Jager et al. 2011). The new criterion C specifies that the mood episode must be present for a majority of the total duration of the illness (Tandon & Carpenter, 2013).

In terms of the dimensionality of schizophrenia, the DSM-V encourage physician to measure schizophrenia in each of the following dimensions: reality distortion (delusions, hallucinations),

negative symptoms, disorganization, cognitive impairment, motor symptoms (e.g. catatonia), and mood symptoms (depression, mania). In addition to this , the inclusion of a, to be investigated, condition called "attenuated psychosis syndrome" is of interest as this condition aims to encapsulate the group of individuals that present with attenuated psychotic symptoms prior to onset of frank psychosis. It is estimated that this group is more than 100 times more likely than the general population to develop schizophrenia in the next 12 months. However, the majority do not go on to schizophrenia. It is thus necessary to conduct further field trials prior to inclusion of this diagnostic category in future versions of the DSM (Carpenter & van Os 2011, Fusar-Poli et al. 2012, Tandon et al. 2012, Regier et al. 2013). The results from the Xhosa population may shed more light on the appropriateness of the new DSM-V diagnosis; similar to the information gleamed from the field trials in major depression.

Looking at the proposed new ICD (ICD-11) criteria for the diagnosis of schizophrenia (Gaebel et al. 2013), it requires at least two symptoms from the following groups: "(a) persistent delusions of any kind; (b) persistent hallucinations in any modality; (c) thought disorder; and (d) distortions of self-experience. Additional symptoms include negative symptoms; disorganized behaviour and psychomotor disorders. As with the DSM-V, the importance of first-rank symptoms were neutralized. However, a difference from the DSM-V is that the ICD 11 retain the one-month duration criterion. This duration criteria has shown good stability and evidence for a change in duration is lacking (Salvatore et al. 2011, Keith & Matthews 1991).

As with the DSM-V, the ICD-11 has dropped the subtypes and replaced it with six symptom specifiers (positive, negative, depressive, manic, and psychomotor symptoms, and cognitive impairment). In ICD-11, course specifiers include first episode, multiple episode cases, full remission, partial remission and continuous symptoms. It seem from the revisions that DSM and ICD has tried to harmonize the criteria of interest for this study i.e. positive and negative

symptoms measured by the SAPS and SANS. It will be of interest to see if the results from our analysis in the Xhosa schizophrenia population can inform us on the applicability of the new diagnostic criteria in this population.

1.5 Aims and objectives of this study

In this study we look at LCA as a method of finding an organizing principle behind the complex set of variables measured in a Xhosa schizophrenia sample in order to potentially organise the participants of the study into meaningful and homogeneous classes. Class membership would be based on the probability of an individual possessing certain observed patterns of responses obtained using diagnostic instruments and identified by the latent class model.

Patients with psychosis can be hypothesized to be in a space with different dimensions including positive symptoms, negative symptoms etc. The problem is to find where in this "space" a specific patient can be found and to group this patient with similar patients in order to develop a cohort that can be used for genetics studies. The importance of the current study in the Xhosa population is that in characterizing the symptoms of schizophrenia with respect to the underlying latent structure, the validity and reliability of further studies designed to shed light on possible illness modifying or causative factors might be enhanced. These factors could include clinical or more complex endophenotypes or intermediate endophenotypes. In addition, this approach can assist in developing or validating cut-off points for symptom measurement tools.

Since subtypes may represent individuals with similar genetic makeup, subtyping schizophrenic subjects by means of latent class analysis may later prove useful in studying the genetic underpinnings of diseases such as schizophrenia. For example, if membership in a particular class

can be shown to be associated with genetic vulnerability linked to a specific genetic variant or group of variants, by virtue of being present in both diseased and non-diseased family members at a higher rate than that found in the general population, such latent class can be considered a possible endophenotype for the disorder.

In this study we performed LCA on selected items of the SANS and SAPS scales in order to subtype Xhosa schizophrenic patients in terms of the latent structure of their symptoms and to compare the findings with those reported in Caucasian samples (Sham et al. 1996, Blanchard et al. 2005). We were specifically interested in the usefulness of the four items found to be concordant in our previous Xhosa sib pair study, namely eye contact, auditory hallucinations, global hallucinations, and delusions of control. The identification of patients with a high probability of belonging to certain latent classes on the basis of their response profiles to the SANS/SAPS can pave the way for further investigations of schizophrenia endophenotypes and intermediate endophenotypes and the elucidation of underlying pathways that could influence the phenomenology of this disease. Because of the limitations of a purely categorical approach, the latent class analysis was further expanded to include FMM. The ability to include information regarding symptom severity will add greatly to the wealth of information that can be gleaned from a study such as this, especially since different genetic influences may result not only in different symptom patterns, but also in different levels of symptom severity.

In summary we aimed to reduce the clinical heterogeneity of this Xhosa schizophrenia sample by generating latent classes using FA, LCA and FMM, and exploring degrees of symptom severity within each class. Covariates such as gender and comorbid cannabis abuse were included, where applicable, in the model in order to further characterise the sample.

1.6 References

Allen AJ, Griss ME, Folley BS, et al. Endophenotypes in schizophrenia: A selective review. Schizophrenia Research 2009 109:24-37.

Amador XF, Kirkpatrick B, Buchanan RW, et al. Stability of the diagnosis of deficit syndrome in schizophrenia. American Journal of Psychiatry 1999 156:637-639.

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Third Edition. Washington, DC: American Psychiatric Association, 1980.

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Third Edition, Revised. Washington, DC: American Psychiatric Association, 1987.

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition. Washington, DC: American Psychiatric Association, 1994.

Andreasen NC, Flaum M, Arndt S. The Comprehensive Assessment of Symptoms and History (CASH). An instrument for assessing diagnosis and psychopathology. Archives of General Psychiatry 1992 49:615–623.

Andreasen NC, Arndt S, Miller D, et al. Correlational studies of the Scale for the Assessment of Negative Symptoms and the Scale for the Assessment of Positive Symptoms: An overview and update. Psychopathology 1995 28(1):7-17.

Andreasen NC, Olsen SA. Negative v. positive schizophrenia: Definition and validation. Archives of General Psychiatry 1982 39:789-794.

Andreasen NC. The Scale for the Assessment of Negative Symptoms (SANS) Iowa City, IA, University of Iowa 1983.

Andreasen NC. The Scale for the Assessment of Positive Symptoms (SAPS) Iowa City, IA, University of Iowa 1984.

Andreasen NC, Grove W. Evaluation of positive and negative symptoms in schizophrenia. Psychiatric Biology 1986 1:108-121.

Andreasen NC. Negative symptoms in schizophrenia. Definition and reliability. Archives of General Psychiatry 1982 39:784−788.

Andreasson S, Allebeck P, Engstrom A, et al. Cannabis and schizophrenia.A longitudinal study of Swedish conscripts. Lancet 1987 2:1483–1486.

Arajarvi R, Haukka J, Varilo T, et al. Clinical phenotype of schizophrenia in a Finnish isolate. Schizophrenia Research 2004 67:195-205.

Arndt S, Alliger RT, Andreasen NC. The distinction of positive and negative symptoms: the failure of a two-dimensional model. British Journal of Psychiatry 1991 158:317-322.

Arseneault L, Cannon M, Poulton R, et al. Cannabis use in adolescence and risk for adult psychosis: longitudinal prospective study. British Medical Journal 2002 325:1212–1213.

Arseneault L, Cannon M, Witton J, et al. Causal association between cannabis and psychosis: examination of the evidence. British Journal of Psychiatry 2004 184:110–17.

Arsenault L, Cannon M, Fisher HL, et al. Childhood trauma and children's emerging psychotic symptoms: a genetically sensitive longitudinal cohort study. American Journal of Psychiatry 2011 168:65-68.

Ballon N, Leroy S, Roy C, et al. (AAT)n repeat in the cannabinoid receptor gene (CNR1): association with cocaine addiction in an African-Caribbean population. Pharmacogenomics Journal 2005 6:126.

Barrett JC, Fry B, Maller J, et al. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005 21(2):263-265.

Bell MD, Lysaker PH, Beam-Goulet JL, et al. Five-component model of schizophrenia: Assessing the factorial invariance of the Positive and Negative Syndrome Scale. Psychiatry Research 1994 52(3):295-303.

Bijl RV, Ravelli A, van Zessen G. Prevalence of psychiatric disorder in the general population: results of The Netherlands Mental Health Survey and Incidence Study (NEMESIS). Soc Psychiatry Psychiatric Epidemiology 1998 33(12):587-95.

Bilder RM, Mukherjee S, Rieder RO, et al. Symptomatic and neuropsychological components of defect states. Schizophrenia Bulletin 1985 11:409-419.

Blanchard JJ, Horan WP, Collins LM. Examining the latent structure of negative symptoms: Is there a distinct subtype of negative symptom schizophrenia? Schizophrenia Research 2005 77:151– 165.

Boks PM, Leask S, Vermunt JK, et al. The structure of psychosis revisited: The role of mood symptoms. Schizophrenia Research 2007 93:178-185.

Brekke JS, DeBonis JA, Graham JW. A latent structure analysis of the positive and negative symptoms in schizophrenia. Comprehensive Psychiatry 1994 35(4):252-259.

Bresnahan M, Begg MD, Brown A, et al. Race and risk of schizophrenia in a US birth cohort: another example of health disparity? International Journal of Epidemiology 2007 36(4):751-758.

Brown AS. The environment and susceptibility to schizophrenia. Progress in Neurobiology 2011 93:23–58.

Buchanan RW, Carpenter WT. Domains of psychopathology: An approach to the reduction of heterogeneity in schizophrenia. Journal of Nervous and Mental Disease 1994 182(4):193-204.

Burke JG, Murphy BM, Bray JC, et al. Clinical similarities in siblings with schizophrenia. American Journal of Medical Genetics 1996 67(3):239-243.

Burks BS. The relative influence of nature and nurture upon mental development: A comparative study of foster parent-foster child resemblance and true parent-true child resemblance. The Twenty-Seventh Yearbook of the National Society for the Study of Education 1928 27:219-316.

Calkins ME, Iacono WG, Ones DS. Eye movement dysfunction in first-degree relatives of patients with schizophrenia: a meta-analytic evaluation of candidate endophenotypes. Brain Cognition 2008 68(3):436-61.

Cardno AG, Jones LA, Murphy KC, et al. Factor analysis of schizophrenic symptoms using the OPCRIT checklist. Schizophrenia Research 1996 22(3):233-239.

Cardno AG, Jones LA, Murphy KC, et al. Dimensions of psychosis in affected sibling pairs. Schizophrenia Bulletin 1999 25(4):841-850.

Caroff SN, Ungvari GS. Guest editorial. Psychiatric Annals 2007 37:7-9.

Carpenter WT, Heinrichs DW, Wagman AM. Deficit and nondeficit forms of schizophrenia: The concept. American Journal of Psychiatry 1988 145:578–583.

Caspi A, Moffitt TE, Cannon M, et al. Moderation of the effect of adolescent-onset cannabis use on adult psychosis by a functional polymorphism in the catechol-O-methyltransferase gene: longitudinal evidence of a gene environment interaction. Biological Psychiatry 2005 57:1117–1127.

Castle DJ, Sham PC, Wessely S, et al. The subtyping of schizophrenia in men and women: a latent class analysis. Psychological Medicine 1994 24:41-51.

Chiu YF, McGrath JA, Thornquist MH, et al. Genetic heterogeneity in schizophrenia II: conditional analysis of affected schizophrenia sibling pairs provide evidence for an interaction between markers on chromosome 8p and 14q. Molecular Psychiatry 2002 7:658-664.

Comings DE, Muhleman D, Gade R, et al. Cannabinoid receptor gene (CNR1): association with IV drug use. Molecular Psychiatry 1997 2:161-168.

Compton MT, Furman AC, Kaslow NJ. Lower negative symptom scores among cannabis-dependent patients with schizophrenia-spectrum disorders: preliminary evidence from an African American first-episode sample. Schizophrenia Research 2004 71:61-64.

Corvin A, Craddock N, Sullivan PF. Genome-wide association studies: a primer. Psychological Medicine 2010 40:1063–1077.

Crow TJ. Molecular pathology of schizophrenia: More than one disease process? British Medical Journal 1980 280:66-68.

Crow TJ. Positive and negative schizophrenia symptoms and the role of dopamine. British Journal of Psychiatry 1981 139:251-254.

Crow TJ. Con: the demise of the Kraepelinian binary system as a prelude to genetic advance. In: Gershon ES, Cloninger CR, eds. Genetic Approaches to Mental Disorders. Washington, DC: American Psychiatric Press 1994:163-192.

Dawson E. Identification of a polymorphic triplet marker for the brain cannabinoid receptor gene: use in linkage and association studies of schizophrenia. Psychiatric Genetics 1995 5:50-51.

Derks EM, Allardyce J, Boks MP, et al. Improvement of phenotyping in genome wide association studies on schizophrenia: an application of latent class factor analysis. Schizophrenia Research 2010:184-185.

DeRosse P, Funke B, Burdick KE, et al. Dysbindin genotype and negative symptoms in schizophrenia. American Journal of Psychiatry 2006 163:532-534.

Detera-Wadleigh SD, Badner JA, Yoshikawa T, et al. Initial genome scan of the NIMH genetics initiative bipolar pedigrees: chromosome 4, 7, 9, 18, 19, 20 and 21q. American Journal of Medical Genetics 1997 74:254-262.

Detera-Wadleigh SD, Badner JA, Berrettini WH, et al. A high-density genome scan detects evidence for a bipolar-disorder susceptibility locus on 13q32 and other potential loci on 1q32 and 18p11.2. Proceedings of the National Academy of Science 1999 96:5604-5609.

Diagnostic and Statistical Manual of Mental Disorders, 4th Edition (DSM-IV 4th Edition), Text Revision. Washington, DC, American Psychiatric Association, 2000.

Drögemöller BI, Wright GE, Niehaus DJ, et al. Whole-genome resequencing in pharmacogenomics: moving away from past disparities to globally representative applications. Pharmacogenomics 2011 12:1717-28.

Dudley KJ, Xiang L, Kobor MS, et al. Epigenetic mechanism mediating vulnerability and resilience to psychiatric disorders. Neuroscience and Biobehavioral Reviews 2011 (e-version).

Du Toit ED, MacGregor KJ, Taljaard DG, et al. HLA-A, B, C, DR and DQ polymorphisms in three South African population groups: South African Negroes, Cape Coloureds and South African Caucasoids. Tissue Antigens 1988 31(3):109-125.

Emrich HM, Leweke FM, Schneider U. Towards a cannabinoid hypothesis of schizophrenia: cognitive impairments due to dysregulation of the endogenous cannabinoid system. Pharmacology Biochemistry Behaviour 1996 56:803-807.

Emsley RA, Niehaus DJH, Mbanga NI, et al. The factor structure for positive and negative symptoms in South African Xhosa patients with schizophrenia. Schizophrenia Research 2001 47:149– 157.

Emsley RA, Roberts MC, Rataemane S, et al. Ethnicity and treatment response in schizophrenia: a comparison of three ethnic groups. Journal of Clinical Psychiatry 2002 63:9 – 14.

Ensink K, Robertson BA, Ben Arie O, et al. Expression of schizophrenia in black Xhosa-speaking and white English-speaking South Africans. South African Medical Journal 1998 88:883–887.

Fanous AH, Neale MC, Webb BT et al. Novel linkage to chromosome 20p using latent classes of psychotic illness in 270 Irish High-Density Families. Biological Psychiatry 2008 64:121-127.

Fanous AH, Van Den Oord EJ, Riley BP, et al. Relationship between a high-risk haplotype in the DTNBP1 (dysbindin) gene and clinical features of schizophrenia. American Journal of Psychiatry 2005 162:1824-1832.

Farmer AE, McGuffin P, Spitznagel EL. Heterogeneity in schizophrenia: a cluster-analytic approach. Psychiatry Research 1983 8(1):1-12.

Fenton WS, McGlashan TH. Natural history of schizophrenia subtypes II. Positive and negative symptoms and long-term course. Archives of General Psychiatry 1991 48:969-977.

Fergusson DM, Horwood LJ, Swain-Campbell NR. Cannabis dependence and psychotic symptoms in young people. Psychological Medicine 2003 33:15–21.

Fisher HL, Jones PB, Fearon P. The varying impact of type, timing and frequency of exposure to childhood adversity on its association with adult psychotic disorder. Psychological Medicine 2010 40:1967-1978.

Foussias G, Remington G. Negative symptoms in schizophrenia: Avolition and Occam's razor. Schizophrenia Bulletin 2008 36:359−369.

Galderisi S, Maj M. Deficit schizophrenia: an overview of clinical, biological and treatment aspects. European Psychiatry 2009 24:493-500.

Girard SL, Xiong L, Dion PA, et al. Where are the missing pieces of the schizophrenia genetics puzzle? Current Opinion in Genetics and Development 2011 21:310-316.

Gottesman II, Gould TD. The endophenotype concept in psychiatry: etymology and strategic intentions. American Journal of Psychiatry 2003 160:636-645.

Gur RE, Calkins ME, Gur RC, et al. The consortium on the genetics of schizophrenia: neurocognitive endophenotypes. Schizophrenia Bulletin 2007 33:49–68.

Gurling HMD, Kalsi G, Brynjolfson J, et al. Genome wide genetic linkage analysis confirms the presence of susceptibility loci for schizophrenia , on chromosome 1q32.2, 5q33.2 and 8p21-22 and provides support for linkage to schizophrenia, on chromosomes 11q23.3-24 and 20q12.1-11.23. American Journal of Human Genetics 2001 68:661-673.

Häfner H, Riecher A, Maurer K, et al. How does gender influence age at first hospitalization for schizophrenia? Psychological Medicine 1989 19(4):903–918.

Hagenaars JA, McCutcheon AL. Applied Latent Class Analysis. Cambridge University Press 2002.

Henquet C, Murray R, Linszen D, et al. The environment and schizophrenia:the role of cannabis use. Schizophrenia Bulletin 2005 31:608–612.

Henquet C, Rosa A, Krabbendam L, et al. An experimental study of catechol-o-methyltransferase Val158Met moderation of delta-9-tetrahydrocannabinol-induced effects on psychosis and cognition. Neuropsychopharmacology 2006 31:2748–2757.

Howes OD, McDonald C, Cannon M, et al. Pathways to schizophrenia: the impact of environmental factors. Int J Neuropsychopharmacology 2004 7:S7-S13.

Ingason A, Rujescu D, Cichon S, et al. Copy number variations of chromosome 16p12.1 region associated with schizophrenia. Molecular Psychiatry 2011 16:17-25.

Jablensky A. The diagnostic concept of schizophrenia: its history, evolution, and future prospects. Dialogues in Clinical Neuroscience 2010 12:271-287.

Kaufmann CA, Suarez B, Malaspina D, et al. NIMH Genetics Initiative millenium schizophrenia consortium: linkage analysis of African-American pedigrees. American Journal of Medical Genetics 1998 81:282–289.

Kay SR, Fiszbein A, Opler LA. The Positive and Negative Syndrome Scale (PANSS) for schizophrenia. Schizophrenia Bulletin 1987 13(2):261-276.

Keefe RS, Harvey PD, Lenzenweger MF, et al. Empirical assessment of the factorial structure of clinical symptoms in schizophrenia: Negative symptoms. Psychiatry Research 1992 44:153−165.

Kelly BD, O'Callaghan E, Waddington JL. Schizophrenia and the city: a review of literature and prospective study of psychosis and urbanicity in Ireland. Schizophrenia Research 2010 116(1):75-89.

Kendler KS, Lieberman JA, Walsh D. The structured interview for schizotypy (SIS): a preliminary report. Schizophrenia Bulletin 1989 15:559-571.

Kendler KS, Myers JM, O'Neill FA, et al. Clinical features of schizophrenia and linkage to chromosomes 5q, 6p, 8p, and 10p in the Irish Study of High-Density Schizophrenia Families. American Journal of Psychiatry 2000 157:402-408.

Kendler KS. The impact of diagnostic misclassification on the pattern of familial aggression and coaggregation of psychiatric illness. Journal of Psychiatric Research 1987 21:55-91.

Kendler KS, Karkowski LM, Walsh D. The structure of psychosis: latent class analysis of probands from the Roscommon Family Study. Archives of General Psychiatry 1998 55:492–499.

Kendler KS. Kraepelin and the diagnostic concept of paranoia. Comprehensive Psychiatry 1988 29:4-11.

Kendler KS, Karkowski-Shuman L, O'Neill FA, et al. Resemblance of psychotic symptoms and syndromes in affected sibling pairs from the Irish Study of High-Density Schizophrenia Families: evidence for possible etiologic heterogeneity. American Journal of Psychiatry 1997 154(2):191-8.

Khashan AS, Abel KM, McNamee R, et al. Higher risk of offspring schizophrenia following antenatal maternal exposure to severe adverse life events. Archives of General Psychiatry 2008 65:146-152.

Kirkpatrick B, Buchanan RW, McKenney PD, et al. The Schedule for the Deficit Syndrome: An instrument for research in schizophrenia. Psychiatry Research 1989 30(2):119-123.

Kirkpatrick B, Buchanan RW, Ross DE, et al. A separate disease within the syndrome of schizophrenia. Archives of General Psychiatry 2011 58(2):165-171.

Kirkpatrick B, Ram R, Amador XF, et al. Summer birth and the deficit syndrome of schizophrenia. American Journal of Psychiatry 1998 155:1221–1226.

Kirkpatrick B, Buchanan RW, Breier A, et al. Case identification and stability of the deficit syndrome of schizophrenia. Psychiatry Research 1993 47:47−56.

Kirkpatrick B. Ross DE, Walsh D, et al. Family characteristics of deficit and non-deficit schizophrenia in the Roscommon family study. Schizophrenia Research 2000 45:57-64.

Kirkpatrick B, Amador XF, Flaum M, et al. The deficit syndrome in the DSM-IV Field Trial: I. Alcohol and other drug use. Schizophrenia Research 1996 20:69-77.

Kirkpatrick B, Fenton WS, Carpenter WT, et al. The NIMHMATRICS consensus statement on negative symptoms. Schizophrenia Bulletin 2006 32:214−219.

Klimidis S, Stuart GW, Minas IH, et al. Positive and negative symptoms in the psychoses. Re-analysis of published SAPS and SANS global ratings. Schizophrenia Research 1993 9(1):11-18.

Knapp M. Costs of schizophrenia. British Journal of Psychiatry 1997 171:509–518.

Knight HM, Pickard BS, Maclean A, et al. A cytogenetic abnormality and rare coding variants identify ABCA13 as a candidate gene in schizophrenia, bipolar disorder, and depression. American Journal of Human Genetics 2009 85:833-846.

Krabbendam L, Myin-Germeys I, De Graaf R, et al. Dimensions of depression, mania and psychosis in the general population. Psychological Medicine 2004 34(7):1177-86.

Kraepeline J. Manic-Depressive Insanity and Paranoia, (Translated by Barclay M.). Livingstone, Edinburgh 1921.

Kraepelin E. Psychiatrie (1896): Reprinted (1971) in part as Dementia-Praecox and Paraphrenia Robert E.Krieger, Huntington, NY.

Kraepelin E. Clinical Psychiatry: A Text-Book for Students and Physicians. [abstracted and adapted from the sixth German edition of Kraepelin's Lehrbuch der Psychiatrie by A. Ross Diefendorf, MD]. New York, NY: The Macmillan Co, 1904.

Kraepelin E. Kraepelin on "paranoid conditions." Gosline HI, trans. Alienist Neurology 1916 37:184-210.

Kraepelin E. Dementia Praecox and Paraphrenia. Huntington, NY: Krieger Publishing, 1971.

Kulhara P, Kota SK, Joseph S. Positive and negative subtypes of schizophrenia: a study from India. Acta Psychiatrica Scandinavica 1986 74:353-379.

Lancon C, Aghababian V, Llorca PM, et al. Factorial structure of the Positive and Negative Syndrome Scale (PANSS): A forced five-dimensional factor analysis. Acta Psychiatrica Scandinavica 1998 98(5):369-376.

Leask SJ, Vermunt JK, Done DJ, et al. Beyond symptom dimensions: schizophrenia risk factors for patients groups derived by latent class analysis. Schizophrenia Research 2009 115:346-350.

Lee KW, Woon PS, Teo YY, et al. Genome-wide association studies (GWAS) and copy number variation (CNV) studies of the major psychosis: what have we learnt? Neuroscience and Biobehavioral Reviews 2012 36:556-71.

Lencz T, Szeszko PR, DeRosse P, et al. A schizophrenia risk gene, ZNF804A, influences neuroanatomical and neurocognitive phenotypes. Neuropsychopharmacology 2010 35(11):2284-91.

Lenzenweger MF, Dworkin RH. The dimensions of schizophrenia phenomenology. Not one or two, at least three, perhaps four. British Journal of Psychiatry 1996 168(4):432-40.

Lenzenweger MF, Dworkin RH, Wethington E. Models of positive and negative symptoms in schizophrenia: an empirical evaluation of latent structures. Journal of Abnormal Psychology 1989 98:62-70.

Leroy S, Griffon N, Bourdel MC, et al. Schizophrenia and the cannabinoid receptor type 1 (CB1): association study using a single-base polymorphism in coding exon 1. American Journal of Medical Genetics 2001 105:749-752.

Liddle PF. Commentary on the Modified Rogers Scale and the "conflict of paradigms" hypothesis. British Journal of Psychiatry 1991 158:337–339.

Lindenmayer JP, Bernstein-Hyman R, Grochowski S. Five-factor model of schizophrenia: Initial validation. Journal of Nervous and Mental Disease 1994 182(11):631-638.

Lindenmayer JP, Grochowski S, Hyman RB. Five factor model of schizophrenia: Replication across samples. Schizophrenia Research 1995 14(3):229-234.

Lysaker P, Bell M, Beam-Goulet J, et al. Relationship of positive and negative symptoms to cocaine abuse in schizophrenia. Journal of Nervous and Mental Disorders 1994 182:109-112.

Macleod J, Oakes R, Copello A, et al. Psychological and social sequelae of cannabis and other illicit drug use by young people: a systematic review of longitudinal, general population studies. Lancet 2004 363:1579–1588.

Mah S, Nelson MR, DeLisi LE, et al. Identification of the semaphoring receptor PLXNA2 as a candidate for susceptibility to schizophrenia. Molecular Psychiatry 2006 11:471-478.

March D, et al. Psychosis and place. Epidemiology Review 2008 30:84-100.

McGrath JA, Nestadt G, Liang KY, et al. Five latent factors underlying schizophrenia: analysis and relationship to illnesses in relatives. Schizophrenia Bulletin 2004 30(4):855-73.

McGuffin P, Farmer A, Harvey I. A polydiagnostic application of operational criteria in studies of psychotic illness. Development and reliability of the OPCRIT system. Archives of General Psychiatry 1991 48:764-770.

McGuffin P, Farmer AE, Gottesman II, et al. Twin concordance for operationally defined schizophrenia. Archives of General Psychiatry 1984 41:541-545.

Menninger K, Ellenberger H, Pruyser P, et al. The unitary concept of mental illness. Bulletin Menninger Clin 1958 22:4-12.

Messinger JW, Tremeau F, Antonius D, et al. Avolition and expressive deficits capture negative symptom phenomenology: Implications for DSM-5 and schizophrenia research. Clinical Psychology Review 2011 31:161−168.

Minas IH, Stuart GW, Klimidis S, et al. Positive and negative symptoms in the psychosis: multidimensional scaling of SAPS and SANS items. Schizophrenia Research 1992 8:143-156.

Minas IH, Klimidis S, Stuart GW, et al. Positive and negative symptoms in the psychoses: principal components analysis of items from the Scale for the Assessment of Positive Symptoms and the Scale for the Assessment of Negative Symptoms. Comprehensive Psychiatry 1994 35:135−144.

Minozzi S, Davoli M, Bargagli AM, et al. An overview of systematic reviews on cannabis and psychosis: discussing apparently conflicting results. Drug and Alcohol Review 2010 29(3):304-317.

Mojtabai R. Duration of illness and structure of symptoms in schizophrenia. Psychological Medicine 1999 29(4):915-924.

Moore TH, Zammit S, Lingford-Hughes A, et al. Cannabis use and risk of psychotic or affective mental health outcomes: a systematic review. Lancet 2007 370:319−328.

Murphy BM, Burke JG, Bray JC, et al. An analysis of the clinical features of familial schizophrenia. Acta Psychiatrica Scandinavica 1994 89(6):421-427.

Nakaya M, Ohmori K. A two-factor structure for the Schedule for the Deficit Syndrome in schizophrenia. Psychiatry Research 2008 158:256−259.

Ndetei DM, Singh A. Schneider's first rank symptoms of schizophrenia in Kenyan patients. Acta Psychiatrica Scandinavia 1983 67:148– 153.

Ng MYM, Levinson DF, Faraone SV, et al. Meta-analysis of 32 genome-wide linkage studies of schizophrenia. Molecular Psychiatry 2009 14:774–785.

Niehaus DJH, Koen L, Laurent C, et al. Positive and negative symptoms in affected sib pairs with schizophrenia: implications for genetic studies in an African Xhosa sample. Schizophrenia Research 2005 79:239.

Northoff G. What catatonia can tell about "top-down modulation": a neuropsychiatric hypothesis. Brain Behavioural Science 2002 25:555-604.

Nurnberger JI Jr, Blehar MC, Kaufmann CA. Diagnostic interview for genetic studies. Rationale, unique features, and training. NIMH Genetics Initiative. Archives of General Psychiatry. 1994 51(11):849-59.

Onaivi ES, Leonard CM, Ishiguro H, et al. Endocannabinoids and cannabinoid receptor genetics. Progress Neurobiology 2002 66:307-344.

Pedersen CB, Mortensen PB. Evidence for a dose-response relationship between urbanicity during upbringing and schizophrenia risk. Archives of General Psychiatry 2001 58:1039-1046.

Peralta V, Cuesta MJ, de Leon J. An empirical analysis of latent structures underlying schizophrenic symptoms: A four-syndrome model. Biological Psychiatry 1994 36(ll):726-736.

Peralta V, Cuesta MJ. Familial liability and schizophrenia phenotypes: a polydiagnostic approach. Schizophrenia Research 2007 96:125-134.

Peralta V, Cuesta MJ. Dimensional structure of psychotic symptoms: an item-level analysis of SAPS and SANS symptoms in psychotic disorders. Schizophrenia Research 1999 38:13−26.

Peralta V, Cuesta MJ. The deficit syndrome of the psychotic illness. A clinical and nosological study. European Archives of Psychiatry and Clinical Neuroscience 2004 254:165−171.

Peralta V, Cuesta MJ. The nosology of psychotic disorders: a comparison among competing classification systems. Schizophrenia Bulletin 2003 29(3):413-425.

Piton A, Gauthier J, Hamdan FF, et al. Systematic resequencing of X-chromosome synaptic genes in autism spectrum disorder and schizophrenia. Molecular Psychiatry 2010 16(8):867-880.

Ratakonda S, Gorman JM, Yale SA, et al. Characterization of psychotic conditions: Use of the domains of psychopathology model. Archives of General Psychiatry 1998 55(1):75-81.

Regier DA, Farmer ME, Rae DS. Comorbidity of mental disorders with alcohol and other drug abuse. Results from the Epidemiologic Catchment Area (ECA) study. JAMA 1990 264:2511 – 2518.

Robins LN, Helzer JE, Croughan J, et al. National Institute of Mental Health Diagnostic Interview Schedule. Archives of General Psychiatry 1981 38:38I.

Rösslera W, Salizeb HJ, van Os J, et al. Size of burden of schizophrenia and psychotic disorders. European Neuropsychopharmacology 2005 15(4):399-409.

Rouillon F, Dansette GY, Le Floch C. Therapeutic management of schizophrenic patients and its cost. Encephale 1994 20:303–309.

Ruscio J, Ruscio AM. Categories and Dimensions Advancing Psychological Science Through the Study of Latent Structure. Current Directions of Psychological Sciences 2008 17(3):203-207.

Semple DM, McIntosh AM, Lawrie SM. Cannabis as a risk factor for psychosis: systematic review. Journal of Psychopharmacology 2005 19 (2):187–194.

Schofield P, Ashworth M, Jones R. Ethnic isolation and psychosis: re-examining the ethnic density effect. Psychological Medicine 2011 41(6):1263-1269.

Schreier A, Wolke D, Thomas K, et al. Prospective study of peer victimization in childhood and psychotic symptoms in a nonclinical population at age 12 years. Archives of General Psychiatry 2009 66:527-536.

Schumacker RE, Lomax RG. A beginner's guide to structural equation modeling. New Jersey: Lawrence Erlbaum Associates, 2004.

Sham PC, Castle DJ, Wessely S, et al. Further exploration of a latent class typology of schizophrenia. Schizophrenia Research 1996 20(1-2):105-115.

Shevlin M, Murphy J, Dorahy MJ, et al. The distribution of positive psychosis-like symptoms in the population: a latent class analysis of the National Comorbidity Survey. Schizophrenia Research 2007 89(1-3):101-109.

Shifman S, Johannessn M, Bronstein M, et al. Genome-wide association identifies a common variant in the reelin gene that increases the risk of schizophrenia only in women. PLoS Genetics 2008 4:e28.

Spitzer RL, Williams JBW, Gibbon M, et al. Structured clinical interview for the DSM-III-R (SCID). New York, New York State Psychiatric Institute, Biometric Research 1987.

Statistisches Bundesam. Krankheitskosten 2002. Wiesbaden 2004.

Stefansson H, Ophoff RA, Steinberg S, et al. Common variants conferring risk of schizophrenia. Nature 2009 460(7256):744-747.

Stilo SA, Murray RM. The epidemiology of schizophrenia: replacing Dogma with Knowledge. Dialogues Clinical Neuroscience 2010 12:305-315.

Sullivan PF, Kendler KS, Neale MC. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. Archives of General Psychiatry 2003 60:1187-1192.

The WHO World Health Report: new understanding, new hope. Geneva, 2001.

Tiwari AK, Zai CC, Müller DJ, Kennedy JL. Genetics in schizophrenia: where are we and what next? Dialogues Clinical Neuroscience 2010 12:289-303.

Toomey R, Kremen WS, Simpson JC, et al. Revisiting the factor structure for positive and negative symptoms: Evidence from a large heterogeneous group of psychiatric patients. American Journal of Psychiatry 1997 154(3):371-377.

Trameau F, Goggin M, Antonius D, et al. A new rating scale for negative symptoms: The Motor-Affective-Social Scale. Psychiatry Research 2008 160:346−355.

Ujike H, Nakata K, Tanaka Y, et al. CNR1, central cannabinoid receptor gene, associated with susceptibility to hebephrenic schizophrenia. Molecular Psychiatry 2002 7:515-518.

Ujike H, Morita A. New perspectives in the studies on endocannabinoid and cannabis: cannabinoid receptors and schizophrenia. Journal of Pharmacological Science 2004 96:376-381.

Ullmann R, Turner G, Kirchhoff M, et al. Array CGH identifies reciprocal 16p13.1 duplications and deletions that predispose to autism and/or mental retardation. Human Mutation 2007 28:674-682.

Van Os J. From schizophrenia metafacts to non-schizophrenia facts. Schizophrenia Research 2011 127 (1-3):16-17.

Van Os J, Gilvarry C, Bale R, et al. A comparison of the utility of dimensional and categorical representations of psychosis. UK700 Group. Psychological Medicine 1999 29(3):595-606.

Ungvari GS, Goggins W, Leung S-K, et al. Schizophrenia with catatonic features ("catatonic features") III. Latent class analysis of the catatonic syndrome. Progress in Neuro-Psychopharmacology and Biological Psychiatry 2009 33:81-85.

van Os J, Bak M, Hanssen M, Bijl RV, et al. Cannabis use and psychosis: a longitudinal population-based study. American Journal of Epidemiology 2002 156:319–327.

Van Os J, Kenis G, Rutten BPF. The environment and schizophrenia. Nature 2010 468:203-212.

Walker EF, Harvey PD, Perlman D. The positive/negative symptom distinction in psychoses: a replication and extension of previous findings. Journal of Nervous and Mental Disorders 1988 176:359-363.

Waller NG, Meehl PE. Multivariate Taxometric Procedures: Distinguishing Types From Continua. Sage, Newbury Park, CA, 1998.

Walsh T, McClellan JM, McCarthy SE, et al. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. Science 2008 320:539-543.

Wang KS, Liu XF, Aragam N. A genome-wide meta-analysis identifies novel loci associated with schizophrenia and bipolar disorder. Schizophrenia Research 2010 124:192-199.

Wessman J, Paunio T, Tuulio-Henriksson A, et al. Mixture Model Clustering of Phenotype Features Reveals Evidence for Association of DTNBP1 to a Specific Subtype of Schizophrenia. Biological Psychiatry 2009 66:990–996.

White L, Harvey PD, Opler L, et al. Empirical assessment of the factorial structure of clinical symptoms in schizophrenia: A multisite, multimodal evaluation of the factorial structure of the Positive and Negative Syndrome Scale. The PANSS Study Group. Psychopathology 1997 30(5):263-274.

Wright GE, Niehaus DJ, Koen L, et al. Psychiatric genetics in South Africa: cutting a rough diamond. African Journal of Psychiatry 2011 14:355-66.

Xu B, Roos JL, Levy S, et al. Strong association of de novo copy number mutations with sporadic schizophrenia. Nature Genetics 2008 40:880-885.

Young MA. Evaluating diagnostic criteria: a latent class paradigm. J. Psychiof Rex 1983 17: 285-296.

Zammit S, Allebeck P, Andreasson S, et al. Self reported cannabis use as a risk factor for schizophrenia in Swedish conscripts of 1969: historical cohort study. British Medical Journal 2002 325:1199.

CHAPTER 2 METHODOLOGY

CONTENT

CHAPTER 2 METHODS

2.1 Methodology overview

This chapter describes: the recruitment and selection of the patient sample; their clinical and psychiatric assessment; background information and principles pertaining to latent class analysis; and the procedures followed in analysing the data.

2.2 Subjects

2.2.1 Recruitment and selection of subjects

Participants were recruited from in- and outpatient hospital services and community treatment centres throughout the Cape Town Metropole as part of a large multisite genetic study, which had been approved by the ethical committee of the University of Stellenbosch. Participants had to be of Xhosa ethnicity, defined as having four grandparents of Xhosa descent. Mental health workers were asked to identify possible participants with clinical features suggestive of schizophrenia, bipolar mood disorder or schizo-affective disorder, who were then screened for suitability using DSM-IV-TR criteria. Inclusion into this study was dependent on a diagnosis of schizophrenia or schizo-affective disorder. Patients were excluded if history and assessment left any doubt as to the diagnosis. The presence of comorbid conditions was only considered as exclusionary if the condition precluded the investigators from making the diagnosis of schizophrenia or schizo-affective disorder. Informed, written consent was obtained before including each subject in the study.

2.2.2 Demographic and clinical assessment of subjects

The primary diagnostic instrument used was the Diagnostic Interview for Genetic Studies (DIGS) version 2.0 (Nurnberger et al. 1994), a structured interview which was developed by the NIMH Diagnostic Centers for Psychiatric Linkage Studies in the United States for accurately recording the phenotypes of the major psychiatric disorders for the purposes of genetic research.

A trained psychiatrist (DJHN, CS, and JM) and a Xhosa speaking psychiatric nurse with extensive clinical experience (IM) assessed each participant using an English version of the DIGS, which was verbally translated into Xhosa. At regular calibration meetings a patient was assessed in team format.

Further assessments (see below for a more detailed description) included a urinary drug screen, the construction of a family tree, and the following standardized instruments: the AIMS (Abnormal Involuntary Movement Scale) assessment (Guy 1976), the Barnes Akathisia rating scale (Barnes 1989), the neurological evaluation scale (NES) (Buchanan et al. 1989), and the Modified Waldrop Scale (MWS) (Waldrop 1968) for minor physical anomalies.

The interviewers also used hospital and clinic charts, where available, to supplement the interview and collateral information (an unstructured family interview) was gathered from family members.

At the end of the assessment a lifetime dimensions of psychosis scale (unpublished data) and a narrative summary were completed for each patient.

Other data captured for analysis purposes included demographic characteristics (age, gender, and educational level), sib pair status, and age of onset, number of episodes, comorbid conditions, medication use history and duration of illness.

2.2.2.1 The Diagnostic Interview for Genetic Studies

This instrument was used in the current study because it provides a comprehensive profile of the phenotypic features displayed by each patient, and classifies subjects into discrete diagnostic categories according to the DSM-III (Diagnostic and Statistical Manual of Mental Disorders 1987) and the diagnosis (by investigators) based on DSM-IV-TR (Diagnostic and Statistical Manual of Mental Disorders 2000). The symptoms and signs of schizophrenia contained in the DIGS are assessed by means of the following sections exploring life-time and current symptom presence: medical history, the modified Mini-mental status examination (completed only if applicable), somatization symptoms, an overview of psychiatric illness, major depression, mania/hypomania, dysthymia, cyclothymia, alcohol abuse and dependence, tobacco use, drug abuse and dependence, marijuana use or abuse, psychosis (contains a section on establishing the presence of a current or past psychotic episode, delusions, hallucinations, disorganized behavior, positive formal thought disorder, catatonic motor behavior, avolition/apathy, alogia, affect, schizophrenia criterion A, onset of first symptoms episode, delineation of the current or most recent episode, prodromal and residual symptoms), schizoaffective disorder and illness patterns, schizotypal and antisocial personality features, comorbidity assessment, self-harm [including suicide], anxiety disorders, eating disorders, pathological gambling, global assessment scale and the interviewer' reliability assessment.

In addition, the SAPS (Schedule for the Assessment of Positive Symptoms) and the SANS (Schedule for the Assessment of Negative Symptoms) (Andreassen et al. 1983, Andreassen et al. 1984) as well as life-time symptomatology were completed.

The DIGS has been widely used in studies of the genetics of psychiatric disorders (Dubertret 2010, Hoon et al. 2004, Tang et al. 2007). Its test- retest reliability has been found to be good for most psychiatric conditions (0.73 to 0.95 except for schizoaffective disorder where the mood symptom duration in comparison to psychotic periods caused disagreement) (Nurnberger et al. 1994).

The SAPS, which consists of 34 items, comprises four symptom subscales (hallucinations, delusions, bizarre behaviour and positive formal thought disorder). A global rating is calculated for each of these four subscales of the SAPS. The SANS comprises 25 items representing 5 global subscale ratings measuring the domains affective flattening, alogia, avolition-apathy, anhedonia-asociality and attention. Each item relating to a particular symptom on the SANS and SAPS is assessed using a 6-point scale ranging from 0 to 5 as severity of the symptom increases. A rating of zero or one indicates the absence or possible presence of the symptom while a rating of 2 to 5 gives an indication of the severity of the symptom. Non-availability of information is rated as "unknown".

In addition to the SANS and SAPS, the DIGS gathered information on demographics, the personal and developmental history of the patient, medical history (including tobacco smoking) and full psychiatric history.

The following SANS / SAPS symptoms were evaluated and rated according to set criteria. The definition applied to each item is as follows and is based on Andreassen et al. (1983, 1984). See

Appendix 5.1 for the full description of criteria required for each symptom to be rated as being present.

A. Delusions

Delusions are firmly held, false beliefs out of keeping with a person's cultural and religious belief systems.

Persecutory Delusions

Persecutory delusions occur when the subject believes himself / herself to be the object of some kind of persecution or conspiracy. The severity is rated according to the duration and complexity of the delusions.

Delusions of Jealousy

The subject believes that his/her partner is involved in an illicit liaison with another person and acts on this belief.

Delusions of Sin or Guilt

The subject falsely believes himself / herself to have committed some unpardonable offense, of being the cause of some terrible disaster or of deserving punishment.

Grandiose Delusions

The subject entertains a false belief that he / she possess powers or abilities in excess of that experienced by others or is actually the incarnation of some famous person.

Religious Delusions

False, unshakable religious beliefs out of keeping with the subject's religious or socio-cultural background.

Somatic Delusions

False belief in abnormalities of the subject's body or bodily function. The presence of concomitant hallucinations should be rated separately.

Ideas and Delusions of Reference

The subject believes that references by others - whether delivered in person or by means of printed or electronic media - are directed at and have a special meaning for him / her.

Delusions of Being Controlled

The subject believes that he / she is being controlled by some force outside of him / herself.

Delusions of Mind Reading

The subject falsely believes that others are able to know what is going on in his / her mind.

Thought Broadcasting

The false belief that the subject's thoughts are audible to him/herself or others. If these thoughts are experienced as being audible outside his/her head, this is also an auditory hallucination.

Thought Insertion

The subject believes that thoughts are being planted in his mind by some external person or agency.

Thought Withdrawal

The subject believes that his / her thoughts are removed as soon as they start to take form in his / her mind.

B. Hallucinations

Hallucinations are abnormal sensory perceptions appearing without a stimulus.

Auditory Hallucinations

The subject hears sounds, most commonly voices, without a stimulus to these sounds.

Voices Commenting

Commenting voices, which are Schneiderian first rank symptoms, have been considered pathognomonic of schizophrenia, though this view is not universally accepted. The subject hears a voice constantly commenting on his / her actions or thoughts. This type of hallucination is rated as a separate type of hallucination distinct from other auditory hallucinations.

Voices Conversing

Another Schneiderian first-rank symptom, voices conversing entails the experience by the subject of hearing voices conversing with one another, usually about the subject. They are also rated as a distinct type of hallucination and not simply rated as an auditory hallucination.

Somatic or Tactile Hallucinations

The subject experiences false bodily sensations (e.g., burning, crawling, tingling) or the perception that body parts have somehow changed in structure.

Olfactory Hallucinations

The subject experiences odours that can be external or believed to be emanating from the subject himself /herself. If the subject cannot smell the odour, but believes that others can, it should be rated as a delusion rather than a hallucination.

Visual Hallucinations

The subject sees objects or people that are not present. Hypnogogic and hypnopompic hallucinations as well as hallucinations occurring in people who have taken hallucinogens should be excluded.

C. Behavioral features

Clothing and Appearance

The subject alters his appearance in some way or dresses in a bizarre or inappropriate way.

Social and Sexual Behavior

The subject engages in activities deemed inappropriate given the social situation.

Aggressive and Agitated Behavior

The subject may, often in an unpredictable manner, engage in aggressive, violent or agitated behaviour.

Repetitive or Stereotyped Behavior

The subject displays repetitive actions or rituals, often with symbolic meaning.

D. Thought disorder

Derailment (Loose Associations)

The subject's spontaneous speech is disjointed and characterised by ideas that are not clearly related to each other. There are unclear connections between sentences and clauses.

Tangentiality

The subject's answers to questions appear only distantly related to the questions asked, or may even be completely unrelated to them. Here derailment occurs in response to questions set to the subject.

Incoherence (Word Salad, Schizophasia)

Speech is sometimes incomprehensible due to unclear constructions within sentences or clauses. Where some lack of clarity of speech is a result of the subject's low educational level or intelligence or due to geographical or social variations in standard language use, this should not be rated as incoherence.

Illogicality

The subject makes unwarranted or illogical inferences which are not necessarily delusional. If the illogicality is related to a delusional system it should be rated as a delusion and not considered part of a thought disorder.

Circumstantiality

The subject takes a long time to get to the point of his or her explanations, often being side-tracked into providing additional information before reaching a conclusion.

Pressure of Speech

The subject's speech, which tends to be loud and forceful, is increased in rate and quantity and is hard to interrupt. The subject may talk without social stimulation. Certain medications may slow

speech, in which case pressure of speech should be rated on the basis of quantity, social appropriateness and loudness.

Distractible Speech

The subject's speech is interrupted by external stimuli.

Clanging

The subject's choice of language relies on the sounds of the words rather than their actual meaning. Intelligibility is hampered by the use of seemingly unnecessary words, rhymes, or a play on words.

E. Negative Symptoms

Unchanging Facial Expression

The subject does not show much emotion, his or her face appearing expressionless.   Although facial expression may be affected by phenothiazine use to an extent, this should not change the rating given to the subject at the time of interview. Medication use should be noted, however.

Decreased Spontaneous Movements

The subject shows very little or no spontaneous movement during the interview.

Paucity of Expressive Gestures

Very few or no gestures or body movements are used by the subject to help express him /herself during the interview.

Poor Eye Contact

The subject avoids eye contact or avoids using his / her eyes to convey meaning during the interview.

Affective Nonresponsivity

The subject does not smile or laugh in response to the interviewer's attempts to bring out such emotions.

Inappropriate Affect

This is not the same as affective flattening or blunting. Here the subject shows emotion or expressions that are not appropriate to the social situation, e.g., smiling while talking about a tragic event. If the subject's affect is socially inappropriate because of anxiety, this should not be rated as inappropriate affect.

Lack of Vocal Inflections

The subject's speech is monotonous, without the normal changes in volume or emphasis expected from changes in subject matter under discussion.

Alogia

Poverty of Speech

The subject gives very little spontaneous information, and even when allowed enough time to formulate or explain him- or herself, the subject's answers remain brief or even absent.

Poverty of Content of Speech

Unlike circumstantiality, in which the subject provides more than the requested amount of information, a subject with poverty of speech either communicates very little useful information,

even though the amount of speech may be normal, or may take a long time to provide any useful information.

Thought Blocking

The person's train of thought is interrupted so that he or she cannot complete the sentence. The subject either volunteers the fact that the thought he / she was about to express has been forgotten, or gives this as the reason when questioned by the interviewer.

Increased Latency of Response

Although the subject is aware of the interviewer's questions he / she takes longer than normal to reply.

Grooming and Hygiene

The subject's grooming and general hygiene are poor.

Impersistence at Work or School

The subject finds it difficult to maintain routine occupational, educational or other tasks.

Physical Anergia

The subject fails to participate in any physical activity.

Recreational Interests and Activities

The subject's ability to enjoy activities is restricted, or he /she may engage only in passive, undemanding pastimes.

Sexual Interest and Activity

The subject's shows a lack of interest in or desire for sexual activities.

Ability to Feel Intimacy and Closeness

The subject exhibits little need for or does not have the ability to engage in close relationships with family members."

Relationships with Friends and Peers

The subject tends to prefer spending time alone rather than socialising with friends.

Social Inattentiveness

The subject tends to be uninvolved and distant during social interactions and may not be able to engage in pastimes such as reading, playing games or watching television because of poor concentration.

2.2.2.2 Additional clinical assessments

Apart from the DIGS, each person underwent a comprehensive assessment which included the following:

A. the construction of a family tree with clinical notations regarding possible psychiatric or medical illnesses of significance,

B. the AIMS (Abnormal Involuntary Movement Scale) assessment (Guy 1976). The AIMS was developed by Guy in 1976 for the evaluation of abnormal involuntary movement and consists of seven items, namely: movements of the 1. muscles of facial expression, 2. lips

and perioral area, 3. jaw, 4. tongue, 5. upper extremities, 6. lower extremities, 7. neck, shoulder and hip areas, and three global ratings of the severity, incapacitation and patient awareness of the movements. Each item has four anchors (0-4) that rate the severity of the movement in the specific anatomical area.

C. The Barnes Akathisia rating scale (Barnes 1989). The BARS, one of the most widely used scales for assessing akathisia, was developed to provide a standardized means to assess akathisia.   The four-item Barnes Akathisia Rating Scale consists of an objective observation item (4 anchors), a subjective awareness item (4 anchors), a degree of distress item (4 anchors) and a global score (with 5 anchors) of the overall severity of the akathisia.

D. The neurological evaluation scale (NES) (Buchanan et al. 1989): This scale was developed for the assessment of different neurological variations seen in schizophrenia and consists of 26 items. The 26 items (14 of these tested on both sides of the body) cover sensory integration, motor coordination, and sequencing of complex motor acts. In addition, short-term memory, frontal release signs, eye movement abnormalities and cerebral dominance is tested. Each item is scored based on 3 anchors.

E. Assessment of minor physical anomalies and craniofacial measurements. The Modified Waldrop Scale (MWS) (Waldrop 1968) was used to assess minor physical anomalies in the eye region, ears, oral cavity and extremities. For the purpose of this assessment each individual had video and still images taken in a standardized setup (standard camera-patient distance). The images included a frontal and profile view of the head, palmar and dorsal views of the hands (fingers spread and unspread) and dorsal views of the feet with toes slightly spread (see Koen et al. 2006 for a detailed description). Each image was then examined by a clinical status-blinded clinical geneticist (GDJ) and rated on the modified Waldrop scale. The following craniofacial measurements were done by means of caliper

measurements: head circumference, trichion to glabella, glabella to nasion, glabella to subnasale, glabella to stomion, glabella to gnathion, tragion to trichion, tragion to subnasale, tragion to gnathion.

F. Five panel drug screen that tested for the presence of cocaine, metampethamine, MDMA, heroin and THC (Dis-Chem Drug Test).

G. 30-40 ml venous blood taken for genetic studies (see Wright et al. 2010 for a full description of DNA extraction method).

2.3 Data analysis

2.3.1 Latent Class Analysis: overview and background

A comparison of LCA and factor analysis regarding their roles in identifying unobserved or 'latent' constructs underlying observable symptoms or other measurable features has already been touched on in Chapter 1. In this chapter, the principles of LCA will be discussed in a heuristic fashion, following which, for the benefit of the more mathematically inclined reader, the statistical basis of LCA will be further conveyed by means of the equations used in the analyses. The section ends off with a description of the statistical procedures followed in the current study.

2.3.1.1 The aims of LCA

The aim of latent class analysis is to identify homogeneous latent classes of individuals. It assumes that an underlying categorical latent variable causes a response on a number of observed variables, such as the categorically defined item responses obtained from an instrument such as the SANS / SAPS. It involves estimating the smallest number of classes that explain the original (or manifest) variables obtained from the instrument. A particular model is fitted to the observed data, which is in the form of a contingency table. Under the null hypothesis the model represents the population from which the data was obtained. To test the model fit, cell counts expected under the assumption of conditional independence are compared with observed frequencies in each cell. The best fit occurs when the number of classes included in the model results in the null hypothesis *not* being rejected. Categorical concomitant variables (e.g., gender) can be included in the model as groups, and then the parameters are fitted conditional on the groups. Covariates that predict latent class membership can also be incorporated in the model. This is similar to incorporating covariates in a logistic regression, but now the outcome is latent rather than directly observed.

Thus LCA assumes that subjects in a given dataset fall into distinct, homogeneous classes which are termed "latent" because they are not directly obvious from a casual perusal of the data at hand. LCA assumes local independence, meaning that the observed variables are independent; conditional on the latent variable (i.e. each variable in a given latent class is independent of the other variables in that class).

In the context of schizophrenia research, consider a setting in which (using loosely defined criteria) symptoms a, b, c and d occur in patients with one of 3 subtypes of schizophrenia. Suppose that subtype X is associated with the presence of symptoms a, b and c, while subtype Y is associated with symptoms b, c and d and finally subtype Z with symptoms a, c and d. LCA will attempt to create patterns of associations in the symptoms listed above and thereby "create" latent classes ("subtypes" in the above example). LCA will always attempt to achieve classes in which

the symptoms are "conditionally independent" and the observed variables are statistically independent. The association between the observed variables is therefore explained solely by the classes of the latent variable (McCutcheon 1987).

Latent class models thus allow us to look at both unobserved latent variables (variables that are measured indirectly and with measurement error) as well as explanatory variable (where the focus is a group of variables and we are looking for an explanation of the relationships between them). Each individual's responses thus have two components, the true latent class and error. Indeed, latent class analysis can be viewed as a subset of structural equation modelling. The term "latent" means that an error-free latent variable, usually categorical in nature, is postulated. The latent variable is always indirectly measured by using two or more observed variables. The observed (or indicator) variables are subject to error.

As can be seen in the right hand drawing (Fig 2.1) the assumption of independence is violated as the variables X_1 and X_2 are not related through the latent variable alone, but also through the errors (e_2 and e_3).

Figure 2.1: Assumption of independence
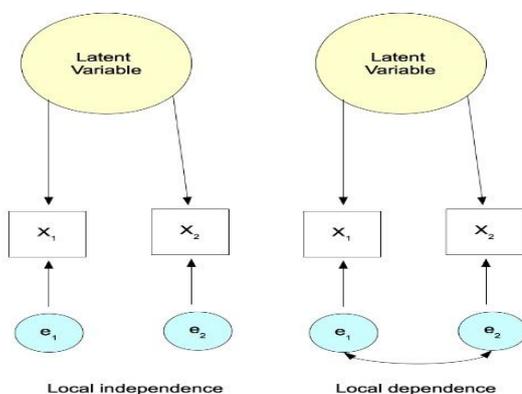
It is important to note that the assumption of independence does not mean that the observed variables in the set of data are independent, but that the latent classes are seen as locally independent of each other.

It has been proposed that LCA could be useful in providing a parsimonious and easily understood format for data interpretation and could thus help in identifying endophenotypes of schizophrenia (Castle et al. 1994). LCA can provide a hypothetical "probability-based" framework for subtyping schizophrenia and thus provide a basis for modelling proposed endophenotypes and examining their relationships with clinical variables. Subgroups of persons identified as belonging to a particular latent class or subtype of schizophrenia can be compared in terms of the existence of certain traits or characteristics. Thus, hypotheses can be tested regarding genetic influences on latent class membership, or associations can be sought between latent class membership and outcomes such as prognosis or treatment response. As continuing research broadens our understanding of the determinants and neurobiology of schizophrenia, targeted interventions may conceivably be developed which can be applied to specific subgroups of persons with schizophrenia.

2.3.1.2 History of the development and applications of LCA

The problem of evaluating the relationships (dependence versus independence) between two or more observed variables has a long history spanning at least two centuries. Where only two variables are involved this is a fairly simple process. However, in the social sciences, especially when dealing with responses to diagnostic rating instruments, different statistical methods are needed to make sense of the information obtained. Peirce (Peirce 1884) was the first to introduce latent structure models as a possible mechanism for dealing with the complexities of the

relationships between variables. However, it was only during the latter part of the 20th century that latent class model development gathered momentum. Although Lazarsfeld and Henry in 1968 discussed latent class analysis in depth and realised the potential of this method in social science research, it was only when Goodman (1974) was able to implement a method for obtaining maximum likelihood estimates of the latent class models that it achieved popularity as method of data exploration. Its applications were further broadened by placing it within the framework of log-linear models (Forman 1985, 1992), thus allowing for model fitting (Rindskopf 1986). This was followed by the introduction of covariates in 1988 into latent class analyses (Dayton and Macready 1988)

Although latent class analysis can be used to assess longitudinal data as well, for the purpose of understanding the statistical background of this study we will limit this discussion to cross-sectional data.

2.3.1.3 Constructing a contingency table of the possible response patterns

In order to complete a LCA the first step is to compile a contingency table (also known as a cross-tabulation).   As an example, consider the contingency table produced when subjects are categorised by their responses to three variables (e.g., hallucinations, delusions and flat affect), each rated dichotomously as being present (1) or absent (0).

There are 8 (2x2x2 or 2^3) possible response patterns that can be obtained:

(1 = presence of the attribute or symptom; 0 = symptom absent)

Response Pattern	Hallucinations	Delusions	Flattened affect
1	1	1	1

2	1	1	0
3	1	0	1
4	1	0	0
5	0	0	0
6	0	0	1
7	0	1	0
8	0	1	1

The three-way contingency table could be represented as follows:

	Flattened Affect present		Flattened Affect absent	
	HALLUCINATIONS PRESENT	HALLUCINATIONS ABSENT	HALLUCINATIONS PRESENT	HALLUCINATIONS ABSENT
Delusions present				
Delusions absent				

Each of these 8 cells would represent one of 8 mutually exclusive response patterns and each subject would be represented as one count in a specific cell, depending on his or her responses on the three indicator variables.

Thus each of these cells will contain the total number of participants that agreed to a specific symptom (variable) combination. With a larger number of variables, the number of cells can be very large (e.g., for 12 dichotomous variables there are 2^{12} or more than 4 000 possible response patterns). Before the introduction of LCA, one would have had to rely on visual inspection of the

table. LCA allows us to now organise and interpret the data in this table. The latent classes identified by the LCA therefore represent a group of participants that showed similar patterns of responses across the variables.

2.3.1.4 Assessing the goodness-of-fit of the model

The likelihood ratio chi square statistic (G^2) is used to compare estimated and observed cell proportions in the multi-way contingency table. G^2 approximates a chi square distribution with large samples and sufficient numbers of counts in each cell. Under the null hypothesis, the model fits the data. If the null hypothesis is rejected, it implies that the model does not provide a good fit to the data. Generally, the more classes are added to the model, the lower the G^2 value becomes until the null hypothesis is not rejected. When we now test the proposed model for fit, it should show a high degree of fit if the estimated cell proportions are close to the observed cell proportions. We therefore look for a model where the null hypothesis is not rejected. The larger the G^2, the smaller the probability that the null hypothesis will be supported. The G^2 value is interpreted relative to the degrees of freedom (as discussed more fully below).

2.3.1.5 Estimating population parameters from the sample data

LCA uses the responses obtained from the sample to estimate two sets of population parameters.

a. Gamma (γ) parameters (probability of membership of or belonging to a latent class). These are estimates of how prevalent each particular class is in the population under investigation, i.e., the probability that a randomly selected member of the population would fall in a specific class. The

estimated values for this parameter sums to 100 % across the latent classes, taking rounding errors into account. For example, if in a 3-class model, the proportions of the population falling in two of the classes have been estimated, the proportion in the 3rd class is assumed to be 1 minus the sum of the probabilities of belonging to the first 2 classes. This implies that in LCA each individual belongs to one and only one latent class.

b. Rho (ρ) parameters (conditional or item response probabilities) refer to the probability of a specific response to each variable conditional on membership in a given latent class. Once latent classes have been identified, the researcher is responsible for labelling each of these classes with a descriptive label that conveys the essence of this group in terms of clinical significance within the field of study. The posterior probabilities in Latent class analysis (LCA) refer to the probability that a specific observation is classified in a given class. The rho parameters are posterior probabilities, in that they represent the probability that a randomly selected subject, *given that* he/she belongs to a particular class, will manifest a certain symptom, e.g., flattened affect. The response probabilities are extremely important to the researcher as they are used to assign labels to the latent classes. For example, if the item response probabilities to items measuring negative symptoms tend to be higher in a particular class than in the other classes, this class might be given a descriptive label conveying the negative symptom "nature" of the subjects belonging to that class. These latent classes (with labels) often have qualitative and quantitative differences. The rhos give an indication of the strength of the relationship between each variable and each latent class and therefore the set of rhos gives us an indication of how well an individual fits into a latent class. It is important to take cognisance of the fact that the gamma's and rho's are probabilities and that the laws of probability apply to both.

Thus, the latent class model consists of estimated latent class prevalences and item response probabilities and these estimations are used to calculate the expected cell frequencies of the

contingency table. To estimate the parameters in the latent class analysis some form of iterative procedure such as expectation-maximization (EM) algorithm can be used (Dempster 1977). In essence EM finds maximum likelihood (ML) parameter estimates and ML estimates represent the parameter values for which data are most likely to be observed (Hagenaars et al. 2002). As this is an iterative process the researcher can use two methods to stop the procedure. The researcher can set the maximum number of iterations allowed or (and more importantly) can stop the iterations once a set of parameter estimates maximizes or almost maximizes the likelihood function (handbook LCA) also referred to as convergence. LCA uses maximum absolute deviation (MAD) between parameter estimates in two successive iterations as convergence index. The closer to convergence (as the procedure nears ML) the smaller the difference in parameter estimates between successive iterations becomes. A commonly used convergence value for MAD is $\leq$ 0.000001. Given these two criteria for stopping the iterations there are two outcomes: the procedure can converge – that is the convergence criteria were met before we ran out of maximum iterations or secondly, the procedure did not converge, meaning that the maximum allowable number of iterations were reached before the convergence criteria could be met. Bayesian estimation is an alternative to the use of ML estimation of parameters and seems to be more flexible (Chung et al. 2008).

2.3.1.6 Freely estimated and restricted parameter estimation

When estimating parameters it is important to realize that one can get freely estimated parameters (any value between 0 and 1) or restricted parameters (either fixed or constrained). The latter are not estimated during the procedure and the value of the parameter needs to be set prior to starting the procedure and cannot be changed during the procedure of iteration.

As more classes are added to the model, the more parameters need to be estimated and the more the unknown information increases. This amount can be lowered if one imposes restrictions on the parameters. A freely estimated (unrestricted) model allows individuals to endorse some responses that are not typical of the class they would be likely to be allocated to in the LCA. For example, supposing there were 3 dichotomous variables, delusions, hallucinations and flattened affect, a freely estimated model would permit subjects who experience hallucinations also in some cases to experience delusions and flattened affect.

Restriction of parameters decreases the number of parameters that are estimated (and improves the chances of achieving model fit). A restricted parameter can be either fixed or constrained. In the case of a fixed parameter, before estimation commences, a parameter is fixed to a specified value in the range 0 to 1. Thus that parameter is not estimated but fixed. Sometimes parameters can be constrained to be equal to each other (i.e., are assumed to show complete agreement across or within classes), taking on a value from 0 to 1. Only one parameter is estimated for the whole set ("equivalence set") of constrained parameters. A latent class model may be freely estimated or a mixture of unrestricted, fixed and constrained parameters

2.3.1.7 Model identification in LCA

A model is said to be "identified" if there are unique estimates of all parameters (i.e., one optimal solution exists). In order for the model to be identified, the model must have degrees of freedom greater or equal to 1 and not be under-identified. In essence this means that the "known" information exceeds the "unknown" information. The "unknown" information refers to the estimated parameters. For each class that is added to the model, more parameters (i.e., rho and gamma probabilities) need to be estimated. Thus, as the number of latent classes increases, the identification problems increase. The degrees of freedom depend on the number of independent

parameters (latent class probabilities and conditional probabilities) estimated and the number of latent classes specified by the model. For a 3 class model with 4 items, each with a dichotomous response, the number of conditional parameters = 3 * ((2-1) + (2-1) + (2-1) + (2-1)) = 12. Because the item response probabilities must add up to one, 1 is subtracted for each item (if the probability of one response (e.g., 'yes') is known (in the case of a dichotomous variable), the probability of a 'no' response is also known). The number of latent class prevalences (γ parameters) to be estimated = the number of classes − 1 (because the probabilities add up to 1). The total number of unique parameters estimated = the number of unique (or nonredundant) conditional probabilities + the number of unique latent class prevalences, which in the above example = 12 + 2 = 14.

Whether a model can be identified is determined by the number of degrees of freedom. If the degrees of freedom are less than 1, the model is not identified. The degrees of freedom can be calculated by subtracting from the number of possible response patterns in the multiway contingency table the number of unique parameters estimated and from this answer subtracting 1. Hence in the example above the df = 2^4 − 14 − 1 = 1. Since this is > 0, the model can be identified. If a model is not identified, one can decrease the problem by putting restrictions on parameters or placing them in equivalence sets. In addition, the sample set also needs consideration in this respect. The larger the sample, the more the information, the less sparse the data and the better the information available for exploring the relationship between the observed variables and the latent variable. This basically means that the better the homogeneity pattern and latent class separation of the item response probabilities are, the greater the amount of information that is available. One can increase information by either increasing the sample size or by reducing the number of variables from the contingency table, but this can also limit information via its impact on homogeneity and latent class separation, can lead to elimination of important variables and even result in negative degrees of freedom. Positive degrees of freedom is one of the ways to see whether a model needs to be simplified (i.e. simplify the model when degrees of freedom are negative). Another approach is to apply parameter restrictions as previously discussed. Parameter

restriction helps to simplify models, can be used to test a priori hypothesis or to test hypotheses about the equivalence of parameters. Parameter restrictions are used to test for equivalency across groups in terms of item-response probabilities or latent class prevalences. It is important to note that parameter restrictions will never decrease G^2 and most of the time increase G^2. Taking all of these changes to the model into consideration, it is still essential that the final model must make conceptual sense.

2.3.1.8 Choosing the latent class model

Choosing a latent class model is not a straightforward process. In essence the investigator needs to decide how many latent classes he or she should specify and whether any parameter restrictions are necessary. To assist the researcher in this decision the following should be considered:

13. Parsimony

14. model interpretability

15. test of absolute model fit

16. assessment of relative fit of competing models

17. cross-validation

Parsimony refers to the principle that if all else is equal, a simpler (estimating fewer parameters) model is preferred above a more complex model. Model interpretability refers to the researcher's ability to draw on current knowledge and hypotheses to meaningfully interpret the final model.

2.3.1.9 Relative model fit: comparing models

One can also look at relative model fit. In this scenario two models are compared to see which one fits best to a dataset (taking parsimony into account). There are two ways of doing this; via a likelihood-ratio difference test or by comparing the information criteria of each successive model. The likelihood-ratio difference test tests the null hypothesis that both models fit equally well. In case the null hypothesis is supported, then the more parsimonious (simpler) model is preferred. However, this test can only be applied if the two models are nested, i.e. if one model differs from the other only in terms of parameter restrictions and the two models have the same number of latent classes.

2.3.1.10 Information Criteria as model fit criteria

The alternative approach to using only the likelihood ratio chi square test as a criterion of model fit is to look at information criteria. These criteria include the Akaike information criterion (AIC; Akaike 1987) and the Bayesian information criterion (BIC; Schwartz 1978). Both these criteria impact on the G^2 statistic by reducing it in relation to the number of parameters estimated in your model and even the sample size.

AIC = G^2 +2P (P=number of estimated parameters)

BIC = G^2 + [log(N)]/P

In both these cases a smaller value indicates better fit and parsimony. However, because the BIC takes into account the sample size, it is possible that the two criteria might differ in model it supports.

2.3.1.11 Dealing with missing data

It is important to note that incomplete response patterns cannot be used to calculate the number of degrees of freedom. The latter is partly dependent on the number of cells in the contingency table. To get some idea of the response patterns one can visually inspect the table and look for cells with the largest number of individuals in it. However, one should be careful as missing data (and especially missing data that does not appear to be randomly missing), might skew conclusions drawn from your visual interpretation.

Missing data causes several problems; for example, individuals in the study who could not provide their correct age may differ qualitatively in their ability to answer questions on other issues as well. In addition, if a large number of individuals have missing data then the sample size can be significantly reduced. Corrections need to be made for missing data to address these concerns. The three possible underlying types of missing data are important when we consider how to manage missing data namely "missing completely at random" or "missing at random" or "missing not random" (Little & Rubin 2002). The last mentioned category, where the reason for the missing data is related to a variable outside of the dataset, is particularly important to detect, as "missing completely at random" or missing at random variables can be handled by LCA software but missing not at random cannot be handled by LCA software. The following two methods can be used to manage missing data:

1. Full-information maximum likelihood (FIML) and
2. Multiple imputation (MI) (Schafer 1997).

In FIML complete data and partially complete data on individuals are analyzed together and adjustments are made based on all the available information, the researcher only indicating to the

Page | 115

analysis software whether missing data is present. However, if the missing data occurs in a grouping variable or a covariate, then that individual's data is not analysed.   In MI on the other hand, so called "plausible" random values are imputed in the missing variable slots in multiple datasets. The advantage of MI is that so-called auxiliary variables can be used to help increase the accuracy of the MI procedure and that missing covariates do not exclude an individual from the analyses. Auxiliary variables are variables that will not form part of the final LCA model but may speak towards the reason for the missing variables. The disadvantage of the MI approach is that the model fit is much more complex and involves considering the model fit within each of the imputed datasets.

The G^2 statistic is influenced by missing data since the G^2 value calculation comprises of two components, firstly the lack of model fit component and secondly the degree of departure from MCAR (missing completely at random). G^2 can be inflated by the latter and therefore corrections for this are needed to arrive at an adjusted G^2. Now a p-value can be obtained by comparing the adjusted G^2 to the reference chi-square distribution that relates to the degrees of freedom in your model. However, this result can be impacted on by sparseness, also defined as the problem of small average expected cell counts. In cases where sparseness is significant i.e. the total sample size divided by the size of the contingency table is less than 5, the G^2 statistic becomes unknown and thus model fit becomes difficult to test. To address this problem two possible solutions are available i.e. parametric bootstrap (McLachlan & Peel 2000) and posterior predictive checks. The reader is referred to these papers for a detailed discussion on the methodology and application of these methods. It is also important to note that one should consider using multiple starting values (at least 10) and then look at their convergence consistency. This number can be increased to 100 or more if the convergence pattern is suboptimal.

2.3.1.12 Statistical basis of LCA

In order to understand the reasoning behind latent class analysis the following will assist the reader in appreciating its underlying statistical basis. In this example two observed variables are measured (dichotomous or polytomous).

1. Variables:

Variable A has I number of classes (i=1,2,3,...,I) and variable B J number of classes (j= 1,2,3,...,J). A third variable C is a latent or unobserved dichotomous or polytomous variable that has T number of classes (t=1,2,3,...,T).

1. Formula

$\pi_{ijt}{}^{ABC}$ = joint probability that an observation is in class i on variable A and in class j on variable B and class t on variable C.

$\pi_{it}{}^{\bar{A}C}$ = the conditional probability that an observation is in class i on variable A, given that the observation is in class t on variable C.

$\pi_{jt}{}^{BC}$ = the conditional probability that an observation is in class j on variable B, given that the observation is in class t on variable C.

$\pi_{t}{}^{C}$ = the probability that an observation is in class t on variable C.

Therefore: $\pi_{ijt}{}^{ABC} = \pi_{t}{}^{C} \pi_{it}{}^{\bar{A}C} \pi_{jt}{}^{BC}$, for i = 1,2,3,...,I; j=1,2,3,...,J; t=1,2,3,...,T.

Given the class level on variable C, variables A and B are viewed as conditionally independent of each other according to the stated model:

$\pi_{ijt}{}^{ABC} = \pi_{ijt}{}^{ABC} / \pi_t{}^C = \pi_{it}{}^{\bar{A}C} \pi_{jt}{}^{BC}$ (where $\pi_{ijt}{}^{ABC} = \pi_{ijt}{}^{ABC} / \pi_t{}^C$ is the conditional probability that an observation can be found in class i on variable A and in class j on variable B, given that the observation is in class t on variable C.

LCA is thus based on the statistical concept of likelihood. Cases are thus not absolutely assigned to a specific class, but rather to a probability of belonging to a class and thus a class can be characterized by a pattern of conditional probabilities. In other words, the chance that variables take on certain values. The method depends on the estimation of parameters for class profiles and the size of each class and can deal with all types of data including binary, count and continuous data.

2.3.2 Analytical procedures used in the current study

In this study the SANS/SAPS items were dichotomized into present/absent categories and LCA was conducted using SPSS (version 16.0 and add-on module).

2.3.2.1 SANS / SAPS Variables selected for the analysis

For the purposes of this study, only the following 12 SAPS / SANS global subscale scores or individual items were included in the latent class analysis. These included:

1. Eye contact,
2. Affective non-responsiveness,
3. Spontaneous movement,
4. Grooming,
5. Recreation.

6. Auditory hallucinations,

7. Commenting voices,

8. Global hallucinations score

9. Thought insertion,

10. Control delusions,

11. Mind reading,

12. Global delusions score

The reasons for choosing these variables in the latent class analysis were:

1. In the original sib-pair study (Niehaus et al. 2005), a review of the available literature on subtyping in schizophrenia performed by the author identified these items as being relevant, and a theoretical basis for their selection was constructed for each item.

2. These 12 items showed high concordance rates in the sib-pair study (Niehaus et al. 2005). See Niehaus et al. (2005) for a description of the theoretical basis for the inclusion of these items in studies aiming to identify intermediate endophenotypes for schizophrenia.

2.3.2.2 Estimating the number of latent classes

The number of latent classes was estimated using the following model fit criteria: the Bayesian Information Criterion (BIC), the Akaike Information Criterion (AIC) and the Likelihood ratio test. Lower values of the AIC and BIC indicate the best fitting model. Large differences in BIC and AIC when comparing different models indicate good fit for the model with the lowest values.

2.3.2.3 Subgroup comparisons

In this study we were interested in finding out whether there are differences between subgroups in the data. In this case we need to remember that the grouping variable will be an observed variable such as cannabis use or abuse or gender.

How do we interpret differences between the groups? LCA differs from FA in that it involves assigning subjects to distinct categories rather than identifying related variables and thus in LCA the item response probabilties may be identical between the groups and thus we can directly interpret the latent class prevalences because differences are quantitative and not qualitative.

2.3.2.4 Parameter estimation

We estimated different sets of parameters from the model. The Gamma parameters functioned as class membership probabilities, summing to 1 over the classes. Class membership probabilities estimate the proportion of the population falling in each class. The gamma parameters were calculated as functions of the covariates. The second set of parameters, the rho parameters, functioned as item-response probabilities conditional on latent class membership. The item response probabilities represent the probability of a particular variable (i.e., item in a questionnaire) being endorsed or manifested by an individual, given that he or she has been allocated to a specific latent class.

2.3.2.5 Criteria for evaluating and labelling item-response probabilities

Homogeneity within a given class and separation (pattern of probabilities clearly differentiating among the latent classes) were the two criteria used for evaluating the overall pattern of item-response (conditional) probabilities. The beta parameters are logistic regression coefficients for the covariates, predicting class membership, where the last class is the reference class.

2.3.2.6 Effects of covariates and subgroups

We also wanted to investigate the effects of gender, cannabis, duration of illness (defined as period from first reported behavioural changes or psychosis to date of interview) and sib-pair status (whether or not subjects had an affected sib in the sample) on the model fit. For example, we wanted to test whether the proportion of persons in each latent class varied according to the level of a categorical group (e.g., gender) to which they belonged. We did this by restricting the parameters to be equal across the gender groups (i.e., male and female) in one model and not in another. The G^2 difference test was used to test the significance of the differences between the two models. A significant test would suggest that membership in the various latent classes varies by gender. These differences might manifest themselves in either the item-response probabilities or the latent class probabilities. If the item-response probabilities are equal across groups (i.e. measurement invariance), it means that the latent class prevalences can be compared directly since the interpretation of the latent classes is identical across groups.

This phase of the study describes the application of latent class analysis (LCA) in a sample of 734 Xhosa-speaking schizophrenic subjects using factor analytically derived variables previously identified in an independent sample of this population (Niehaus et al. 2005). LCA was performed

Page | 121

according to the abovementioned guidelines on the following 8 SANS and SAPS items identified by preliminary exploration of the data: eye contact, auditory hallucinations, global hallucinations score, global delusions score, grooming, affective non-responsiveness, spontaneous movement, and commenting voices.

However, in schizophrenia, as previously discussed, the debate about whether the underlying structure of schizophrenia is categorical or dimensional is an ongoing one. In the categorical view, schizophrenia can be represented as diagnostic categories that indicate a dichotomous view of ill versus healthy or belonging to subtype one or two. Some of the reasons for the popularity of this viewpoint are that it meets clinical needs and allows for ease of monitoring and health care planning for health authorities and insurers (Muthén 2006). Alternatively, schizophrenia can be considered dimensional in nature and as such each individual displays a certain amount (severity) of disease, thus producing a continuous distribution.

Given the previously discussed advantages of FMM (including the use of both a categorical and dimensional components) it may be an appropriate model to investigate the underlying structure of psychiatric conditions (Yung 1997, McLachlan & Peel 2000, Muthén 2006,). Despite this advantage FMM has not yet been used extensively in this field (Muthén & Asparouhov 2006a, Muthén & Asparouhov 2006b). The main reasons for this might be that there is very little research on how a well-fitting model should be interpreted and even how this model should be applied in practice. It is however important that we take note of this method as the categorical versus dimensional debate is an ongoing issue in psychiatry and DSM is a good example of a shift from categorical (DSM-IV) to more dimensional (DSM-V to be released shortly).

2.4 Factor Mixture Modeling

With the use of FMM the structure underlying schizophrenia is considered categorical (using latent class variables) and also dimensional (using continuous latent variables) (see figures below):

Figure 2.2 LCA (A), FA (B) and FMM (C) structure (Adapted from Clark et al. 2009)
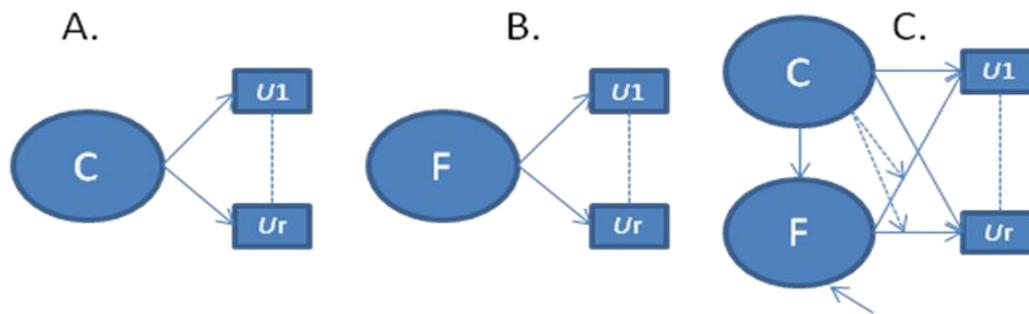
The latent class variable allows for the classification of individuals into groups or classes while the factor gives an indication of severity. The factor gives quantitative scores in the form of the factor scores. The FMM thus differ from FA in that all the parameters have the potential to be different across the classes. This is illustrated in figure 3 above by the dashed lines showing that the factor structure of the model can be different in each class.

The following formulas therefore represent a factor mixture model for $k = 1, 2, \ldots, K$ latent classes with dichotomous items (Clark et al. 2009):

$$y^*_{ik} = \tau_k + \Lambda_k \boldsymbol{\eta}_{ik} + \varepsilon_{ik},$$

$$\boldsymbol{\eta}_{ik} = \alpha_k + \zeta_{ik},$$

where,

$$\zeta_{ik} \sim N(\mathbf{0}, \boldsymbol{\Psi}_k)$$

and

yij = { 1 if $y*ij > 0$

0 *otherwise*.

The *k* subscript is the important change from FA modeling as this indicates class specificity. As all the parameters can be class specific, there is more flexibility in the model than with FA and allows for variations differing in the amount of measurement invariance. The advantage of not needing conditional assumption of independence allows one to explore the severity of the illness within a class and to quantify this value in terms of a factor score (Lubke & Muthén 2005, Muthén et al. 2006). In summary FA can thus be viewed as FMM with one latent class containing all the individuals in the sample while LCA is FMM with a factor covariance matrix that is zero (Clark et al. 2009).

Muthén (2008) divides factor mixture models into four branches or types depending on the amount of measurement invariance and whether the factor in each model is parametric or not. Of these branches the first 3 are of significance for this study. Branch 1 has measurement invariance and parametric factor distributions, while branch 2 has measurement invariance and non-parametric factor distributions. In branch 3 some or all the measurement parameters are non-invariant and the factor distribution is parametric.

FMM has been applied in genetic studies on micro-array gene expression data and on twin heritability data (McLachlan et al. 2004; Muthén et al. 2006).

Clark and Muthén (2009) in their paper "Models and strategies for factor mixture analysis" explain the different models that can apply. The first model, seen as part of branch two (Muthen 2008), is the latent class factor analytic (LCFA) model where the only changeable parameter

across classes is the factor mean as illustrated by the following formula (indicated by the subscript k on α):

$$y*_{ik} = \tau + \Lambda \eta_{ik} + \varepsilon_{ik},$$

$$\eta_{ik} = \alpha_k, \quad (6)$$

As the item thresholds and factor loadings do not vary across classes the disease is measured the same way across all classes. As the factor covariance matrix, Ψ, is fixed at zero, the implication is that there is no measure of severity in the disorder.

Figure 2.3: FMM-1 diagram and factor distribution plot (Adapted from Clark et al. 2009)
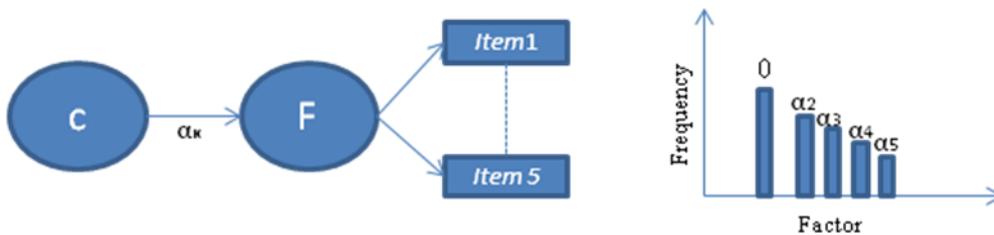
The model is illustrated on the left side of figure 4. In this model the latent class variable points to the factor and thus indicates that class membership is dependent on the participant's location on the factor. The plot on the right side shows an example of the factor means versus the frequency of each class. From this diagram it can be seen that this model has a non-parametric factor distribution as the factors have no variance. On a clinical level this implies that participants only differ from one another in terms of having different amounts of the disorder and the limited number of amounts of illness is the number of classes. The factor is thus categorical with each latent class representing a category.

The second model (Branch one type; Muthen 2008)is also called a mixture factor analysis and differs from the first model only in terms of the fact that the factor variances and covariances to zero is not set to 0, but are freely estimated in each of the classes. The following formula represents this model:

$y^*ik = \tau + \Lambda \eta ik + \varepsilon\ ik,$

$\eta ik = \alpha k + \zeta ik,$

where,

$\zeta ik \sim N(0, \Psi k)$. (7)

This model is illustrated in the diagram below.

Figure 2.4: FMM-2 diagram and factor distribution plot (Adapted from Clark et al. 2009)
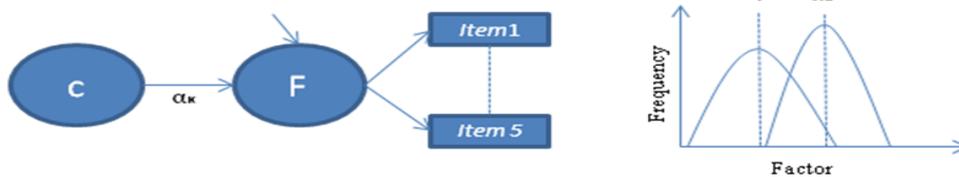
The major difference between this model and the first is that there is now a residual (see arrow pointing to latent factor) and thus within-class factor variance is modeled. There are therefore many possible amounts of the disorder a participant can have. Variations of this model is now possible, for example the factor variance can be allowed to be freely estimated but held constant across the classes. The factor is continuous, non-normal and a mixture of normal distributions located at different points on the factor distribution. Each normal distribution is thus a latent class.

In both model 1 and 2 the factor loadings and item thresholds are invariant across classes and therefore it means that the factor has the same meaning at both the low and high values in the

population. In both these models the latent classes are used to model the non-normality of the factor in the population. In real-life situations these two models seldom fit real data well as they are too restrictive because of the need for invariant factor loadings and thresholds.

The third model (an example of branch 3, Muthen 2008) has invariant factor loadings and an invariant factor covariance matrix but the item thresholds are allowed to vary across classes. The following formula explains this concept (Clark et al. 2009):

$$y^*ik = \tau k + \Lambda \eta ik + \varepsilon ik \,,$$

$$\eta ik = \zeta i \,,$$

where,

$$\zeta ik \sim N(\mathbf{0}, \mathbf{\Psi}). \quad (8)$$

(the factor mean does not appear in the formula above as it is set to 0 for identification purposes)

It can be seen in figure 6 that the main difference between this model and the previous ones is that the arrows starting from the latent class variable now points to the items, and not the factor.

Figure 2.5: FMM-3 diagram and factor distribution plot (Adapted from Clark et al. 2009)
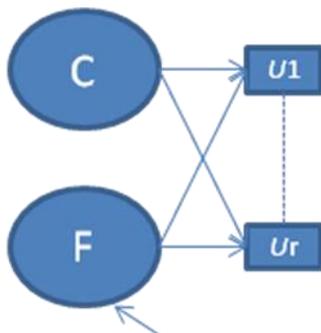
This implies that classes are now formed by the item responses and not by the factor.

The fourth model variation is similar to the third model except for that fact that the factor

covariance matrix is not invariant and can therefore change across the classes. This implies that there is may be different amounts of severity within each class. In practical terms this means that for a participant within a narrowly defined familial schizophrenia study, the "unaffected" (not fulfilling full DSM-IV TR criteria) class may have less variance because individuals are showing no and very little symptoms while a class with "Affected" participants might have more variation as there is likely to be a greater range of symptoms.

The mathematical basis of this model is illustrated below (Clark et al. 2009):

$y*_{ik} = \tau_k + \Lambda \eta_{ik} + \varepsilon_{ik}$,

$\eta_{ik} = \zeta_{ik}$,

where,

$\zeta_{ik} \sim N(\mathbf{0}, \Psi_k)$. (9)

The fifth and final model is the least restrictive model. In this model the item thresholds, factor loadings, and factor covariance matrix are all allowed to vary across classes.

This variance is illustrated by the following formula (Clark et al. 2009):

$y*_{ik} = \tau_k + \Lambda_k \eta_{ik} + \varepsilon_{ik}$,

$\eta_{ik} = \zeta_{ik}$,

where,

$\zeta_{ik} \sim N(\mathbf{0}, \Psi_k)$. (10)

It is thus clear that the FMM models differ in one important measure: measurement invariance. In order to interpret the results of this modeling the reader should have a clear understanding of what measurement invariance refers to. Meredith's (1993) definition of strong factorial invariance is

considered appropriate for use in our study setting (Clark et al. 2009, Little 1997, Widaman & Reise 1997). This definition requires equality of factor loadings, λ, and item thresholds, τ, across latent classes. Applying this definition to the data, different scenarios can thus arise depending on invariance in each of these items in the definition (Clark et al. 2009):

a. the factor loadings can be non-invariant in the latent classes and therefore a unit increase in the factor score will not lead to a similar increase of the dependent variable in the different classes. When this happens it might be interpreted to mean that there are many items with large differences in their factor loadings between classes and thus each class has a different underlying factor. However it is also possible that if there are only a limited number of items with differences in their factor loadings between classes, then those items (with differential loadings) may function differently in each class.

b. the item thresholds can be unequal across latent classes. In this scenario one latent class scores consistently higher or lower than the others, independent of scores on the factor. It is important to note that this implies that the observed differences between classes may not be entirely due to differences in the factor.

Taking this into consideration, it means that in FMM models 3 and 4 (with the non-invariance of the item thresholds) and FMM model 5 (with non-invariance of both items thresholds and the factor loadings) there is a violation of the strong factorial invariance. This then means that the same factor does not apply to the whole population and each latent class has a different factor, with a different interpretation. This creates a situation where there are several populations, represented by the latent classes and each with their own distributions.

To identify items which violate measurement invariance, we can use three possible methods (Clark et al. 2009):

a. compare the estimated and observed item means. A difference between these two indicate item non-invariance.

b. examine the within-class residuals for each item's mean from an FMM 1 or 2. A non-invariant item may have a larger residual than invariant items.

c. conduct a series of analyses in which one item's thresholds are held invariant while the thresholds for the other items vary freely across the classes. Then test whether there is a significant difference between the items whose thresholds are allowed to vary and the others. If the difference is significant, then it means that those item thresholds are non-invariant. This process should be followed with each item and the more times an item has a significant difference test, the higher the likelihood that that item should have freely varying item thresholds.

In addition to understanding the above concepts, it is also important to realize that, similar to FA, FMMs can have a confirmatory or exploratory factor measurement structure. Confirmatory FMMs uses a simple measurement structure. This means that substantive theory is used to define factors. These factors are then measured only by specific symptom items. The factor have no influence on the remaining items. Latent variables need to be based in substantive theory and the researcher needs to formalize their measurement hypothesis beforehand. This approach can of course aid in the search for more parsimonious models (Asparouhov & Muthén 2009). The problem is however that the simple structure and ignoring cross-loadings, may lead to a too parsimonious model and poor model fit.

When the researcher is dealing with situations where limited measurement knowledge is available and a complex model is expected, exploratory FMM might be the better choice (Asparouhov & Muthén 2009) given the disadvantages of the confirmatory FMM.

FMM is still a relatively new statistical model and therefore no exact steps are available, but rather just guidelines (Muthén 2006, Muthén & Asparouhov 2006, Muthén et al. 2006). The guidelines suggest that the data be first analyzed using latent class, then factor analytic and then factor mixture models. The fit among these three types are then investigated. A suggested strategy to decide on the number of classes and factors is to first fit LCA models with increasing number of classes and FA models with increasing number of factors. This then is followed by trying to fit an FMM with two classes and one factor to the data. Subsequent models are then used with increased number of classes. This is then followed by a step where the number of factors would be increased to two and the classes increased in subsequent models. The problem is to decide when to stop increasing the number of factors and classes. It is suggested that the endpoint in model building should be determined by the number of classes from the best fitting LCA model and the number of factors from the best fitting FA model (Clark et al. 2009). Now the best FMM model for the data will be selected. Following this the model should be compared to the best fitting LCA and FA models. This comparison will allow the researcher to explore if a more parsimonious solution can provide a better fit and explanation of the data. These models are compared on the basis of information criteria. The lowest value will indicate the best fitting model type.

This exploratory phase of the study describes the application of factor mixture modelling (FMM) in a sample of 734 Xhosa-speaking schizophrenic subjects using factor analytically derived variables (mean scores based) previously identified in an independent sample of this population (Niehaus et al. 2005). FMM was performed according to the abovementioned guidelines on the following 10 SANS and SAPS items identified by preliminary exploration of the data: poor eye contact, auditory hallucinations, global hallucinations score, conversing voices, global delusions score, grooming, affective non-responsiveness, global bizarre behaviour, decreased spontaneous movement, global alogia and grooming behaviour.

2.4 References

Agresti A. Categorical data analysis. Wiley, New York, 1990.

Akaike H. Factor analysis and AIC. Psychometrika 1987 52:317-332.

Andreasen NC. The Scale for the Assessment of Negative Symptoms (SANS) Iowa City, IA, University of Iowa 1983.

Andreasen NC. The Scale for the Assessment of Positive Symptoms (SAPS) Iowa City, IA, University of Iowa 1984.

Barnes TRE. A rating scale for drug-induced akathisia. British Journal of Psychiatry 1989 154:672-676.

Buchanan RW, Heinrich DW. The neurological evaluation scale (NES): A structured instrument for the assessment of neurological signs in schizophrenia. Psychiatry Research 1989 27:335–350.

Castle DJ, Sham PC, Wessely S, et al. The subtyping of schizophrenia in men and women: a latent class analysis. Psychological Medicine 1994 24:41-51.

Chung H, Lanza ST, Loken E. Latent transition analysis: inference and estimation. Statistics in Medicine 2008 73:1834-1854.

Clark SL, Muthén B, Kaprio J, et al. Models and strategies for factor mixture analysis:

Two examples concerning the structure underlying psychological disorders. 2009. www.statmodel.com.

Dayton CM, Macready GB. "Concomitant-Variable Latent Class Models," Journal of the American Statistical Association 1988 83:173-178.

Diagnostic and Statistical Manual of Mental Disorders, 4th ed. Text Revised (DSM-IV-TR). American Psychiatric Association,Washington, DC, 2000:803–805.

Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society 1977 Series B 39:1-38.

Dubertret C, Bardel C, Ramoz N. A genetic schizophrenia-susceptibility region located between the ANKK1 and DRD2 genes. Progress in Neuro-Psychopharmacology and Biological Psychiatry 2010 34(3):492-499.

Formann AK. "Constrained Latent Class Models: Theory and Applications," British Journal of Mathematical and Statistical Psychology 1985 38:87-111.

Formann AK. Linear logistic latent class analysis for polytomous data. Journal of the American Statistical Association 1992 87:476-486.

Goodman LA. Exploratory Latent Structure Analysis Using Both Identifiable and Unidentifiable Models. Biometrika 1974 61:215-231.

Guy WA. Abnormal Involuntary Movement Scale (AIMS). ECDEU Assessment Manual for Psychopharmacology. U.S. Department of Health Education and Welfare, Washington, DC, 1976:534–537.

Hagenaars JA, McCutcheon AL. Applied Latent Class Analysis. Cambridge University Press, 2002.

Hoon Jeong S, Joo E-J, Min Ahn Y. Association study of dopamine transporter gene and schizophrenia in Korean population using multiple single nucleotide polymorphism markers. Progress in Neuro-Psychopharmacology and Biological Psychiatry 2004 28(6):975-983.

Koen L, Niehaus DJH, De Jong G, et al. Morphological features in a Xhosa schizophrenia population. BMC Psychiatry 2006 6:47.

Lazarsfeld PF, Henry NW. Latent Structure Analysis. Boston, Houghton Mifflin, 1968.

Little RJ, Rubin DB. Statistical analysis of missing data. Second Edition. Wiley New York, 2002.

McLachlan GJ, Do KA, Ambroise C. Analyzing microarray gene expression data. Wiley New York, 2004.

McCutcheon AC. Latent Class Analysis. Beverly Hills: Sage Publications, 1987.

Muthen B, Asparouhov T. Item response mixture modeling: application to tobacco dependence criteria. Addiction Behavior 2006 31(6):1050-1066.

Muthen B. Latent variable hybrids: Overview of old and new models. In Hancock GR and Samuelsen KM. (Eds.). Advances in latent variable mixture models. Information Age Publishing, Inc. Charlotte, NC, 2008.

Niehaus DJH, Koen L, Laurent C, et al. Positive and negative symptoms in affected sib pairs with schizophrenia: implications for genetic studies in an African population. Schizophrenia Research 2005 79:239– 249.

Nurnberger JI Jr, Blehar MC, Kaufmann CA. Diagnostic interview for genetic studies. Rationale, unique features, and training. NIMH Genetics Initiative. Archives of General Psychiatry 1994 51(11):849-859.

Peirce CS. The numerical measure of the success of predictions. Science 1884 4:453-454.

Rindskopf R, Rindskopf W. "The Value of Latent Class Analysis in Medical Diagnosis," Statistics in Medicine 1986 5:21-27.

Schafer JL. Analysis of incomplete multivariate data. CRC Press, Boca Raton, Florida, 1997.

Schwartz G. Estimating the dimension of a model. The Annals of Statistics 1978 6:461-464.

Tang Y-L, Gillespie CF, Epstein MP. Gender differences in 542 Chinese inpatients with schizophrenia. Schizophrenia Research 2007 97:88-96.

Waldrop MF, Pedersen FA, Bell RQ. Minor physical anomalies and behavior in preschool children. Child Development 1968 39:391–400.

Wright GE, Niehaus DJ, Drögemöller BI. Elucidation of CYP2D6 genetic diversity in a unique African population: implications for the future application of pharmacogenetics in the Xhosa population. Annals of Human Genetics 2010 74(4):340-50.

CHAPTER 3 RESULTS

Content

CHAPTER 3 RESULTS

3.1 Overview of Chapter 3

In this chapter relevant demographic and clinical data and the results of the latent class analysis are presented. Clinical features as defined by the SANS and SAPS are divided into lifetime symptoms and those occurring at or within 30 days of the interview.

3.2 The patient sample

A sample of 737 (604 [82%] male and 133 [18%] female) Xhosa subjects with schizophrenia, from the Western and Eastern Cape provinces of South Africa, participated in the latent class analysis study. Included in this number were 214 participants (173 [80.8%] male and 41 [19.2%] female) who formed part of the sibling pair sample used in the initial study (Niehaus et al 2005). The majority of subjects (n=578 [78.4%]) were recruited from the Western Cape Province. Among this group, 65.2% (n=377) were permanently residing in the Southern Metropole, including the Overberg and Worcester areas, and 19.2% (n=111) in the Northern Metropole, including the Paarl and Stellenbosch areas. The remainder of the sample lived in the Eastern Cape. The majority of the Eastern Cape group, living mainly in Port Elizabeth, East London and the rural areas surrounding these cities, contributed only 32 patients to the study. Three residents of the Eastern Cape were working in Johannesburg and their interviews were conducted in Gauteng. A large group of participants (17.3%; n=128) had no permanent residence and moved between areas. Two hundred and thirty two probands (31.5%) reported an urban upbringing (defined as being raised for the first 18 years of life in a city of more than 500 000 inhabitants) while 201 (27.3%) had a rural upbringing, mostly in rural Eastern Cape. In 304 cases (41.2%) we were unable to make a clear distinction between rural and urban upbringing on the available information.

3.3 Characteristics of the subjects

3.3.1 Demographic and other background characteristics

See Table 3.1 for a summary of the patient characteristics. At the time of interview, the majority of patients were single (79.4%). Of the remaining participants, 11.5% were married, 5.8% separated or divorced, and 2.8% widowed.

In 66.7% of cases (n=492) we were able to confirm their permanent accommodation status: 234 (47.6%) lived in a family home, 180 (36.5%) were shack dwellers in informal settlements, 34 (6.9%) lived in their own homes, 25 (0.5%) lived in rented houses, 10 (2%) were living in a hospital, hostel or shelter, 5 (1.0%) in a rented room and 3 (0.6%) were homeless. Four hundred and fifty one individuals were able to provide information on the length of stay at their current permanent residential address. The mean duration of stay at the current residence was 3 years (155.0 weeks, SD 152.7, range 1 week to 16 years). Indeed 33.7% of the sample had stayed at their current permanent residence for one year and 19.5% had stayed for longer than 5 years. The residences where the individuals were currently staying had an average of 3.12 rooms (SD 1.6; range 1-8; information available for 489 of the cases) and 4.68 (SD 2.44; range 1-14) people per residence. The levels of crowdedness within residences were similar to that they had experienced during their developmental years (i.e. up to the age of years) with a mean number of 3.05 (SD 2.30; range 1-15) people per room (least crowded period) and 3.86 (SD 3.21; range 1-25) people per room (most crowded period).

The mean maternal age at birth of probands (data available only for 246 cases) was 27.26 years (mean; SD 7.73; range 15-54) with the mean paternal age at birth (available in 160 cases) 35.47 years (mean; SD 10.00; range 17-68). Sibships ranged in size and in 485 cases sibship

information was available. Of these, 142 (19.3%) were eldest children and this included twenty

three (4.8%) probands who were single sibships (see Fig 3.1)

Fig 3.1 Sibship size (excluding the proband)

Sibship size excluding the proband
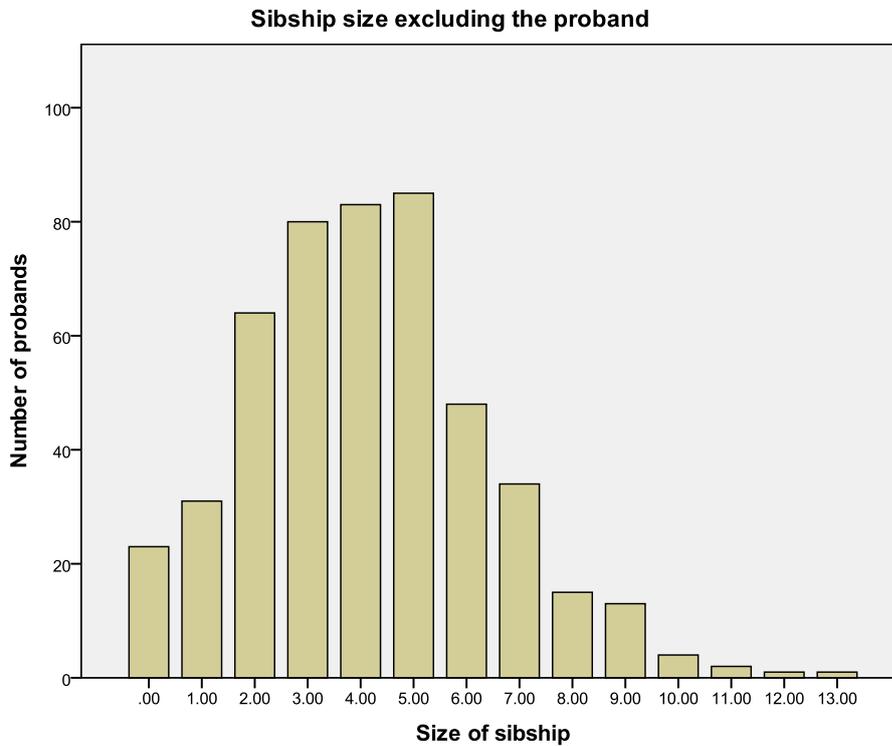
More than 90% of the participants were affiliated to a religious movement or church, of which the

protestant movement was the most common (59.7%, n=440). Sixty eight percent of the

participants received disability grants and 4.6% were either gainfully employed or students

(0.95%). Slightly fewer than 5% of the participants had failed to complete at least 1 year of

schooling, while only 1.6% had attained a level of education of grade twelve or higher (see Fig

3.2).

Fig 3.2 Highest level of education

Highest level of education (n=719)

* 0 = no official schooling received, ** 13 = one year post grade 12 education, ***15 = 3 years post grade 12 education.
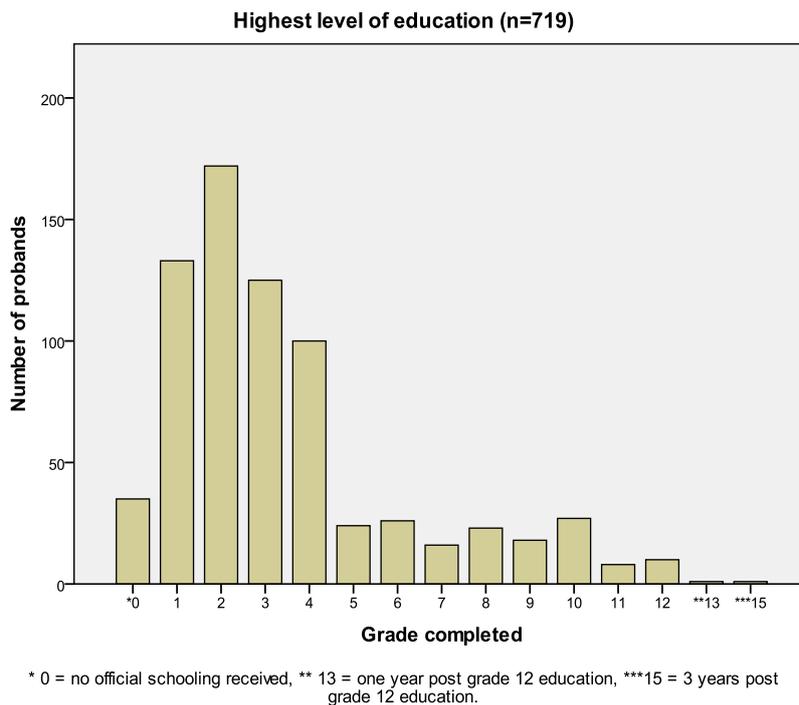
The median number of grades completed was 3 (mean 3.47, SD 2.8) and the mean age at leaving formal education was 18.37 years (SD 3.27; range 5-30 years). In 59.3% (n=437) information was available (self-reported and confirmed by collateral information) regarding school performance. 31.1% (n=136) of this group (n=437) reported average marks below 50%, 26.5% (n=116) reported marks between 50 and 60%, 10.3% scored between 60 and 70% and 0.2% (n=1) reported marks above 70%. In 422 (57.3%) of cases we were able to ascertain reported failure in grades. One hundred and thirty seven (32.5%) of this group (n=422) never failed a grade, while 48 (11.3%) failed three or more times during their school career. Indeed, 28 (6.6%) failed grade 1, 15 (3.6%) grade 2, 26 (6.2%) grade 3 and 26 (6.2%) grade 4. Grade 7 was failed by 8.8% (n=37) and grade 10 by 6.2% (n=26).

The age at interview (n=734) ranged from 15 to 70 years (mean 35.9 years SD 10.60). The mean duration of illness (n=721) was calculated as 12.4 years (SD 9.41; range 6 months to 50 years)

Page | 142

(Fig 3.3). The mean age at onset (n=723) was 23.32 years (SD6.65; range 10 to 55 years) (Fig 3.4)

(Fig 3.5).

Fig 3.3 Duration of illness (in years)

Duration of illness (n=721)

Duration of illness defined as total time elapsed from age of onset
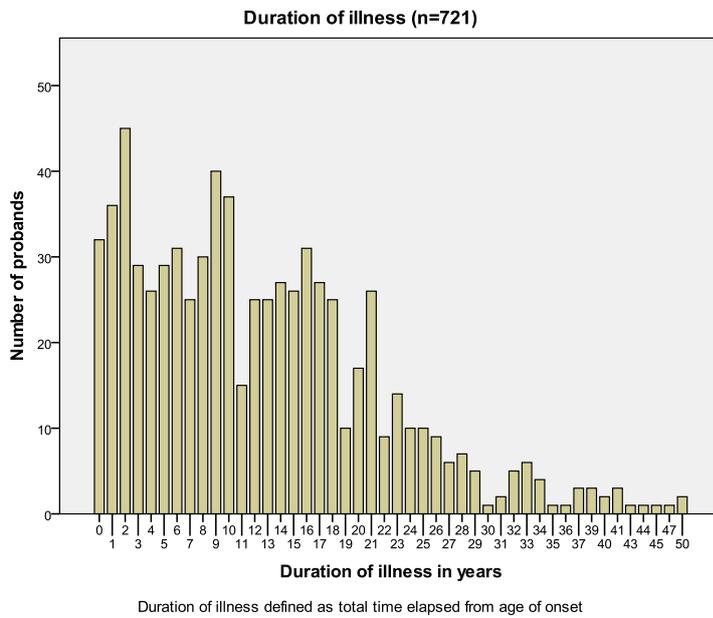
Fig 3.4 Age of onset of illness

Age at onset (n=724)

Age at onset defined as first reported presence of overt positive or negative symptoms
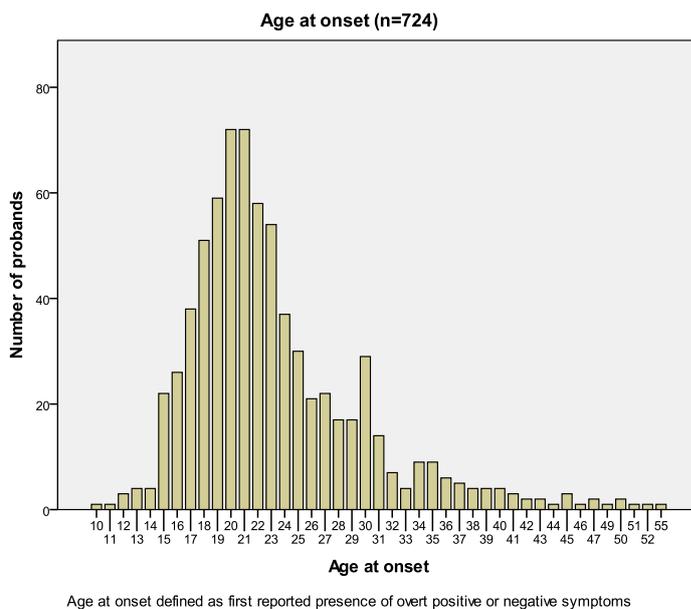
Fig 3.5 Age at interview

Age at interview (n=734)
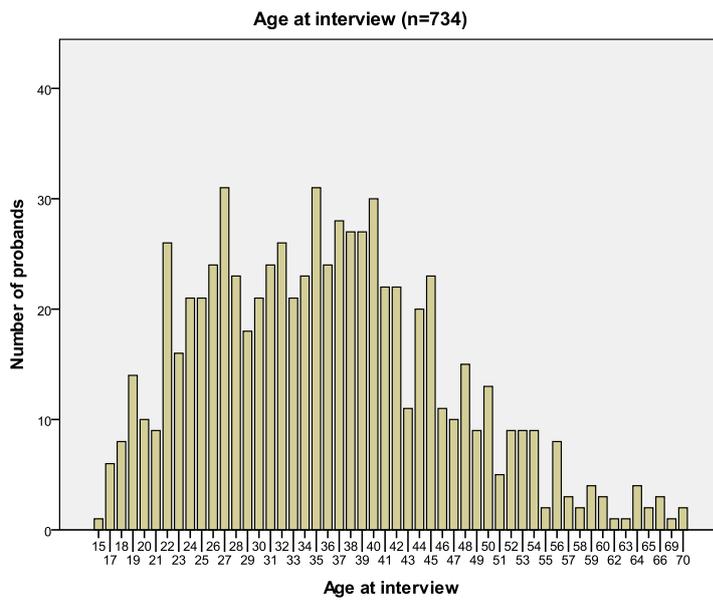
Table 3.1 Baseline characteristics of the 737 subjects

		n	%
Gender	Male	604	82.0
Marital status	Single	584	79.3
Employment status	DG	501	68.0
	Employed	27	3.7
	Student	7	1.0
Education*	Completed high school	12	1.6
	No schooling	35	4.7
Early developmental course (n=712)			
	Total number of developmental insults	42	5.9
	Perinatal complications	16	2.3
	Delayed milestones	10	1.4
	School difficulties	3	0.4

		mean (sd)	range
Age	at interview (y)	35.86 (10.60)	15-70
	of onset (y)	23.32 (6.64)	10-55
Duration	of illness (y)	12.40 (9.42)	0.5-50

* See paragraph on education for full details. DG = receiving disability grant.

3.3.2 Clinical assessment

3.3.2.1 Comorbid medical conditions and early developmental history

Early developmental incidents occurred in 5.9% of the participants (n=42). Of these, 16 cases of antenatal and intra-partum complications (including pre-eclampsia, forceps delivery and prematurity) were reported. One patient was born with dysmorphic feet and thirteen patients reported various childhood afflictions including possible meningitis (n=1), kwashiorkor (n=1), alcohol abuse during pregnancy (n=1), polio (n=1), febrile convulsions (n=1), loss of consciousness after trauma (n=2), loss/death of mother during or shortly after birth (n=3), complicated twin delivery (n=1) and one child had significant anger management issues during childhood. Ten patients were described as having had delayed milestones and two had repeatedly failed early school grades in the absence of slow milestones (see also Table 3.1).

3.3.2.2 Longitudinal assessment of schizophrenia course and symptomatology

Table 3.2 gives a summary of the patients' longitudinal course.

Table 3.2: Clinical features of schizophrenia: longitudinal course

		n	%
Onset	Acute	158	41.8
	Insidious	55	14.6
Precipitating life event		85	22.4
Lifetime delusions		708	96.1
Lifetime hallucinations		722	98.0
		mean (median)	SD (range)
Number of psychotic episodes		3.2 (2)	2.5 (0-20)
Number of psychiatric admissions		3.3 (1)	2.9 (0-20)

Psychosis

In 51.3% (n=378) of cases a history could be elicited regarding the prodromal period. 41.8% of the 378 patients had an acute onset (less than 1 week before onset of overt psychosis), while 14.6% had an insidious onset lasting more than 6 months. 22.4% of subjects reported a stressful life-event as the precipitating factor. The most common events included significant losses (death, financial) (6.4%), marital or relational conflict (5.4%), cultural rituals (2.6%), court cases, riots, MVA's and assaults (2.3%), pregnancy (2.2%), stress associated with schooling (1.6%) and drug use (0.5%).

Most patients (47.3%) reported 1 or 2 life-time episodes of psychosis (mean 3.2; SD 2.54; range 0 to 20) and the number of hospitalizations ranged from zero (6.0% of the sample) to twenty times (0.3% of the sample; mean 3.13; SD 2.93).

Very few participants (1.9%) denied residual symptoms. Establishing the exact duration of residual symptoms was difficult within this patient group and in 18.7% of cases no calculation could be made. Nine percent of the participants were floridly psychotic (as defined by a score of 8 or more on the sum of the global items for hallucinations and delusions on the SAPS) at the time of the interview.

Hallucinations and Delusions

98% (722 of 737) of the probands had experienced hallucinations as part of their life-time course of illness. The vast majority had experienced auditory hallucinations (96.6%; n=712), followed by visual hallucinations (52.5%; n=387), tactile hallucinations (28.4%; n=209), olfactory hallucinations (26.3%; n=194) and least commonly gustatory hallucinations (17.5%; n=129).

96.1% (708 of the 737) of the subjects had experienced delusions as part of their illness across their life-time. 88.5% of the total sample had had paranoid delusions during their lifetime, while delusions of reference (53.5%), grandiose (45.7%), religious (35.1%), mind reading (34.3%), broadcasting (26.3%), somatic delusions (25.9%) and delusions of control (22.1%) were also found in a substantial proportion of the probands. The least common delusions involved thought insertion (19.7%), thought withdrawal (16.8%), delusions of sin or guilt (16.4%), bizarre delusions (12.5%), jealousy (6.2%), erotomanic (5.2%) and nihilistic delusions (4.7%).

Behavioral features

Life-time behavioral abnormalities were common (90.2%), with aggression (verbal and non-verbal) occurring in a substantial number of subjects (73.4%). Bizarre behavior, including hoarding and arson, had occurred in 55.3% of the patients and catatonia was reported in 8% of the sample.

Thought disorder

Thought disorder occurred in 61.8% of the sample across their lifetime, with incoherence (29.3%) derailment (28.9%), tangentiality (27.3%), illogical thought pattern (20.1%) and circumstantiality (20.55) reported as distinct components.

3.3.2.3 Current mental state as assessed using the SANS / SAPS items of the DIGS

Table 3.3 gives a summary of the current mental state of the 737 patients in terms of whether they scored at least 2 on the SANS and SAPS. The median value for each item is also given. Global scores are given in Table 3.4. For a complete breakdown of each item and the criterion by which each symptom is rated, see Appendix 5.1.

Table 3.3 Frequency of positive responses to the SANS/SAPS items endorsed by the 737 Xhosa schizophrenia subjects in the month prior to the interview

	n	%	median score
Persecutory delusions	329	45.9	0
Delusions of jealousy	12	1.8	0
Delusions of sin or guilt	27	3.8	0
Grandiose delusions	131	18.5	0
Religious delusions	99	14.0	0
Somatic delusions	66	9.3	0
Delusions of reference	177	25.0	0
Delusions of control	74	10.5	0
Mindreading delusions	107	15.2	0
Thought broadcasting	80	11.4	0
Thought insertion	63	8.9	0
Thought withdrawal	57	8.1	0
Auditory hallucinations	353	49.2	1
Commenting voices	170	24.2	0
Conversing voices	215	31.5	0
Somatic / tactile hallucinations	86	12.1	0
Olfactory hallucinations	82	11.6	0
Visual hallucinations	128	18.1	0
Clothing and appearance	155	21.6	0
Social and sexual behaviour	196	21.2	0
Agitated/aggressive behaviour	214	29.7	0
Repetitive/stereotyped behaviour	50	6.9	0

Derailment	212	29.5	0
Tangentiality	203	28.3	0
Incoherence	164	22.8	0
Illogicality	145	20.2	0
Circumstantiality	163	22.7	0
Pressure of Speech	35	4.9	0
Distractible Speech	37	5.2	0
Clanging	3	1.1	0
Unchanging facial expression	562	78.6	2
Decreased spontaneous movement	415	56.8	2
Paucity of expressive gestures	463	64.5	2
Poor eye contact	398	54.6	2
Affective nonresponsivity	429	58.8	2
Inappropriate affect	134	18.4	0
Lack of vocal inflections	450	62.1	2
Poverty of speech	357	49.0	1
Poverty of content of speech	424	68.3	2
Thought blocking	19	2.6	0
Increased latency of response	111	15.4	0
Poor grooming/hygiene	434	59.5	2
Impersistence	604	83.3	3
Physical anergia	486	66.9	2
Decreased recreational interest	610	87.5	3
Decreased sexual interest*	343	73.6	3
Decreased need for intimacy	567	80.0	3

Restricted Relationships	622	85.9	3
Social Inattentiveness	149	20.4	0

NOTE: Each item was rated on a scale from 0 to 5.

Because of missing values, denominators were generally below 730. The response rate for the "Decreased sexual interest" item* was only 63%, but response rates for all other items were more than 95%.

Table 3.4 Global Scores on the SAPS and SANS

	n	%	median
Global Delusions score	378	53.1	2
Global Hallucinations score	360	50.5	2
Global Behavioural changes	280	39.0	0
Global Thought disorder	277	38.6	0
Global Affective changes	599	82.1	2
Global Alogia	446	61.2	2
Global Avolition/apathy	629	86.2	2
Global Asociality	665	91.5	3

Negative Symptoms

Negative symptoms are reported based on the SANS assessment and cover affective changes, alogia, avolition-apathy, asociality and anhedonia. Varying degrees of affective changes were reported (SANS based) in the vast majority of this group (88.1%) as reflected by the global score for affective changes. Concentration difficulties were reported in 25% of the sample.

3.4 Treatment

The participants used a wide variety of medications. Only a small minority (4.3%) denied taking their prescribed or suggested medication. In 10% of subjects, reliable information regarding medication use could not be obtained, because neither the patients nor the records could provide us with this information. Depot antipsychotics were still by far the most common treatment chosen by medical practitioners. Surprisingly, fewer than 5% of patients used clozapine. The depot antipsychotics were commonly used in combination with oral antipsychotic medication. Four cases received clozapine in combination with depot antipsychotics. Slightly more than 20% of the participants used anticholinergic medication. Mood stabilizers were given to eight patients and antidepressants to three.

3.5 Substance abuse and dependency

Thirty six percent (n) of participants had a history of possible substance use/abuse or dependency (cannabis and/or alcohol). Alcohol abuse was present in 16.1% of the participants. Only 1.6% of the sample admitted to symptoms consistent with a diagnosis of alcohol dependence. Cannabis abuse was more prevalent, with abuse or use diagnosed in 34.9% of cases. Only 3.4% admitted to cannabis dependence. Most participants (73%) had a history of tobacco smoking.

3.6 Comorbid mood and anxiety disorders

Seventeen percent of the sample were diagnosed or suspected of suffering from mood disorders (adjustment disorder with a depressed mood, dysthymia, major depression, hypomania or mania) while 4.7% had symptoms of anxiety disorders (panic disorder, phobias, OCD).

3.7 Latent class analysis

Contingency table

The 8 dimensional contingency table produced by cross-tabulation of the 8 dichotomously defined SANS/SAPS variables namely eye contact, auditory hallucinations, global hallucination score, global delusion score, grooming behavior, affective non-responsivity, spontaneous movement and commenting voices (see section 2.3.2.1, Methods) consisted of 256 cells representing 116 distinct observed response patterns (Appendix 5.2). The most frequently occurring combinations of responses were a cell with 82 observations (11.7%; the combination with all negative symptoms), a cell with 66 observations (9.5%; the combination with all positive symptoms) and 114 cells with the other combinations of symptoms (See appendix 5.2).

Choosing the best-fitting latent class model

Using the dichotomous responses to the 8 SANS/SAPS variables, various goodness-of-fit indices, including the likelihood ratio chi square test (G^2), the Akaike information criterion (AIC), the Bayesian information criterion (BIC) and Log-likelihood (l) were used to test inferences regarding

the number of latent classes that account for the heterogeneity existing in the Xhosa schizophrenia population.

In terms of the Akaike information criteria (AIC), Bayesian information criteria (BIC) and Log-likelihood (/), the latent class analysis revealed a best fit for an 8-item model with 4 classes (Table 1, See below for all models). In the unrestricted model the BIC and AIC reached a minimum of 219.4 and 380.2 respectively with the 8-item, 4-class model, but increased to 511.8 and 631.2, respectively with the 8-item, 3-class model. Similarly, in the 8-item, 4 class model the Log-likelihood was -2666.2 and -2821.4 for the 8-item, 3 class model, respectively. The variables thought insertion, control delusions, mind reading and recreation were thus excluded from the final model (see table 3.5).

Table 3.5 Fit table for all 8-variable models

Number of variables	Classes	G^2	df	AIC	BIC	*aBIC*	*l*
8	5	127.3	211	215.3	417.4	277.7	-2655.1
8	4	149.4	220	219.4	380.2	269.1	-2666.2
8	3	459.8	229	511.8	631.2	548.6	-2821.4

G^2: Likelihood-ratio statistic; AIC: Akaike information criterion; BIC: Bayesian information criterion; l: Log-likelihood; Chi diff: Likelihood-ratio difference test

The 8 variable model (see above) in the dataset (n=730; 7 had missing values) showed 115 response patterns and converged in 68 iterations using the maximum absolute deviation convergence method and a convergence criterion of 0.000001.

Table 3.6 Parameter estimates under the 4 class model for the 8 schizophrenia symptoms

Parameter Estimates				
Class membership probabilities: Gamma estimates (standard errors)				
Class	1	2	3	4
	0.26 (0.02)	0.18 (0.02)	0.31 (0.02)	0.25 (0.02)
Item response probabilities: Rho estimates (standard errors)				
Class	1	2	3	4
Auditory hallucinations	0.0010	0.9960	0.9888	0.0066
	(0.0027)	(0.0115)	(0.0159)	(0.0063)
Commenting voices	0.0004	0.7182	0.5730	0.0004
	(0.0015)	(0.0441)	(0.0361)	(0.0015)
Global halluc-inations score	0.0387	0.9988	0.9862	0.0357
	(0.0219)	(0.0033)	(0.0079)	(0.0159)
Global delusions score	0.2028	0.8475	0.8722	0.2283
	(0.0330)	(0.0337)	(0.0235)	(0.0322)
Eye contact	0.9409	0.1868	0.7981	0.0910
	(0.0210)	(0.0394)	(0.0329)	(0.0236)
Affective non-responsiveness	0.9298	0.1539	0.9644	0.0852
	(0.0222)	(0.0533)	(0.0186)	(0.0230)
Spontaneous movement	0.9470	0.1292	0.8944	0.0947
	(0.0187)	(0.0394)	(0.0301)	(0.0258)
Grooming	0.6851	0.4742	0.8522	0.2689
	(0.0350)	(0.0483)	(0.0257)	(0.0341)

Estimated latent class prevalences

Examination of the gamma parameters (i.e., latent class probabilities) reveals that 18 and 31 percent of the population of Xhosa schizophrenia subjects respectively are estimated to belong to the smallest and biggest of the 4 classes. About halve of the subjects are estimated to belong to either one of the other two classes.

Item-response probabilities

The item response probabilities represent probability of a particular variable (i.e., item) being endorsed or manifested by individuals, given that the individuals belong to the specific latent class. Item response probabilities are conditional probabilities and therefore cannot be looked at in isolation but one has to look at the pattern of item-response probabilities across all responses.

Two criteria defined a strong relationship:

1. Item response probabilities varied considerably over the latent classes. For example, the marginal probability for eye contact was 0.55, but conditional probabilities varied from 0.09 (class 4) to 0.94 (class 1).

2. Item responses were close to 0 or 1: In class 1 the probability of exhibiting poor eye contact (SANS) was 0.94 and that of having auditory hallucinations (SAPS) 0.001.

We used the item response probabilities to interpret the classes and to give the classes descriptive labels:

Class 1 (Negative symptom group) - poor eye contact, affective non-responsiveness, lack of spontaneous movement, problems with grooming: predominantly negative symptoms

Class 2 (Positive symptom group) - auditory hallucinations, global hallucination score, global delusion score, commenting voices: predominantly positive symptoms

Class 3 (Mixed symptom group) - negative and positive symptoms

Class 4 (Low symptom group) - a no/low symptom group

Predicting membership in the 4 latent classes (8-variable model).

Each subject was assigned to the latent class to which he or she had the highest probability of belonging, based on their response patterns. Table 3.7 shows the degree of uncertainty in the classification. This indicated that the mean for class 2 was slightly lower than that of the other classes (SD slightly larger), indicating that membership to this class was slightly more unsure than that of the other three classes.

Table 3.7 Degree of uncertainty in classification (Mean posterior probability in each latent class)

Class	N	mean	SD	Min	max
1 Negative symptom group	189	0.96	0.09	0.55	0.9997
2 Positive symptom group	135	0.92	0.14	0.50	0.9996
3 Mixed symptom group	223	0.96	0.09	0.53	0.9988
4 Low symptom group	183	0.98	0.05	0.68	0.9999

Entropy E=0.92, implying that 92% of subjects are correctly classified (good latent class separation)

3.8 Comparing the latent classes in terms of demographic and other variables

Latent classes

The latent classes were further examined by looking at the demographic and other variables of participants belonging to each group. In order to assess this we utilized a class certainty of 0.55 as cut-off point. 37.2% belonged to the mixed symptom group, 22.5% to the no/low symptom group, 20.6% to the negative symptom group and 19.7% to the positive symptom group.

Table 3.8 Comparison of selected variables across groups

	1. Negative symptom group	2. Positive symptom group	3. Mixed symptom group	4. Low/no symptom group
Gender				
Male	88.90%	72.80%	87.70%	78.00%
Female	11.10%	27.20%	12.30%	22.00%
Marital status				
Married/Widowed	11.20%	12.60%	10.20%	21.20%
Single	82.40%	78.60%	80.50%	75.40%
Divorced/Separated	5.50%	6.80%	8.80%	3.30%
Age				
Age at interview	36.53 (SD11.03; Range 18-70)	33.60 (SD 10.77; Range 15-60)	33.48 (SD 10.73; Range 17-66)	37.47 (SD 10.97; Range 18-64)

Age at onset	23.74 (SD 7.28; Range 13-51)	23.03 (SD 6.37; Range 10-52)	23.05 (SD 6.63; Range 12-49)	23.76 (SD 7.55; Range 11-55)
Duration of illness	12.79 (SD9.31; Range 0-50)*	10.42 (SD 9.97; Range 0-50)*	10.23 (SD 8.98; Range 0-47)*	13.58 (SD 9.96; Range 0-41)*

* 0 indicates duration of illness less than 1 month

Comparing the latent classes in terms of the above variables revealed that group 4 (37.5 yrs) was significantly older at the time of interview than group 1(36.5) (p=0.002; DF 242.63) and group 2 (33.6 yrs) (p=0.021; DF 215.78). Group 3 was significantly younger than group 1 (p=0.009; DF 215.938). Group 4 had a significantly longer duration of illness than group 3 (p=0.003; DF 226.72) and group 2 (0.022; DF 216.275). The duration of illness of group 3 was significantly shorter than that of group 1 (p=0.020; DF 212.917). In all four latent classes the proportion of men was higher than that of women, the largest differences occurring in groups 1 and 3 (88.9% and 87.7% men, respectively) and the smallest in groups 2 and 4 (72.8% and 78.0% men, respectively). The difference in gender distribution across the groups was significant (p=0.002). Marital status did not differ significantly across groups.

There were no significant differences among the groups in terms of number of episodes of psychosis or number of hospitalizations.

Table 3.9 Comparing the prevalence of clinical variables across the latent classes

	1. Negative symptom group	2. Positive symptom group	3. Mixed symptom group	4. Low/no symptom group	P
Number of hospitalizations (mean; SD)	3.21 (SD 2.88)	2.97 (SD2.65)	3.53 (SD 3.56)	3.39 (SD 2.80)	NS
Number of episodes (mean; SD)	3.26 (SD 2.48)	3.30 (SD 2.57)	3.50 (SD3.12)	3.74 (SD 2.51)	NS
Lifetime presence of symptoms (%)					
Persecutory delusions	91.43	89.11	90.63	89.66	NS
Delusions of sin	25.47	6.86	11.64	10.34	p<0.001
Grandiose delusions	40.57	49.51	43.16	42.24	NS
Religious delusions	34.91	39.60	24.87	27.59	p=0.042
Delusions of reference	50.94	62.00	52.36	60.34	NS
Delusions of control	11.32	26.73	18.52	20.69	p= 0.042
Delusions of	31.13	33.66	27.66	36.75	NS

Page | 160

mind reading					
Delusions of thought broadcasting	25.47	24.75	22.22	23.68	NS
Delusions of thought insertion	18.87	23.76	15.34	17.95	NS
Delusions of thought withdrawal	16.03	16.83	13.76	15.38	NS
Bizarre delusions	2.83	16.83	12.23	8.62	p= 0.007
Somatic delusions	19.81	23.76	24.21	15.52	NS
Nihilistic delusions	1.89	0.99	2.12	1.72	NS
Erotomanic delusions	5.66	7.00	4.28	6.90	NS
Delusions of jealousy	9.43	5.94	4.23	6.03	NS
Auditory hallucinations	95.33	100.00	100.00	93.04	p<0.001
Visual hallucinations	46.67	64.71	51.31	48.28	p= 0.037
Olfactory	32.38	22.55	22.75	26.09	NS

hallucinations					
Tactile hallucinations	29.52	18.63	24.47	19.13	NS
Gustatory hallucinations	28.57	10.78	11.64	11.30	p<0.001
Behavioral disturbance	90.74	85.44	92.75	88.14	NS
Aggression	76.85	69.90	76.68	71.19	NS
Bizarre behavior	63.89	47.57	61.34	43.22	p= 0.001
Thought disorder	66.98	62.75	81.25	50.00	p< 0.001
Catatonic behavior	19.40	18.40	27.70	17.80	NS
Mood disorder	8.41	33.00	20.10	22.03	p<0.001
Anxiety disorder	0.00	9.70	3.08	10.17	p< 0.001
Suicide attempt	6.48	19.42	12.31	15.20	NS
Stressor prior to onset of illness	25.93	29.13	26.15	25.42	NS

Table 3.10: Hypothesis test summary for age at onset of illness, age at interview and duration of illness

<p align="center">Hypothesis Test Summary</p>

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of age at interview is the same across categories of the Group	Independent Samples Kruskal- Wallis Test	.004	Reject the null hypothesis
2	The distribution of age of onset is the same across categories of the Group	Independent Samples Kruskal- Wallis Test	.820	Retain the null hypothesis
3	The distribution of duration of illness is the same across categories of the Group	Independent Samples Kruskal- Wallis Test	.002	Reject the null hypothesis

Asymptotic significances are displayed. The significance level is .05
Age_at_i = age at interview
Onset_ag = age at onset
Dur_ill = Duration of illness

The prevalence of developmental abnormalities/events (as discussed earlier) was significantly higher in the positive symptom group (12.37%) versus the mixed symptom group (4.9%) (p=0.048). However, only a small number of participants were able to provide a reliable history regarding these occurrences.

Comparing latent classes in terms of substance use

The use or abuse of substances showed more use or abuse of cannabis in the positive and mixed symptom groups and more alcohol abuse in the mixed symptom group.

Table 3.11 Substance use or abuse across latent classes

	Negative symptom group	Positive symptom group	Mixed symptom group	Low/no symptom group
Alcohol abuse	11 (10.58%)	24 (23.76%)	53 (28.19%)*	24 (20.69%)
Alcohol dependence	0 (0.00%)	1 (0.99%)	3 (1.55%)	4 (3.42%)
Cannabis use or abuse	46 (42.59%)	54 (52.43%)	101 (51.79%)	49 (41.52%)
Tobacco smoking	79 (73.15%)	73 (70.87%)	150 (76.92%)	88 (74.58%)

*p=0.002

LCA with covariates

The above analysis comparing the latent classes did not take into account the uncertainty in the classification of the individuals to the groups. The latent class analysis can incorporate covariates in the analysis using a multinomial logistic regression approach, thus taking into account the uncertainty in the classification.

First we need to check that the item response profiles are consistent for the covariate groups. The G^2 difference test for item responses was not significant for gender, a history of cannabis use or abuse (lifetime, DIGS) or the presence of an affected sibling, implying that the item-response probabilities are not different across groups (Table 3.12).

Table 3.12: Fit table for 8-variable model (Assume equal item probabilities) with covariates (sibpair status, gender, cannabis use or abuse)

Number of variables	Groups	Classes	Restriction on gamma parameters	G^2	df	AIC	BIC	l	Chi diff df=3	p
8	2	4	Sibpair unequal*	280.7	473	356.7	531.2	-2639.7		
8	2	4	Sibpair equal	345.4	476	415.4	576.1	-2672.1	64.7	0.0001
8	2	4	Gender unequal*	262.8	473	338.8	513.3	-2659.8		
8	2	4	Gender equal	287.8	476	357.8	518.6	-2672.3	25	0.0001
8	2	4	Cannabis unequal*	268.5	473	344.5	519.0	-2654.2		
8	2	4	Cannabis equal	294.6	476	364.6	525.4	-2667.3	26	0.0001
8	1	4	Duration of illness					-2613.11		

G^2: Likelihood-ratio statistic, AIC: Akaike information criteria, BIC: Bayesian information criteria, l: Log-likelihood, Chi diff Likelihood-ratio difference test

The G^2 difference for gammas (i.e. latent class membership probabilities) for the gender, cannabis and sibpair groupings are significant (p=0.0001; table 3.12), implying that the gammas are different across groups (see below). Table 3.13 shows the class membership probabilities based on gender and sibpair status.

Table 3.13 Class membership probabilities

	Class membership probabilities				
Model		Negative symptom group	Positive symptom group	Mixed symptom group	Low symptom group
Baseline		25.6 (1.7)	18.1 (1.8)	31.2 (2.1)	25.1 (1.7)

model					
Groups	Gender Female	21.0 (3.8)	28.3 (4.4)	22.9 (4.1)	27.8 (4.3)
	Gender Male	26.5 (1.9)	15.3 (1.9)	33.6 (2.3)	24.6 (1.9)
	Sibpair	39.6 (3.5)	11.7 (2.4)	17.1 (2.7)	31.6 (3.3)
	Non-sibpair	19.5 (1.9)	20.6 (2.3)	37.3 (2.6)	22.6 (1.9)

Duration of illness as covariate

LCA with covariates examine predictors of latent class membership and prediction is done using a (multinomial) logistic regression approach. The logistic regression analysis produces an estimate of the effect for each latent class in comparison to the reference latent class. Illness duration was categorized into 0-6 years and 7-50 (maximum duration on dataset) to explore the impact of long duration of illness on latent class membership.

Table 3.14: Duration of illness categories; 0-6 years and 7-50 years

Duration in years	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Missing data	16	2.17	16	2.17
7-50	493	66.89	509	69.06
0- 6	228	30.94	737	100.00

The prevalence of each of the classes based on the duration of illness is illustrated below.

Table 3.15: Duration of illness as covariate. Prevalence of classes

Duration of illness (years)		Negative symptom group	Positive symptom group	Mixed symptom group	Low symptom group	Total
0- 6	Frequency	34	57	90	47	228
	Row %	14.91	25	39.47	20.61	
7-50	Frequency	152	74	126	136	488
	Row %	31.15	15.16	25.82	27.87	

For the logistic regression of the latent class variable on duration of illness, the Beta estimates (standard errors) for the four classes (using class four as reference) were 0.1093 (0.1251), -0.6420 (0.1626), -0.0557 (0.1303) and the parameter estimates were 0.2662, 0.2569 and 0.2254 respectively. Calculating Odds Ratios showed that parameter estimates (intercept) [95% confidence interval, lower and upper boundaries] relative to class 4 (low symptom group) were 1.1155 (0.8730, 1.4255) for negative symptom group, 0.5263 (0.3826, 0.7238) for positive symptom group, 0.9458 (0.7326, 1.2211) for mixed symptom group compared to low symptom group. For duration of illness the lower and upper boundaries were 0.3895 and 1.1058 for class one (negative symptom group), 1.4317 and 3.9193 for class 2 (positive symptom group) and 1.2834 and 3.1050 for class 3 (mixed symptom group). The covariate duration of illness contributed significantly to the prediction of latent class membership ($p<0.0001$; df=3). The 95% CI for the odds ratios for class 2 (positive symptom group) and class 3 (mixed symptom group) exclude 1. This tells us that the odds of belonging to the positive symptom group as well as the mixed symptom group are higher for the duration 0-6 years group than for the 7+ years duration group (2.2 and 2.1 times greater respectively).

The null hypothesis $H0:\beta1=0$, where beta is the vector of regression coefficients, can be tested by comparing the fit of two models. Model 1 is the baseline model without the covariate and model 2

is the corresponding model that includes the covariate (duration of illness). Then $-2(l_1-l_2)$ is in theory distributed as a χ^2 with df=p_2-p_1. In terms of the intercept, 1.1155 is the odds of membership in the negative symptom group in relation to the low symptom group (the reference group) for duration 7-50 years. In terms of the regression coefficient, the odds ratio 2.22 indicate that 0-6 years duration of illness had a higher probability than 7-50 years duration of illness of membership in the positive symptom (and mixed symptom) latent class, i.e. odds for 0-6 years is 2.2 times the odds for 7-50 years duration of illness.

Latent class analysis with groups

Group analysis was performed with cannabis use or abuse, sib pair status and gender as the groups of interest.

Table 3.16:   Fit statistics for test of measurement invariance: cannabis use or abuse

	G	df	p	AIC	BIC	l
1.Item-response probabilities vary across cannabis groups	62.7	73	0.8	170.7	418.7	-1989.6
2.Item response probabilities equal across cannabis groups/unequal gammas	93.3	97	0.587	153.3	291.1	-2004.9
Diff (1-2)	30.6	24	0.166			
3. Gammas equal across cannabis groups	120	100	0.084	174	298	-2018.3
Diff (2-3)	26.7	3	0.000			

G^2 difference for item responses test is not significant implying that the item-response probabilities are not different across groups (Chi-square difference=30.6, df=24, p=0.167). However, the G^2 difference tests for gammas are significant, implying that the gammas are different across groups (Chi-square difference=26.7, df=3, p=0.0001) (Table 3.17).

Table 3.17: Parameter Estimates for groups: cannabis use or abuse

Class membership probabilities: Gamma estimates (standard errors)				
Class	1 (Negative symptom class)	2 (Positive symptom class)	3 (Mixed symptom class)	4 (Low or no symptom class)
Group 1 Cannabis use or abuse	0.27	0.29	0.17	0.27
	(0.0213)	(0.0220)	(0.0203)	(0.0230)
Group 2 No use or abuse	0.2041	0.1690	0.2353	0.3916
	(0.0258)	(0.0242)	(0.0301)	(0.0336)

Fit statistics for groups based on gender was calculated (Table 3.18).

Table 3.18: Fit statistics for test of measurement invariance: gender

	G^2	df	p	AIC	BIC	L
Item-response probabilities vary across gender groups	66.9	73	0.679	174.9	422.9	-2005.6
Item-response probabilities equal across gender groups	77.5	97	0.928	137.5	275.3	-2010.9
Difference	10.6	24	0.992			

G^2 difference test for item-response probabilities is not significant implying that the item-response probabilities are not different across groups (Chi-square difference=10.6, df=24, p=0.992). However, the G^2 difference test for gammas is significant, implying that the gammas are different across gender groups (Chi-square difference= 29.3, df=3, p=0.0001).

Table 3.19: Parameter Estimates for groups: gender

Class membership probabilities: Gamma estimates (standard errors)				
Class	1 (Negative symptom class)	2 (Positive symptom class)	3 (Mixed symptom class)	4 (Low or no symptom class)
Group 1	0.2110	0.3101	0.2278	0.2510
Male	(0.0366)	(0.0437)	(0.0399)	(0.0392)
Group 2	0.2567	0.1685	0.3320	0.2428
Female	(0.0188)	(0.0185)	(0.0220)	(0.0183)

Finally the fit statistics were calculated for the groups based on sibpair status (Table 3.20).

Table 3.20: Fit statistics for test of measurement invariance: sibpair status

	G^2	df	P	AIC	BIC	l
Item-response probabilities & gammas vary across sibpair groups	71.9	73	0.8	179.9	427.9	-1969.4
Item response probabilities equal across sibpair groups gammas vary*	113.1	97	0.587	173.1	310.9	-1989.9
Difference	30.6	24	0.166			
Item response probabilities equal & gammas equal across sibpair groups	180.5	100	0.000	234.5	358.5	-2023.6
Difference	67.4	3	0.000			

G^2 difference for item responses test is not significant implying that the item-response probabilities are not different across groups (Chi-square difference=30.6, df=24, p=0.167). However, the G^2 difference test for gammas is significant, implying that the gammas are different across groups (Chi-square difference=67.4, df=3, p=0.0001).

Table 3.21: Parameter Estimates for groups based on sibpair status

Class membership probabilities: Gamma estimates (standard errors)				
Class	1 (Negative symptom class)	2 (Positive symptom class)	3 (Mixed symptom class)	4 (Low or no symptom class)
Group 1 Sibpair	0.3941	0.1272	0.1710	0.3077
	(0.0349)	(0.0245)	(0.0274)	(0.0330)
Group 2 Non-sibpair	0.1888	0.2212	0.3699	0.2200
	(0.0181)	(0.0221)	(0.0246)	(0.0190)

The question of whether latent class prevalences are identical across groups was addressed by comparing the fit of Model 1, in which the item-response probabilities are constrained to be equal across groups and the latent class prevalences are free to vary, against the fit of Model 2, in which both the item-response probabilities and the latent class prevalences are constrained to be equal across groups. The hypothesis that Model 2 fits as well as Model 1 (i.e., that the class prevalences are not different across groups) was tested via the difference G^2.

In summary, the patients' gender, use of cannabis and sibpair status did not significantly impact on the item response probabilities. However, from the results it seems that male gender increases the probability of belonging to class 2 (positive symptom class), while the absence of a cannabis use

or abuse history increases the probability of belonging to class 4 (Few or no symptom class).

Affected sibpair status increased the probability of belonging to class 1 (negative symptom class).

3.9 Factor mixture modeling

We used the item response probabilities to interpret the classes and to give the classes descriptive labels (see profile of response probabilities tables below):

Class 1("subgroup 1") -   eye contact, affective non-responsiveness, spontaneous

                                        movement, grooming

Class 2("subgroup 2") -   auditory hallucinations, global hallucination score, global

                                        delusion score, commenting voices

Class 3 ("subgroup1+ 2") -   negative and positive symptoms

Class 4 ("subgroup 3") -   is a no/low symptom group

From the 10 selected symptoms (based on previous results) above, the 5 most common response patterns were identified.

Table 3.22: Five most common response patterns for selected symptoms

Items from SAPS and SANS	%	RP 1	RP 2	RP 3	RP 4	RP5
Decreased movement	56.3	Yes	No	Yes	Yes	Yes
Poor eye contact	54.1	Yes	No	Yes	Yes	Yes
Affective non responsiveness	58.5	Yes	No	Yes	Yes	Yes
Global alogia	60.6	Yes	No	Yes	Yes	Yes
Grooming	59.0	Yes	No	No	Yes	Yes
Auditory hallucinations	49.5	No	No	No	Yes	Yes
Global hallucinations	50.8	No	No	No	Yes	Yes
Conversing voices	31.8	No	No	No	Yes	Yes
Global delusions	53.1	No	No	No	Yes	Yes
Global bizarre behaviour	39.1	No	No	No	Yes	No
Frequency		64	61	34	33	29

RP = response pattern

In order to benefit optimally from the available data, the latent class analysis with four classes was repeated incorporating the items on the original numerical scale (0 to 5).

Table 3.23 Mean Item Scores obtained for the SANS and SAPS

Items scale 0 to 5	Mean	SD
Auditory hallucinations	1.6	1.7
Global hallucinations	1.6	1.7
Global delusions	1.6	1.6
Global bizarre behavior	1.2	1.5
Decreased spontaneous movement	1.5	1.4
Poor eye contact	1.4	1.3
Affective nonresponsiveness	1.6	1.4
Global alogia	1.8	1.5
Grooming & hygiene	1.6	1.4

The solution obtained using the numerically scaled items did not materially differ from the model obtained using dichotomized items (symptom present versus absent) used in the prior latent class analysis (see Fig 3.6 below):

Fig 3.6: Allocation to each of the 4 classes according to dichotomised SANS/SAPS items
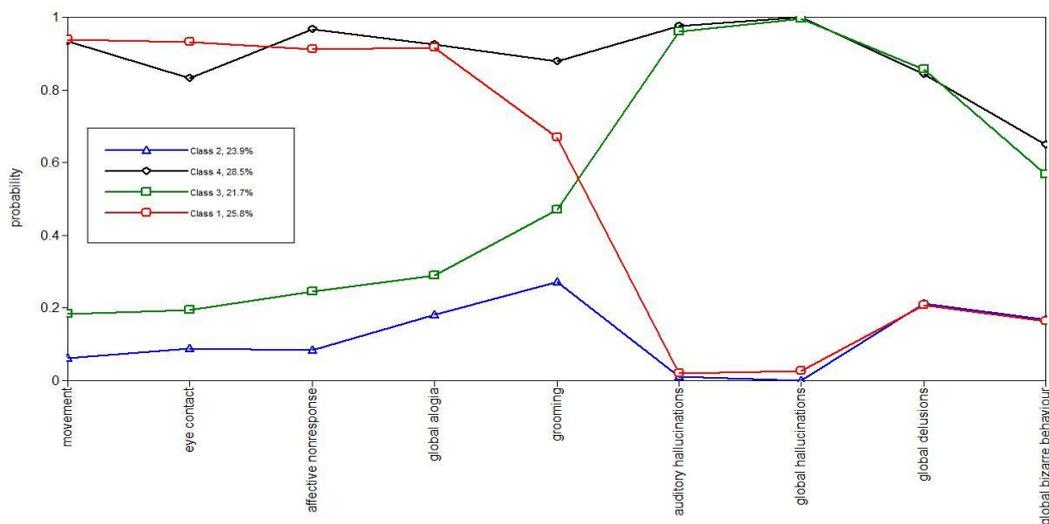
Fig 3.7: Allocation to each of the 4 classes according to numerically scaled SANS/SAPS item scores.
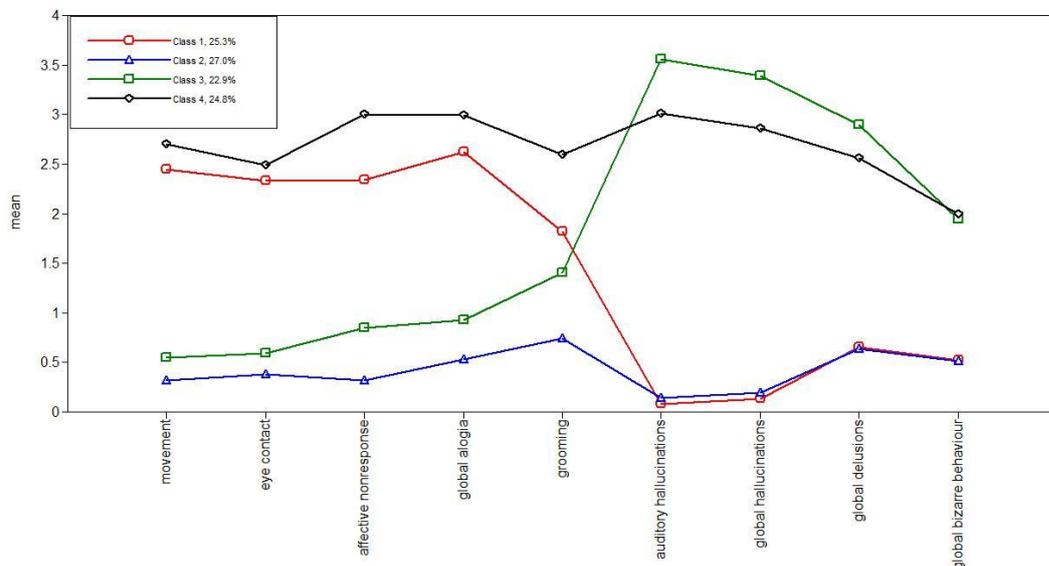
The sizes of the classes remained constant between the two LCA models. So, although the number of classes and the size of each did not differ significantly between the dichotomized and mean score models, there was a clearer separation in the "positive symptom" class and the "mixed symptom" class.

Once the LCA model was established, a factor model was fit on top of the LCA model, allowing individual to have different factor values within the classes. The table shows the solution for the LCA model, the factor model and the factor mixture model.

Table 3.24: Factor Mixture Model

Model		Log-L	Par	BIC	aBIC	LMR
LCA	3c	-9992	38	20232	20111	
LCA	4c	-9517	48	19348	19195	
FA	1 F	-10371	27	20918	20833	
FA	2F	-9471	28	19125	19037	
FMM-3	3c, 1 F	-9090	47	18487	18338	
FMM-3	**4c, 1 F**	**-8754**	**57**	**17882**	**17701**	**0.04**
FMM-3	3c, 2 F	-8965	48	18243	18091	
FMM-3	**4c, 2 F**	**-8745**	**58**	**17870**	**17686**	**0.03**

LMR Lo-Mendell-Ruben test

The best fitting FA model (2 Factors) and the best fitting LCA (4 classes) served as the starting point for the factor mixture model building process (FMM models in the table). This process revealed that there was no material superiority for the two factor model (FMM-3; 4c, 2F) above a one factor model (FMM-3; 4c, 1F) (Table 3.24), but both are superior to their 3-class counterparts (LMR; $p=0.03$ and $p=0.04$). Both models will therefore be investigated before choosing the best

model (Table 3.24). The fit statistics for the FMM-3 4 class, one factor model was better than that of the 2 factor model and the model is illustrated below.

Figure 3.8: FMM-3 4c, 1F model
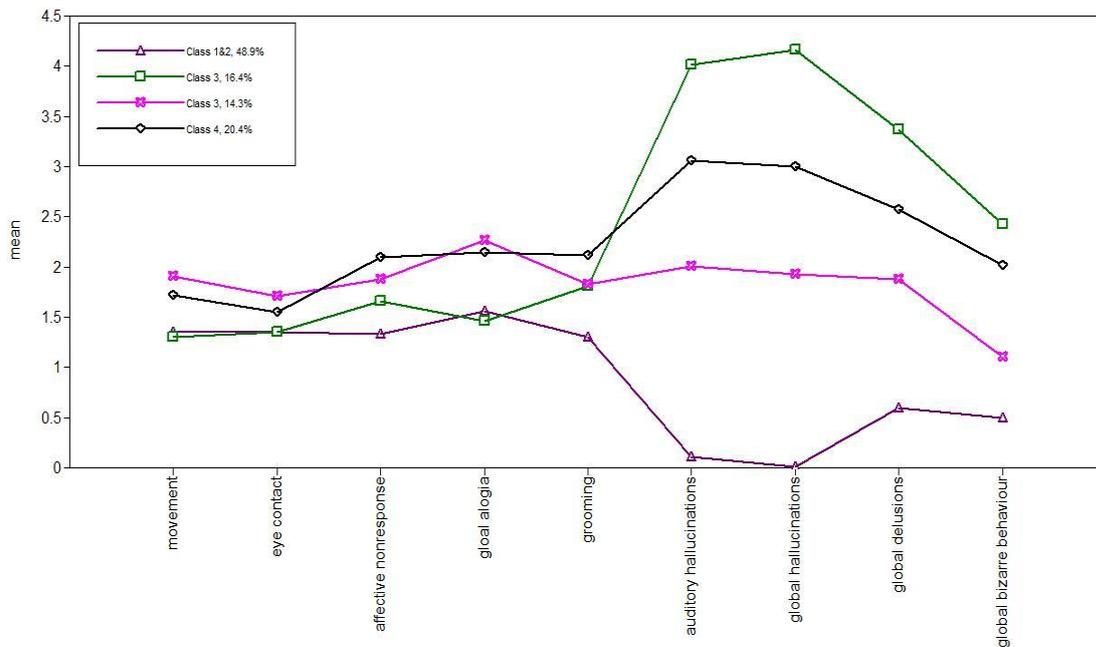
This solution indicates that four classes can be distinguished that could be labelled based on the positive symptoms profile alone:

Class 1(purple): Low symptom group (48.9%). As the name indicates this class was associated with the lowest positive symptom profile (average below 1). This class basically combined the "no/low symptom" class and the "negative symptom" class of the previous LCA model, indicating that the negative symptoms do not determine the classes for the FMM model.

Class 2 (green): High positive symptom group (16.4%): This class showed the highest mean profile (above 3) in auditory hallucinations, global hallucinations, global delusions and bizarre behaviour.

Page | 176

Class 3 (pink): Mixed symptom group (14.3%): This class consist of individuals with an average profile between 1 and 2 for positive symptoms with the second lowest bizarre behaviour mean score across the groups.

Class 4 (black): Medium high positive symptom group (20.4%): This group corresponds to the fourth class in the FMM-3 4 class 2 factor model and represents a group with an average profile between 2 and 3.

The striking finding of this solution is that negative symptoms did not predict group placing. Upon investigation of the factor scores of each of these classes (Figure 3.9 below), it becomes clear that class 1 (purple) factor scores shows a bimodal distribution with a large variation. It seems that class one is responsible for the largest variation in factor scores.

The diagram on the next page shows the FMM3 4c, 1F Factor scores:

Figure 3.9: FMM-3 4c, 1F factor scores
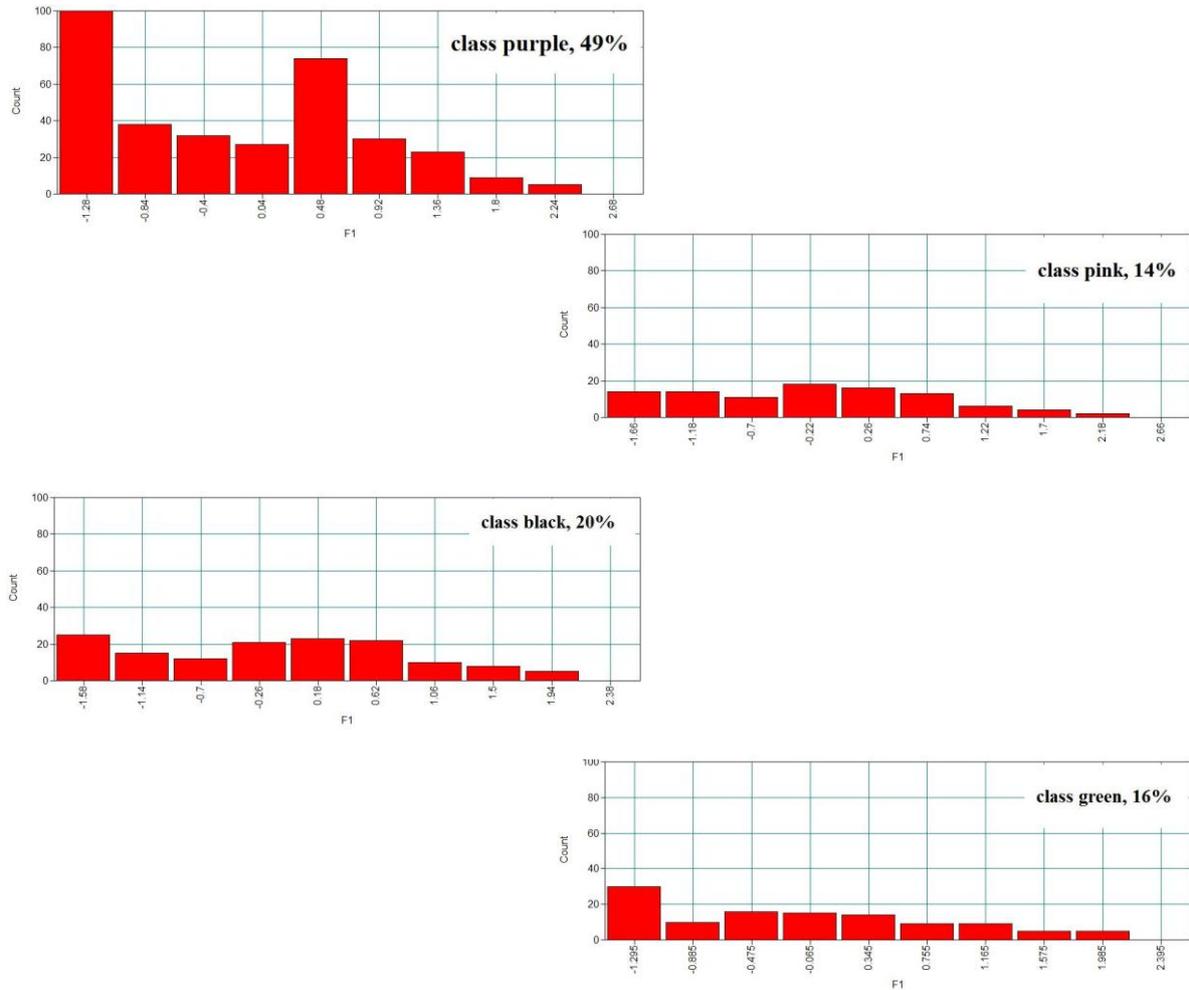
Since the low positive symptom group had a bimodal distribution as seen in the distribution plot (see fig 3.9), the FMM model with 2 factors was also considered. The scatterplot for the two factors fitting this model is shown below (fig 3.10). It is clear that, compared to the first factor, the second factor has a small negligible variance, whereas most of the variance is explained by the 1st factor.

Fig 3.10: Grammatical representation of FMM-3 (4 classes, 2 factors) factor scores
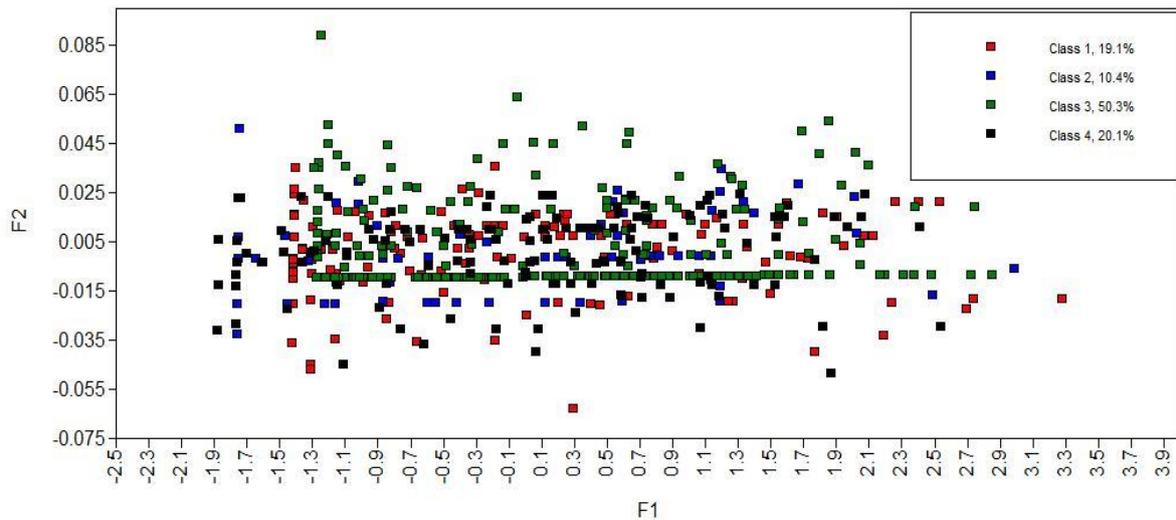
The factor can thus be interpreted as a quantitative score, formed as a weighted sum of the number of negative symptoms present at assessment, and this score can be used as a measure of severity of disease, within each class. Group status is only predicted by the presence of the positive symptoms in the model.

In summary the results show that factor mixture modelling uncovered a heterogeneous latent variable structure that fits the data well with the latent classes capturing distinct behaviours and factors capturing severity variations.

Chapter 4 Discussion

Content

CHAPTER 4 DISCUSSION AND CONCLUSIONS

4.1 Demographic data

The Xhosa schizophrenia sample selected for inclusion in this study was similar to that of the previous study (Niehaus et al. 2008) in terms of demographic and clinical variables, except for a higher prevalence of substance use or abuse, and a lower prevalence of comorbid anxiety disorders. In both studies the ratio of men to women was 4:1. It is not a unique finding that male subjects outnumber female subjects in genetic studies, the reported proportion of men ranging from just over 50% (Huang et al. 2011) to 58% in an unspecified schizophrenia sample (Peralta et al. 2003) to 70% in the Kraepelin schizophrenia group of Derks et al. (2012). The reason for the overrepresentation of male participants in genetic studies of schizophrenia remains an enigma. Possible explanations include a recruitment bias linked to a lower tolerance for symptomatology in male family members and a smaller likelihood of female schizophrenia patients being willing to participate in genetic studies. Future studies should focus on the underlying reason for this bias as it may well impact on treatment strategies.

The mean age at interview of the current study subjects falls in the same 36-40 year age range reported in our previous study, as well as by others (Peralta et al. 2003, McGrath 2004). Similarly, mean age of onset was in the 21-27 year range reported in our previous study (Niehaus (2008) and by Peralta et al. (2003). Our sample had a much lower HLOE than Peralta et al. (2003) (4 v. 9, respectively)]. Nearly 80% of participants in both our studies were single, similar to the data reported by Blanchard et al. (2005). Over two-thirds of both of our studies' subjects were in receipt of a disability grant. The low levels of formal education found in both of our studies corroborate the findings of Arnold et al. (2004) that non-European schizophrenic patients (in their case African-American) have lower levels of education.

The mean numbers of psychotic episodes were 2.5 and 3.2 for our previous and current study (Niehaus et al. 2008) representing an average of 2.6 and 3.1 hospitalizations, respectively, which seems to be consistent with the average of 3.4 hospitalization episodes reported by Peralta et al. (2003).

4.2 Diagnostic Assessment

Where two percentages are given, the following results refer to the 2008 and current schizophrenia studies of the author, respectively. Persecutory (88% and 89%), grandiose (55% and 46%) and reference (51% and 54%) delusions were common in both groups (Niehaus 2008), as were auditory (97%, 97%), visual (57%, 53%) and tactile (46%, 28%) hallucinations. Aggression (verbal and physical) was the most common ($> 70\%$) behavioral disturbance in both groups. Thought disorder occurred in approximately 60% of subjects in both samples. Seventy eight percent and eighty eight percent of the subjects in both studies had affective changes. Alogia (73% and 67%) and avolition and apathy (80% and 86%) occurred in both groups. It seems from this comparison that the two groups show similar symptom prevalence despite the fact that the present study was biased towards lower familial loading in terms of its recruitment strategy. Following work done by Chang et al. (2011), it is important to compare these findings with those of other ethnic groups. Chang et al. (2011) reported that in a group of 219 psychiatric inpatients (DSM-IV diagnosis of schizophrenia or schizoaffective disorder) consisting of 91 African-American, 32 Latino, 25 Euro-American, and 71 Chinese-American patients, the last-mentioned group presented with significantly fewer psychotic symptoms than both the African-American and Latino patients. The Euro-Americans fell between these groups. This finding does not stand in isolation as Barrio et al. (2003) also support the presence of more positive symptoms in African American patients than Euro-American patients. The African-American patients also showed more negative symptoms than non-African Americans as reported by Mark et al. (2003). These findings are

consistent with those of Arnold et al. (2003), Strakowski et al. (1996a) and Strakowski et al. (1996b), who found that African-American patients are more likely to be diagnosed with more severe psychotic symptoms (including first-rank symptoms) than Euro-American patients. The major disadvantage in trying to interpret these studies is that detailed prevalence/incidence rates of specific symptoms are not provided in the papers. Weissman et al. (2000), in comparing Latino and Euro-American patients, reported that Mexican Americans experienced fewer paranoid delusions than Anglo-Americans, and the latter group reported a higher rate of hearing critical voices. No difference in the prevalence of religious delusions was noted. Anglo-Americans had significantly greater levels of self-neglect and blunted affect than Mexican Americans. A number of older studies (sample sizes ranging from 45 - 133 subjects) reported highly variable rates of hallucinations in schizophrenia patients, with 66 - 84% experiencing auditory, 12-46% visual, 4-54% tactile/somatic and 6-38% olfactory hallucinations (Small et al. 1996, Andreasen 1987, Mueser et al. 1990, Bracha et al. 1989, Lewandowski et al. 2009, Goodwin et al. 1971). The interpretation of this data is subject to use of English versions of the assessment tools that was translated by an interpreter. Although a consistent method was followed, the use of oral translation is problematic in terms of pure linguistics (Smith et al. 2013) and in terms of the structure (from a dyadic to a triadic relationship) of the interaction (Swartz et al. 1998). This could have impacted on the quality of the data and validation studies are needed for a Xhosa version of all the assessment tools.

4.3 Treatment and substance use data

Treatment of schizophrenia was similar across our two groups, with typical antipsychotics dominating. Fewer than 5% were prescribed clozapine. A smaller proportion of subjects in our

current study were assessed as having a history of substance abuse or dependence than in our previous study (16% v. 35% for cannabis and 7% v. 16% for alcohol abuse / dependence).

Nicotine use was present in more than 80% of our study population. This is not surprising, as nicotine use is not only very common in schizophrenic subjects (estimated to be two to fourfold the rate in the general population) (Kumari and Postma 2005), but their daily consumption and nicotine content of the products also seem to be higher, this pattern remaining true across cultural boundaries (de Leon et al. 1995, de Leon et al. 2002a; Glassman 1993, Lohr and Flynn 1992, Uck et al. 2004, Olincy et al. 1997a, Strand and Nyback 2005). Indeed, schizophrenia patients seem to be using tobacco at a higher rate than found in bipolar spectrum disorders (de Leon et al. 2002b, Uck et al. 2004). Several reasons have been postulated for these findings, including self-medication of side effects and negative symptoms, improvement of cognitive deficits and, possibly, increased genetic vulnerability (Kumari and Postma 2005).

4.4 Mood and Anxiety disorders

Approximately one sixth of subjects had a history suggestive of mood disorders, this rate being at the lower end of the spectrum of estimates reported in follow up studies, which ranged from 25% (McGlashan and Carpenter 1976, Siris 1991) to 60% (Martin et al. 1985). Kraepelin's traditional framework views schizophrenia as a 'non-affective' psychosis (Upthegrove et al. 2010). However, Bleuler considered depressive symptoms as being characteristic of schizophrenia and viewed hallucinations and delusions as secondary manifestations of schizophrenia (Bleuler 1924). Bleuler's ideas regarding the importance of depressive symptoms in schizophrenia have wide support, though they are not generally viewed as a core symptom (Zisook et al. 1999, Birchwood et al 2000, , Siris 2000, An der Heiden et al 2005). Depressive symptoms can appear across the illness course (Koreen et al. 1993, Hafner et al 1999, Schultze-Lutter et al. 2007, Birchwood et al

2000) and often follow the course of positive psychotic symptoms (Koreen et al 1993, Lancon et al 2001, Oosthuizen et al 2006). However, given the make-up of our sample (a mix of acute and chronic inpatients and outpatients with a selection bias towards "non-affective" psychosis), it seems likely that our lower rates of depression should not only be ascribed to a recruitment bias but can also suggest that depression is less common between acute psychotic episodes than in the earlier stages of illness (Lançon .et al. 2001). Co-occurring depression also increases the risk of suicide attempts (Fenton 2000) and one cannot estimate the impact of mortality in the depressed subgroup on recruitment.

A history suggestive of anxiety disorders was obtained in 11% and 5% of our previous and current study, respectively. A number of clinical studies have reported on the presence of comorbid anxiety disorders in schizophrenia. However, the criteria used for selecting subjects for inclusion into these studies were not uniform (e.g., some used inpatients and others outpatients), and samples were generally not large enough to provide precise estimates (Argyle 1990, Goodwin 2003). Results are therefore highly variable. Co-occuring obsessive compulsive disorder was reported to range from 0% (Tibbo 2003) to 35% (Bayle 2001); panic disorder ranged from 3.3% (Tibbo 2003) to 43% (Labbate 1999); social anxiety disorder ranged from 8.2% (Goodwin 2003) to 36% (Pallanti 2004), while posttraumatic stress disorder was reported by Pallanti (2004) to be only 1.3%. Although anxiety symptoms in schizophrenic patients do occur relatively commonly, the concept is not well understood in terms of the relationship with psychosis.

Comorbidity thus remains a complex construct within the study of schizophrenia. Given the current hierarchical approach of our diagnostic systems, it seems logical that the rate of comorbidities would be low. However, if one neglects psychopathological refinement in the diagnostic system, the comorbidity rate will tend to be inflated. This is compounded by the lack of

a gold standard diagnostic method across research studies. It should thus be expected that the results reported in the literature will be heterogeneous and even contradictory.

4.5 Latent class analysis

The results of the LCA of the SANS / SAPS data suggest that homogeneous classes of Xhosa schizophrenic patients exist. These subtypes may reflect distinct underlying biological, genetic or environmental mechanisms. Latent class analysis identified a model that supported a 4-class structure. Class 1 showed predominantly negative symptoms, class 2 had high rates of positive symptoms, class 3 (comprising nearly a third of the cases) was a mixed symptom group and members of class 4 showed no or few symptoms. The results showed both good homogeneity within a given class (item response probabilities generally being close to 0 or 1) and separation (item response probabilities varying across classes). Of the items of interest extracted from the concordance study by Niehaus et al. (2005), namely eye contact, auditory hallucinations, global hallucinations and delusions of control, only the last mentioned failed to be included as a separate variable in the 8-variable best fit model, while grooming appeared as a separate variable in the model.

4.6 Class 1: Negative Symptoms

In line with previous findings, the negative symptoms cohered into a distinct class (Blanchard et al. 2005). Messinger et al. (2011) define "Negative symptoms" as "those that refer to the loss or diminution of normal functions, such as expressiveness and motivation". Indeed, negative symptoms have come under the spotlight with the development of the DSM-V diagnostic criteria, and especially blunted affect, alogia, asociality, avolition and anhedonia have become the focus for renewed research as proposed by the National Institute of Mental Health (NIMH) Measurement

Page | 187

and Treatment Research to Improve Cognition in Schizophrenia (MATRICS) consensus panel (Kirkpatrick et al. 2006). The panel suggested that negative symptoms might best be viewed as a two-factor model consisting of factor 1, blunted affect-poverty of speech, and factor 2, anhedonia– asociality–avolition, though some concerns have been expressed about the applicability of anhedonia (Foussias & Remington 2008). The inclusion of anhedonia has been criticized, as laboratory evocative studies using standard stimuli for all participants showed that patients with schizophrenia, including those with primary negative symptoms, experienced levels of pleasure for positive or pleasant stimuli similar to those experienced by controls (Tremeau et al. 2009).

Blanchard et al. (2005) demonstrated that a discrete class of patients exists (28-36% base rate) who showed elevated negative symptom levels, were more likely to be men and had poorer social functioning. This group shares some of the characteristics of the proposed deficit syndrome samples (Kirkpatrick et al. 2000) that also had a higher proportion of males. Kirkpatrick et al. (1998) demonstrated an association between summer births and deficit syndrome of schizophrenia while Blanchard et al. (2005) found a modest increase in the disorganization score of the BPRS in the negative symptom group, but found this to be due to correlation with the SANS scores. This is similar to our initial report on the factor analysis (Niehaus et al. 2005) that showed no separation between the negative and disorganization domains. However, because of uncertainty as to which symptoms should form part of the construct of negative symptoms, measurement scales show little overlap and only flat affect is measured in all well-established scales (Foussias & Remington 2008). Factor analytic studies (mostly using the SANS) of negative symptoms generally support two domains, namely expressive deficits (flat affect, alogia) and avolition (amotivation, anhedonia, asociality) (Emsley et al. 2001, Keefe et al. 1992, Kelley et al. 1999, Malla et al. 2002, Minas,Klimidis et al. 1994, Mueser et al. 1994, Peralta & Cuesta 1999; Sayers et al 1996, Toomey et al. 1997). However the items inappropriate affect, poverty of content of speech, and inattentiveness form a separate factor and do not segregate with other negative symptoms of the

SANS. (Keefe et al. 1992, Malla et al. 2002, Mueser et al. 1994, Peralta & Cuesta 1999, Sayers et al. 1996).

In line with the findings on the SANS, factor analytic studies of the Schedule for the Deficit Syndrome (Kirkpatrick et al. 1989) also support two factors namely "diminished emotional expression" and "avolition" (Kimhy et al. 2006, Nakaya & Ohmori 2008). The Motor-Affective-Social Scale also showed two factors: a "motor-affective" factor (spontaneous and voluntary smiles, coverbal gestures) and a "motor-social" factor (motor retardation, personal hygiene, attendance at groups and activities, and verbal interaction). In this case alogia loaded on both factors (Tremeau et al. 2008).

The separation into two factors seem to hold true in very different study populations including patients who were on medication (Kimhy et al. 2006, Mueser et al. 1994, Sayers et al. 1996,; Tremeau et al. 2008) and who received no medication (Kelley et al. 1999), first-episode psychosis (Malla et al. 2002) and chronic (Keefe et al. 1992) patients. It seems to be true cross-culturally as well and similar results were found in patient from the United States (Keefe et al. 1992, Kelley et al. 1999, Kimhy et al. 2006, Mueser et al. 1994, Sayers et al. 1996, Tremeau et al. 2008), Canada (Malla et al. 2002), Spain (Peralta & Cuesta 1999), South Africa (Emsley et al. 2001), Australia (Minas et al. 1994), and Japan (Nakaya & Ohmori 2008).

Given these findings one can argue that negative symptoms are distinct heterogeneous symptoms of schizophrenia and may offer separate impairments in emotional expressiveness and volition. Expressive deficits can be defined as a specific deficit in social communication and shows some overlap with affective processes. Avolition can be defined as a deficit in functional outcomes. This will be very helpful for genetic research as it can possibly serve as a clinical endophenotype. Specifically expressiveness is useful from a clinical measurement perspective as it can be

observed during a clinical interview. Messinger et al. (2011) suggest that negative symptom (especially expressiveness) ratings only consider the following behavioral manifestations:

a. For expressiveness:

communicative facial expressions,

prosody,

hand coverbal gestures

language output.

b. For volition:

spontaneous motor activity,

grooming/hygiene,

work/recreation/leisure,

and social engagement.

The conceptualization of schizophrenia in the DSM-V supports the importance of these two domains (Kring et al. 2013, Heckers et al. 2013, APA 2013). Considering the findings from our studies thus far the negative symptoms of significant in the Xhosa samples were spontaneous movement, eye contact, affective non-responsivity and global alogia, with grooming as a possible variable of interest. Taking the above into consideration these symptoms/signs include mostly expressiveness symptoms (global alogia, eye contact, affective non-responsivity) with only grooming and spontaneous movement (although there is some overlap with hand co-verbal gestures) from the volition group.

4.7 Class 2 and 3: Positive and Mixed classes

Class 2 had high rates of positive symptoms and class 3 (comprising nearly a third of the cases) was a mixed symptom group. The positive symptoms contained in Class 2 and 3 include the following items: auditory hallucinations, global hallucination score, global delusion score and commenting voices. The presence of positive symptoms is considered an integral part of the understanding of schizophrenia. These two groups thus intuitively fit within our understanding of schizophrenia. Indeed, hallucinations, especially auditory and visual types are experienced by more than 70% of individuals with schizophrenia during the course of their illness (Baethge et al. 2005) and one study reported 70% of patients experienced auditory and visual hallucinations simultaneously (Gauntlett-Gilbert and Kuipers 2003).

Indeed, changes have occurred in our understanding of the importance of specific delusions in the diagnosis of schizophrenia. Schneiderian first-rank symptoms such as delusions of control, thought insertion, withdrawal and broadcasting were traditionally viewed as more important delusions. However, we now know that persecutory and reference delusions actually occur more commonly in schizophrenia (Mellor 1970, Tandon et al. 2009). The same issue arose in hallucinations where the classical Schneiderian first-rank hallucinations namely conversing and commenting voices actually occurs less frequently in schizophrenia than threatening or accusatory voices (Tandon et al. 2009). This finding is mirrored in our study with the most common delusions being life-time paranoid delusions (88.5%) and delusions of reference (53.5%).

Despite the high prevalence of hallucinations and delusions there appear to be a caveat in terms of the several aspects of these symptoms including 'What is the real life prevalence rate of these symptoms" and what is the course of fluctuations and the temporal relationships to these fluctuations? (Oorschot 2012). Maher et al. (1974) argued that delusions stem from the attempts to interpret the anomalous hallucinatory experiences and this early theory seems to be supported

Page | 191

by the strong association between the presence of hallucinations and delusions (Liddle and Barnes 1990, van Os et al. 2000) as well as the increased risk for psychosis if hallucinations are followed by delusional interpretation (Krabbendam and van Os 2005). However, the reverse mechanism might also be possible with delusions affecting the inner experiences and source-monitoring mechanism (Kapur 2003, van 't Wout et al. 2004).

Attempts in subtyping schizophrenia on the basis of symptomatology showed some heterogeneity with paranoid and Schneiderian delusions loading on separate factors in two studies (Liddle 1987, de Leon et al. 1993). Despite these earlier studies subsequent attempts mostly used subscale level scores and thus assumed that all items within the subscale contributed equally to the factor solution (Andreasen et al. 1995). Peralta and Cuesta (1998) questioned the wisdom of such an approach and showed that in their sample the best fit was for a more complex five factor model (albeit the fit not perfect). Their finding supported the earlier data in that Schneiderian delusions had the best variance value, while non-Schneiderian delusions seemed very heterogeneous in nature. Nevertheless hallucinations and delusions did still covary. If a two factor model was fitted, the psychosis factor (hallucinations and delusions) was associated with a later age of onset, while the disorganized factor was associated with a lower age at onset and increased illness severity. These finding support the need for an item level approach as per our study as subscale or total score approaches might obscure important findings for subtyping.

The DSM-V supports a dimensional approach to the diagnosis of schizophrenia and in essence class 2 and 3 might merely be representative of degrees of severity within the spectrum (APA 2013, Heckers et al. 2013). Our study does not include a detailed cognitive assessment or a qualitative assessment of mood or anxiety symptoms. When this study was designed the categorical approach to schizophrenia symptoms was followed. This turned out to be a major drawback as it makes it very difficult to ascertain whether class 2 and 3 might differ in terms of

these other domains (i.e. mood or cognitive impairment) and not merely differs in terms of the severity of psychotic symptoms. The meta-analysis of 104 studies investigating the relationship between neurocognition and disorganization performed by Ventura et al. (2010) may shed some light on this. Ventura et al. (2010) reports a weak relationship between reality distortion and neurocognition (r=−.04; p=.03) overall and in all six of the individual neurcognitive domains. Although these findings do not exclude interactions between reality distortion and neurocognition, they do support a dimensional view of positive symptoms distinguishing disorganization from reality distortion.

It can however be argued that positive symptoms vary to such a degree across the lifespan of a patient, that is almost impossible to use this a method for subtyping schizophrenia. Indeed, Schultz et al. (1996) already looked at this issue. They studied a group of 391 patients (ages 14 through 73) with schizophrenia, schizo-affective disorder and schizophrenifrom disorder in terms of psychotic, disorganised and negative symptoms. They found the effect of age to be significant in a negative direction for positive and disorganized symptoms, but no effect for formal thought disorder or negative symptoms.

Of specific interest is whether family history has any relationship with these positive symptoms. A review by Esterberg et al. (2010) shed some light on this. Using studies that reported on family history and age-at-onset (N=15 studies), age-at-onset and sex (N=12 studies), and/or positive (N=11 studies) and negative symptoms (N=12 studies) they demonstrated a small but significant impact on age-at-onset and negative symptoms, but not positive symptoms. Indeed, patients with a family history of schizophrenia had more negative symptoms. However, one should consider earlier individual studies i.e. McGuffin et al. (1991) and Basset et al. (1993) that suggests a liability/threshold model in which positive and mixed positive/negative forms of schizophrenia differed quantitatively along the same continuum of liability. McGuffin et al. (1991) goes further

to suggest that the positive symptom form may be less severe than the mixed form. Addington and Addington (1991), using the SAPS and SANS in 41 schizophrenia subjects in the acute phase and six months later, found no inverse relationship between positive and negative symptoms and the phase of illness. Again negative symptoms seem more constant across acute and follow-up period. This finding is interesting as our data showed a relatively large group with mixed positive and negative symptoms. In terms of subtyping the mixed group might be the more stable group in terms of symptom profile, but it is intriguing that the positive symptom group (group 2) exist as a separate entity. Future projects should include a longitudinal assessment of this group to see whether they remain truly positive or whether they revert to a low or no symptom group as the disease progress or at different time points within the illness course.

4.8 Class 4: No or few symptoms class

Although the first three classes intuitively fit to current theories on schizophrenia subtypes, the fourth class offers a less obvious explanation or fit to theory. Several possibilities exist to explain this group: firstly that recruitment or diagnostic bias in sib pairs could lead to the over diagnosis of schizophrenia in other sibs exhibiting brief psychotic symptoms linked to substances or organic causes; secondly this class could represent a residual or episodic form of schizophrenia with little inter-episodic symptomatology. However this finding is not unique to the Xhosa population, Kulhara and Chandiramani (1990) used the Andreasen and Olsen (1982) criteria to categorize 98 schizophrenia patients into positive, negative and mixed subtype at uptake and then re-assessed them 18-30 months later for category stability (SANS and SAPS scores reported).   Using the ICD-9 diagnostic criteria 25 of 79 patients were classified as positive subtype, 17 as negative and 37 as mixed. At follow-up 9 of the 17 negative subtype patients were re-assigned to the mixed group but none to the positive group, while the positive group did show patients that changed to the negative group. In addition, they also reported that a significant proportion of the mixed group (7 at uptake and 35 at follow-up) had neither positive nor negative symptoms sufficiently severe to

justify categorization into negative or positive subtypes. This group seems similar to class 4 participants in our analysis. Indeed, the principle component analysis yielded three factors of which only factor 1 (large positive loadings on all negative symptom complexes) and three (thought disordered syndrome) were replicable between uptake and follow-up.

4.9 Gender, cannabis use or abuse and sib pair status

Multi-group models were considered for gender, sibpair and cannabis. Firstly, the models were tested for measurement invariance, i.e. the item response probabilities are identical across groups. The results showed that the same measurement models could be accepted for the groups. The groups were then compared with regards to their latent class probability profiles, i.e. the latent class prevalences. The results indicate that for gender, sibpair and cannabis, the model releasing the equality constraints for the latent class probabilities fit better (p<0.001 for all three), revealing a different profile for the groups. The patients' gender, use of cannabis and sibpair status did not significantly impact on the item response probabilities. However, from the results it seems that male gender increases the probability of belonging to class 1 (negative symptom class), while cannabis use or abuse history increase the probability of belonging to class 2 (positive symptom class).

Our sample reported similar use or abuse of cannabis and this rate is well within the reported rates for first and multi-episode schizophrenia samples (Compton et al. 2004). In interpreting these findings, the reader should take note of the fact that cannabis use or abuse was defined as using cannabis more than 21 times in a single year. This history was ascertained as per DIGS guidelines. Although this might have led to over-inclusion of patients in this group, it is balanced by the fact that previous work from the author's group showed a very low rate of false reporting of cannabis use (Koen et al. 2009). Indeed, in this study, which includes 547 subject from the current study, Koen et al. (2009) reports only one case where the participant denied cannabis use, but was found

to be positive upon urine drug testing, Nevertheless, the data still rely on recall and voluntary reporting where cannabis use was not noted in clinical records or by informants. Ideally, a longitudinal study with regular objective monitoring of cannabis use over the lifetime, preferably with quantitative makers, is required. This is particularly true for periods of developmental sensitivity and thus also holds true for early developmental history. One of the challenges previously discussed (Niehaus 2008) in detail is the recall bias linked to the perception of importance of certain happenings, and the temporal relationship thereof, in the Xhosa population. The impression of the researcher is that this might have played a role in the very low report rate of developmental abnormalities. The reason for this recall bias is unknown but it impact significantly on the ability to do cross-sectional studies in this population. Nevertheless, it is not surprising to find the cannabis use or abuse group to have increased probability of belonging to class 2, as other authors (Salyers & Mueser 2001, Lysaker et al. 1994) have previously reported lower negative symptoms in substance abusers, including a first episode African American sample (Compton et al. 2004). A meta-analysis by Koskinen et al. (2010) showed that cannabis use disorders were commonly found in younger male first-episode schizophrenia patients. However, no association between cannabis use and negative symptomatology has also been reported in other studies (albeit mostly alcohol abuse or dependency, with or without cannabis use or abuse) and more specifically Kirkpatrick et al. (1996) who reported no difference in cannabis use between deficit and non-deficit patients. The proposed mechanism by which cannabis can trigger psychosis is via the cannabinoid receptors. Cannabis is the most potent agonist on the cannabinoid receptor 1 (CB1) and thus regulate dopaminergic communication between the brain stem and the stratum via GABAergic and glutamatergic nerve terminals (Morrison & Murray 2009). The best known study on the role of cannabis in the pathophysiology of schizophrenia is that of Andreasson et al. (1987) and the follow-up study by Zammit et al. (2002). This group studied more than 40,000 Swedish conscripts in a cohort study stretching over 15 years and observed a two-fold increase in risk of schizophrenia if cannabis exposure occurred prior to the time of conscription. This findings

persisted over the next 27 years of follow-up (Zammit et al. 2002) and showed a clear dose-response relationship with a 6-fold increased risk in heavy users.

Van Os et al. (2002) also replicated these findings in a Dutch sample of 4000 and found a three-fold increased risk and also a dose-response relationship although the baseline lifetime cannabis use was found to be more important than use during follow-up (3 years). This suggests that a vulnerable period of exposure may exist. Indeed the Dunedin study of a longitudinal birth cohort showed that use at age 15 predicted a 6 fold increased risk, although the presence of childhood psychotic symptoms decreased the risk to 3 fold increase (Arseneault et al. 2002) for schizophreniform disorder at age 26. The Christchurch study (Fergusson et al. 2003) found cannabis-dependence disorder at age 18 to predict increased risk for new psychotic symptoms over the 20 years of follow-up, independent of psychotic symptoms prior to entering the study. This was also mirrored on the European continent with Henquet et al. (2005) reporting a 1.7 fold increase psychotic symptoms in 2500 subjects followed up over four years in Germany.

Arseneault et al. (2002) reported a pooled odds ratio of 2.3 (95% CI = 1.7–2.9) when they reviewed (meta-analysis) prospective studies of cannabis and schizophrenia. The question of whether this vulnerability can be translated into a genetic vulnerability has led to some interesting studies and Caspi et al. (2005) reported that a valine (Val) to methionine (Met) substitution at codon 158 (Val158Met) of the catecholamine-O-methyl transferase (COMT) gene increased the risk for schizophreniform disorder following adolescent use of cannabis. Homozygotic (Val/Val) individuals were at the highest risk, followed by heterozygotes. The risk for homozygotes was also increased in a study where individuals with prior evidence of psychosis had increased cannabis-induced psychotic symptoms (Henquet et al. 2006).

4.10 Advantages and disadvantages of LCA in this study

The advantage of the latent class analysis approach is that it is data driven (i.e., classes are not determined by the researcher) and multivariate (can include covariates and categorical grouping variables), whereas most studies have a univariate or concept driven approach for example that of Csernansky et al. (1985). In addition this study makes use of structured clinical interviews in a relatively large sample of ethnically homogenous individuals. However, one of the factors to consider in interpreting this study is the fact the assessment of the positive and negative symptoms are cross-sectional and in general it is assumed that longitudinal variables are preferable (McGuffin et al. 1984). Andreasen (1983) developed the Scale for the Assessment of Negative Symptoms. The initial scale measured 25 negative symptoms of which several were dropped after extensive construct and validity studies. The scale groups symptoms into five sub-domains: affective flattening, alogia, avolition/apathy, anhedonia/asociality, and attentional impairment (Andreasen 1989). Nevertheless, negative symptom domains and specifically the well-described deficit schizophrenia seem to have high degree of longitudinal stability (Amador et al. 1999, Galderisi et al. 2009).

Although the researchers aimed to reduce selection bias by including all patients known to the mental health services with a diagnosis of schizophrenia, it is possible that inclusion of other groups of patients (schizo-affective disorder, schizophreniform) would have altered the outcome. However considering the similarities with previous LCAs (Peralta & Cuesta 2003, Kendler et al. 1998, Kendler et al. 1997) in terms of positive, negative, thought and behavioral symptoms, there seems to be substantial stability of some of the classes. Latent class analysis does have limitations as it assumes local independence within the resultant classes and LCA can only describe, and not prove, the existence of classes.

The major limitation of this and other LCA studies is the lack of any heritability estimate. Heritability data could provide a clue whether the LCA derived subgroups of schizophrenia outperform current classification systems in terms a relationship to genotype. Indeed, this study is geared towards the introduction of genetic data into the analysis of schizophrenia subgroups may lead to the identification of schizophrenia subgroups for psychiatric genetic analysis (Cardno et al. 2002). Taken together with the previous studies in this group, these results suggest that the proposed heterogeneity of schizophrenia is also sustained in an African population, both in a sib pair (Niehaus et al. 2006) and in a non-sib pair group (Niehaus et al. 2006, Niehaus 2008, current study). The sib pair group can be viewed as the more "familial" group and taking the findings from Sham et al. (1994) into account, one would have expected that this group would have theoretically impacted on class 1, which, similar to Type A of Sham et al. (1994), have more negative symptoms and male predominance. However, although the presence of an affected sib did increase the chance of belonging to group 1 (negative symptom group), our results differ from that of Sham et al. (1994) in that male gender was not overrepresented in this group. Nevertheless, it can be expected that although the core symptoms remain similar across ethnic boundaries, the genetic basis for schizophrenia might have differences in different ethnic groups. The probability of belonging to a class can be calculated for each individual and this information should be useful in candidate gene studies in this population. Indeed, given the previously reported low rates of obsessive compulsive disorder in the Xhosa schizophrenia group, findings in candidate gene studies may support the idea of epigenetic and genetic differences between ethnic groups. One limitation in the interpretation of these findings is that construction of the family trees, in by definition the ethnicity of the participant, are biased by recall of grandparent origin and the possibility of false paternity. The latter was found to be in the range of 12% in a previous study by the author (Data not published). However, the advantage of working within the Xhosa population is that the genetic architecture of the Niger-Kordofanian linguistic group (which includes the Nguni languages such as IsiXhosa and IsiZulu) seems fairly homogenous (May et al.

2013, Petersen et al. 2013) and this should limit the impact of false paternity or recall bias of ethnic origin, in terms of homogeneity. The next step in the search for subtypes of schizophrenia in the Xhosa population is the introduction of more complex models that also incorporates a measurement of severity to further delineate the interplay between these classes.

4.11 Factor mixture modeling

This is where FMM may contribute to the debate around symptom structure of schizophrenia. Looking at our results the FMM uncovered a heterogeneous latent variable structure that fits the data well with the latent classes capturing distinct behaviours and factors capturing severity variations. The following four classes were identified:

Class 1 (purple): Low symptom group (48.9%). As the name indicates this class was associated with the lowest positive symptom score. In essence this class had higher means for negative symptoms than for positive symptoms. This class combined class 1 (predominantly negative symptoms) and class 4 (no or few symptoms) in the LCA model.

Class 2 (green): High positive symptom group (16.4%). This class showed the highest mean scores in auditory hallucinations, global hallucinations, global delusions and bizarre behaviour. This class represents part of the class 2 group in the LCA model.

Class 3 (pink): Mixed symptom group (14.3%). This class showed a balance between positive and negative symptoms with the second lowest bizarre behaviour mean score across the groups. This class represents the second part of the class 3 group in the LCA model.

Class 4 (black): Medium high positive symptom group (20.4%). This group corresponds to the second part of the third class in the LCA model and represents a group with slightly more positive than negative symptoms.

The striking finding of this solution is that negative symptoms did not predict group placing. Furthermore the factor scores suggest there is large variation in severity within the class 1 (low symptom group). Indeed, class 1 (low symptom group) shows a bimodal distribution, suggesting the possibility of a second factor acting on this group (Figure 4.1). However, the addition of a second factor did not significantly contribute to the model and thus a one factor model is still favoured (Figure 4.1).

Figure 4.1 Top: Distribution plot of one factor model with two peaks. Bottom: Distribution plot of the two factor model showing that the second factor is not significantly contributing to the model.
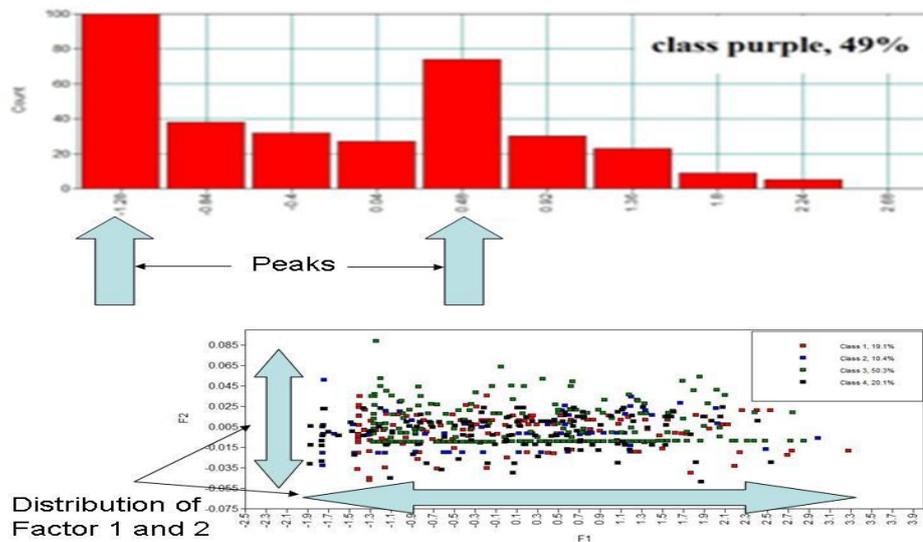
As FMM is a relatively new technique and very little guidance is available on the interpretation of the model within clinical environments, the author can only offer hypothesis that might explain this departure between LCA and FMM results. From the results one hypothesis is that within the negative symptom class a currently unknown variable or variables dictates severity within the class.

In essence these findings do not contradict the changes that occurred in the diagnostic systems (DSM V and ICD 11). The current DSM-V criteria require at least two of the following (one of which must be from A-C):

A. Delusions

B. Hallucinations

C. Disorganized speech

D. Grossly disorganized or catatonic behaviour

E. Negative symptoms.

The FMM model excludes only disorganized speech as a symptom group that does not play a significant role in the subtyping of schizophrenia within this sample. However, in terms of course specifiers of both DSM-V and ICD-11 the idea of acute episode, partial (ongoing symptoms) remission and full remission could reflect the four classes found in the FMM solution. In terms of positive symptoms, it seems that severity of positive symptoms, in terms of presence within a group, within the three groups with significant positive symptoms might differ only in severity of these symptoms. On the other hand, it might be that our findings reflect those of Oorschot et al. (2010). The authors reported that patients with both visual and auditory hallucinations reported levels of delusions intensity and that an increased delusional intensity preceded the onset of episodes of hallucinations. However, since positive symptoms are not necessarily stable over time, using these to subtype schizophrenia might be challenging.

 These results are sobering as even if we were able to model the impact of cannabis use, gender or sibship on each of these groups or classes, it may still be difficult to utilize these groups for genetic subtyping. It is thus important to investigate other avenues to possibly subtype

schizophrenia in a reliable manner. One such obvious mechanism is to explore the "unknown variable/s" that influences severity within the negative symptom class. The current FMM results are however exploratory and no conclusions can be drawn from these findings.

Comparing the FA, LCA and FMM findings (see below Figure 4.2) it looks as if some class membership changes occurred between the LCA and FMM classes with a suggestion that the FMM class 1 is made up by the majority of the subjects in class 4 and class 1 of the LCA. From this it seems as if the two models are similar in structure. It is interesting to note that the FA previously performed by Niehaus et al. (2008) showed that the negative factor accounted for almost half of the total variance in the sample. This might be seen as complementary to our finding that only one factor existed that was linked to a negative symptom "severity" factor.

Figure 4.2: Overview of FA, LCA and FMM findings. Arrows indicate possible group relationships between LCA and FMM classes and factors.
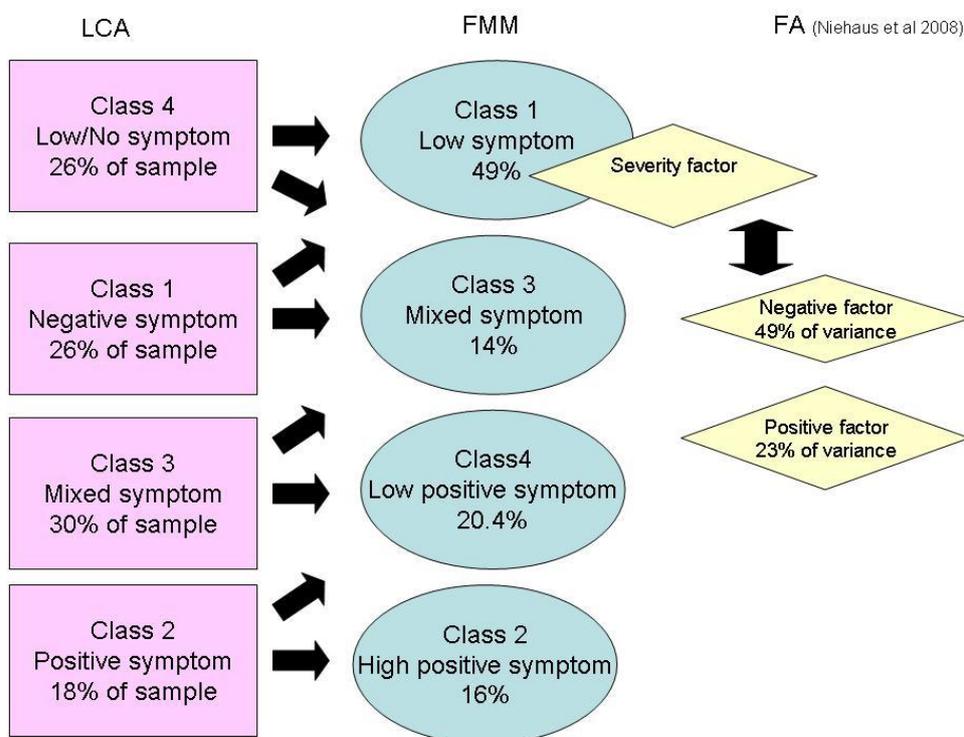
In order to hypothesis as what the variable/s are that may be reflected within the negative symptom "severity" factor, one can take one step back and look at the covariates found to be important in the LCA model such as age of onset, sibship, gender, use or abuse of cannabis and duration of illness (Figure 4.3).

Figure 4.3 Overview of FA, LCA and FMM analysis with covariates of cannabis use or abuse, sibship status, gender and duration of illness (0-6 years and 7-50 years) included.
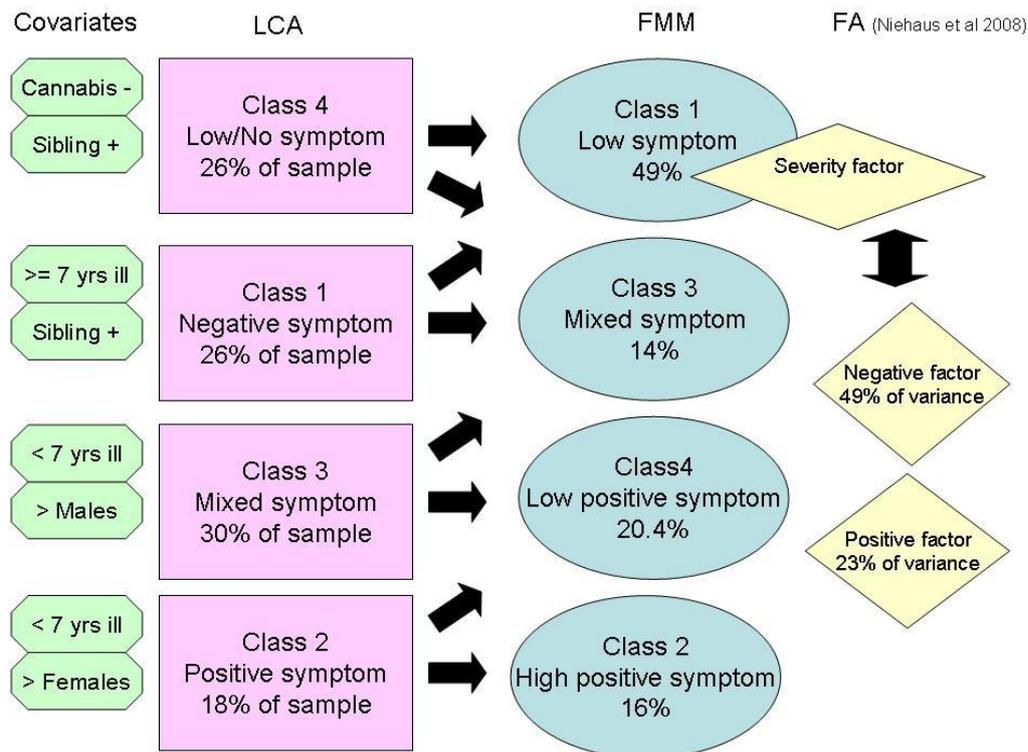
The presence of an affected sib and the absence of cannabis use or abuse history makes it more likely for a subject to be classified in class 4 (low/no symptom group), while a longer duration of illness and male gender makes it more likely to be classified in the class 1 or negative symptom group. If one hypothesis that the covariates cannabis use or abuse, sibship status and duration of illness might maintain their effect in class 1 and 3 of the FMM analysis, given that the subjects in the two LCA groups converged mostly into class 1 and 3. For the FMM model the OR for the presence of an affected sib in class 1 vs class 4 (average class) is 3.7 and the OR for the absence of cannabis use is 2.14 and the presence of males gender is 1.11. It is however important to note that other covariates such as social functioning (Blanchard et al. 2005) might be contributing to the unknown negative symptom "severity" factor.

In light of this it is important to note that the SANS and SAPS does not include ratings on two very important components of schizophrenia namely mood and cognitive symptoms. Although data was collected on the life-time presence of mood symptoms, no cognitive assessment was done in this large genetic study. The cost of including such a detailed assessment on a few hundred patients would be very expensive and not within the financial scope of the researchers. It is important to consider this when interpreting this study as it has been demonstrated that schizophrenia; bipolar disorder and depression share genetic underpinnings (Cardno et al. 2002, Lewis et al. 2003) and that the current classification of psychotic disorders may need to change to include mood and cognitive symptoms as part of the diagnostic criteria. However, as was seen with the results from this study, more complex models than the commonly used LCA and FA might be needed to fit the clinical complexity of psychiatric illness.

4.12 Conclusion

DeLisi (2011) in a commentary on series of papers on the facts known about schizophrenia highlights 10 facts that she perceives as the most likely to be true. These facts include the heterogeneous nature of schizophrenia, albeit it defined by a cluster of symptoms that seem to occur in most individuals with schizophrenia somewhere in their lifetime and shows a pleiotrophic nature (Tandon et al. 2009). Usually has it's onset in late adolescence to early adulthood, has a chronic course and only rarely associated with recovery after a first episode (Tandon et al. 2008). The current biological treatment option cannot cure or prevent the illness (Tandon et al. 2010). To date no specific biological factor has shown high sensitivity and specificity for schizophrenia. Although structural brain abnormalities are frequent in people with schizophrenia, it is not specific to schizophrenia (Keshavan et al. 2008). A wide range of neurotransmitters and receptors (post and presynaptically) abnormalities have been reported (Keshavan et al. 2008). Schizophrenia is highly

heritable but most likely polygenic in nature. Cognitive changes form an integral part of the pathology of schizophrenia and sex differences exist in illness characteristics and severity of course with males having a poorer prognosis.

Given the limited number of "facts" on schizophrenia, the author suggests that choosing a valid phenotype may be of particular importance but also very difficult. These phenotypes or "endophenotypes" should however show high sensitivity for the development of schizophrenia and should preferably not be present in "well" relatives. Indeed Keshavan et al. (2008) suggested that phenotypes should have specificity to the etiology as well as sensitivity. In addition, researchers should be aware that they do not become caught up in their specific hypotheses irrespective of the data. In this regard data-driven analysis may go help to question hypothesis (DeLisi 2011). Van Os (2011) argues for a reconceptualization of schizophrenia with a possible separation between clinically useful diagnostic criteria and research in psychiatry. Our study suggests that clinical measurement of symptoms in schizophrenia samples may be of limited value in establishing an endophenotype for genetic exploration, but importantly illustrates that negative symptoms in schizophrenia may harbor intriguing clues to severity of illness or other related factors (i.e. functioning, cognitive profile) in schizophrenia.

4.13 References

Addington J, Addington D. Positive and negative symptoms of schizophrenia

Their course and relationship over time. Schizophrenia Research 1991 5: 51-59.

Amador XF, Kirkpatrick B, Buchanan RW, et al. Stability of the diagnosis of deficit syndrome in schizophrenia. American Journal of Psychiatry 1999 156:637-639.

An der Heiden W, Konnecke R, Maurer K, et al. Depression in the long-term course of schizophrenia. European Archives of Psychiatry and Clinical Neuroscience 2005 255:174–184.

Andreasen NC. The diagnosis of schizophrenia. Schizophrenia Bulletin 1987 13:9-22.

Andreasen NC. The Scale for the Assessment of Negative Symptoms (SANS). University of Iowa, Iowa City, IA, 1983.

Andreasen NC, Olsen S. Negative v positive schizophrenia. Definition and validation.

Archives of General Psychiatry 1982 39(7):789-94.

Andreasen NC. The Scale for the Assessment of Negative Symptoms (SANS): conceptual and theoretical foundations. British Journal of Psychiatry 1989 7(Suppl):49-58.

Andreasson S, Allebeck P, Engstrom A, et al. Cannabis and schizophrenia. A longitudinal study of Swedish conscripts. Lancet 1987 2:1483–1486.

Andreasen NC, Arndt S, Alliger R, et al. Symptoms of schizophrenia: Methods, meanings, and mechanisms. Archives of General Psychiatry 1995 52:341–351.

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (Fifth ed.). Arlington, VA: American Psychiatric Publishing, 2013:5–25.

Argyle N. Panic attacks in chronic schizophrenia. British Journal of Psychiatry 1990 157:430–433.

Arnold LM, Keck PE, Collins J, et al. Ethnicity and first-rank symptoms in patients with psychosis. Schizophrenia Research 2004 67(2-3):207–212.

Arndt S, Andreasen NC, Flaum M, et al. A longitudinal study of symptom dimensions in schizophrenia: prediction and patterns of change. Archives of General Psychiatry 1995 52:341-51.

Arseneault L, Cannon M, Poulton R, et al. Cannabis use in adolescence and risk for adult psychosis: longitudinal prospective study. British Medical Journal 2002 325:1212–1213.

Baethge C, Baldessarini RJ, Freudenthal K, et al. Hallucinations in bipolar disorder: characteristics and comparison to unipolar depression and schizophrenia. Bipolar Disorder 2005 7(2):136–145.

Barrio C, Yamada AM, Atuel H, et al. A tri-ethnic examination of symptom expression on the positive and negative syndrome scale in schizophrenia spectrum disorders. Schizophrenia Research 2003 60(2-3):259-69.

Bassett AS, Collins EJ, Nuttall SE, et al. Positive and negative symptoms in families with schizophrenia. Schizophrenia Research 1993 11: 9- 19.

Bayle FJ, Krebs MO, Epelbaum C, et al. Clinical features of panic attacks in schizophrenia. European Psychiatry 2001 16:349–353.

Birchwood M, Iqbal Z, Chadwick P, et al. Cognitive approach to depression and suicidal thinking in psychosis. I: ontogeny of post-psychotic depression. British Journal of Psychiatry 2000 177:516–528.

Blanchard JJ, Horan WP, Collins LM. Examining latent structure of negative symptoms: Is there a distinct subtype of negative symptom schizophrenia? Schizophrenia Research 2005 77:151-165.

Bleuler E. Textbook of Psychiatry. MacMillan, New York, 1924.

Bracha HS, Wolkowitz OM, Lohr JB, et al. High prevalence of visual hallucinations in research subjects with chronic schizophrenia. American Journal of Psychiatry 1989 146:526-528.

Caspi A, Moffitt TE, Cannon M, et al. Moderation of the effect of adolescent-onset cannabis use on adult psychosis by a functional polymorphism in the catechol-O-methyltransferase gene:

longitudinal evidence of a gene X environment interaction. Biological Psychiatry 2005 57(10):1117-27.

Chang N, Newman J, D'Antonio E, et al. Ethnicity and symptom expression in patients with acute schizophrenia. Psychiatry Research 2011 185(3):453-5.

Compton MT, Furman AC, Kaslow NJ. Lower negative symptom scores among cannabis-dependent patients with schizophrenia-spectrum disorders: preliminary evidence from an African American first-episode sample. Schizophrenia Research 2004 71:61-64.

Csernansky J, Kaplan J, Hollister L. Problems in classification of schizophrenics as neuroleptic responders and non-responders. Journal of Nervous and Mental Disorders 1985 173:325.

De Leon J, Diaz FJ, Rogers T, et al. Initiation of daily smoking and nicotine dependence in schizophrenia and mood disorders. Schizophrenia Research 2002b 56(1-2):47–54.

De Leon J, Cuesta MJ, Peralta V. Delusions and hallucinations in schizophrenic patients. Psychopathology 1993 26:286–291.

De Leon J, Dadvand M, Canuso C, et al. Schizophrenia and smoking: an epidemiological survey in a state hospital. American Journal of Psychiatry 1995 152:453–455.

De Leon J, Tracy J, McCann E, et al. Schizophrenia and tobacco smoking: a replication study in another US psychiatric hospital. Schizophrenia Research 2002 56(1-2):55–65.

DeLisi LE. Moving on in Schizophrenia Research to the next decade: Commentary on Keshavan, Nasrallah and Tandon. Schizophrenia Research 2011 127:14-15.

Derks EM, Allardyce J, Boks MP, et al. Kraepelin Was Right: A Latent Class Analysis of Symptom Dimensions in Patients and Controls. Schizophrenia Bulletin 2012 38(3):495-505.

Emsley RA, Niehaus DJH, Mbanga NI, et al. The factor structure for positive and negative symptoms in South African Xhosa patients with schizophrenia. Schizophrenia Research 2001 47:149– 157.

Esterberg ML, Trotman HD, Holtzman C, et al. The impact of a family history of psychosis on age-at-onset and positive and negative symptoms of schizophrenia: A meta-analysis. Schizophrenia Research 2010 120:121–130.

Fenton WS. Depression, suicide and suicide prevention in schizophrenia. Suicide and Life Threatening Behaviour 2000 30:34–49.

Fergusson DM, Horwood LJ, Swain-Campbell NR. Cannabis dependence and psychotic symptoms in young people. Psychological Medicine 2003 33:15–21.

Foussias G, Remington G. Negative symptoms in schizophrenia: Avolition and Occam's razor. Schizophrenia Bulletin 2008 36:359−369.

Galderisi S, Maj M. Deficit schizophrenia: an overview of clinical, biological and treatment aspects. European Psychiatry 2009 24:493-500.

Gauntlett-Gilbert J, Kuipers E. Phenomenology of visual hallucinations in psychiatric conditions. Journal of Nervous and Mental Disorders 2003 191(3):203–205.

Glassman AH. Cigarette smoking: implications for psychiatric illness. American Journal of Psychiatry 1993 150(4):546–553.

Goodwin DW, Alderson P, Rosenthal R. Clinical significance of hallucinations in psychiatric disorders. A study of 116 hallucinatory patients. Archives of General Psychiatry 1971 24:76-80.

Goodwin RF, Amador XF, Malaspina D, et al. Anxiety and substance use comorbidity among inpatients with schizophrenia. Schizophrenia Research 2003 61:89–95.

Hafner H, Loffler W, Maurer K, et al. Depressive, negative symptoms, social stagnation and social decline in the early course of schizophrenia. Acta Psychiatrica Scandinavia 1999 100:105–118.

Heckers S, Barch DM, Bustillo J, et al. Structure of the psychotic disorders classification in DSM5, Schizophrenia Research 2013 in press.

Henquet C, Murray R, Linszen D, et al. The environment and schizophrenia: the role of cannabis use. Schizophrenia Bulletin 2005 31:608–612.

Henquet C, Rosa A, Krabbendam L, et al. An experimental study of catechol-o-methyltransferase Val158Met moderation of delta-9-tetrahydrocannabinol-induced effects on psychosis and cognition. Neuropsychopharmacology 2006 31:2748–2757.

Huang GH, Tsai HH, Hwub HG et al. Patient subgroups of schizophrenia based on the Positive and Negative Syndrome Scale: composition and transition between acute and subsided disease states. Comprehensive Psychiatry 2011 52:469–478.

Hughes JR, Hatsukami DK, Mitchell JE, et al. Prevalence of smoking among psychiatric outpatients. American Journal of Psychiatry 1986 143(8):993–997.

Kapur S. Psychosis as a state of aberrant salience: a framework linking biology, phenomenology, and pharmacology in schizophrenia. American Journal of Psychiatry 2003 160(1):13–23.

Keefe RS, Harvey PD, Lenzenweger MF, et al. Empirical assessment of the factorial structure of clinical symptoms in schizophrenia: Negative symptoms. Psychiatry Research 1992 44:153−165.

Kelley ME, van Kammen DP, Allen DN. Empirical validation of primary negative symptoms: Independence from effects of medication and psychosis. The American Journal of Psychiatry 1999 156:406−411.

Kendler KS, Karkowski-Shuman L, O'Neill FA, et al. Resemblance of psychotic symptoms and syndromes in affected sibling pairs from the Irish Study of High-Density Schizophrenia Families: evidence for possible etiologic heterogeneity. American Journal of Psychiatry 1997 154(2):191-8.

Kendler KS, Karkowski LM, Walsh D. The structure of psychosis: latent class analysis of probands from the Roscommon Family Study. Archives of General Psychiatry 1998 55:492–499.

Keshavan MS, Tandon R, Boutros NN, et al. Schizophrenia, "just the facts": what we know in 2008 Part 3: neurobiology. Schizophrenia Research 2008 106(2–3): 89–107.

Kimhy D, Yale S, Goetz RR, et al. The factorial structure of the schedule for the deficit syndrome in schizophrenia. Schizophrenia Bulletin 2006 32(2):274-8.

Kirkpatrick B, Amador XF, Flaum M, et al. The deficit syndrome in the DSM-IV Field Trial: I. Alcohol and other drug use. Schizophrenia Research 1996 20:69-77.

Kirkpatrick B, Buchanan RW, McKenney PD, et al. The Schedule for the Deficit syndrome: an instrument for research in schizophrenia. Psychiatry Research 1989 30(2):119-23.

Kirkpatrick B, Fenton WS, Carpenter WT, et al. The NIMHMATRICS consensus statement on negative symptoms. Schizophrenia Bulletin 2006 32:214−219.

Kirkpatrick B, Ram R, Amador XF, et al. Summer birth and the deficit syndrome of schizophrenia. American Journal of Psychiatry 1998 155:1221–1226.

Kirkpatrick B, Ross DE, Walsh D, et al. Family characteristics of deficit and non-deficit schizophrenia in the Roscommon family study. Schizophrenia Research 2000 45:57-64.

Koen L, Jonathan R, Niehaus DJH. Cannabis use and abuse correlates in a homogeneous South African schizophrenia population. South African Journal of Psychiatry 2009 15:8-12.

Koreen AR, Siris SG, Chakos M, et al. Depression in first episode schizophrenia. American Journal of Psychiatry 1993 150:1643.

Koskinen J, Löhönen J, Koponen H, Rate of cannabis use disorders in clinical samples of patients with schizophrenia: a meta-analysis. Schizophrenia Bulletin 2010 36(6):1115-30.

Krabbendam L, van Os J. Affective processes in the onset and persistence of psychosis. European Archives of Psychiatry and Clinical Neuroscience 2005 255(3):185–189.

Kring AM, Gur RE, Blanchard JJ, Horan WP, et al. The Clinical Assessment Interview for Negative Symptoms (CAINS): final development and validation. American Journal of Psychiatry 2013 170(2):a165–172.

Kulhara P, Chandiramani K. Positive and negative subtypes of schizophrenia. A follow-up study from India. Schizophrenia Research 1990 3(2):107-16.

Kumari V, Postma P. Nicotine use in schizophrenia: The self medication hypotheses. Neuroscience and Biobehavioral Reviews 2005 29:1021–1034.

Labbate LA, Young PC, Arana GW. Panic disorder in schizophrenia. Canadian Journal of Psychiatry 1999 44:488–490.

Lançon C, Auquier P, Reine G, et al. Relationships between depression and psychotic symptoms of schizophrenia during an acute episode and stable period. Schizophrenia Research 2001 47:135–140.

Lewandowski KE, DePaola J, Camsari GB, et al. Tactile, olfactory, and gustatory hallucinations in psychotic disorders: a descriptive study. Annals of Academic Medicine Singapore 2009 38:383-385.

Lewis CM, Levinson DF, Wise LH, et al. Genome scan meta-analysis of schizophrenia and bipolar disorder, part II: Schizophrenia. American Journal of Human Genetics 2003 73(1):34-48.

Liddle PF, Barnes TR. Syndromes of chronic schizophrenia. British Journal of Psychiatry 1990 157:558–561.

Liddle PF. The symptoms of chronic schizophrenia: a re-examination of the positive-negative dichotomy. British Journal of Psychiatry 1987 151:125-151.

Lohr JB, Flynn K. Smoking and schizophrenia. Schizophrenia Research 1992 8:93–102.

Lysaker P, Bell M, Beam-Goulet J, et al. Relationship of positive and negative symptoms to cocaine abuse in schizophrenia. Journal of Nervous and Mental Disorders 1994 182:109-112.

Maher BA. Delusional thinking and perceptual disorder. Journal of Individual Psychology 1974 30(1):98–113.

Malla AK, Takhar JJ, Norman RM, et al. Negative symptoms in first episode non-affective psychosis. Acta Psychiatrica Scandinavica 2002 105:431−439.

Mark TL, Palmer LA, Russo PA, et al. Examination of treatment pattern differences by race. Mental Health Services Research 2003 5(4):241-50.

May A, Hazelhurst S, Li Y, et al. Genetic diversity in black South Africans from Soweto. BMC Genomics 2013 14:644-658.

Martin RL, Cloninger CR, Guze SB, et al. Frequency and differential diagnosis of depressive syndromes in schizophrenia. Journal of Clinical Psychiatry 1985 46:9–13.

McGuffin P, Farmer AE, Gottesman II, et al. Twin concordance for operationally defined schizophrenia. Archives of General Psychiatry 1984 41:541-545.

McGlashan TH, Carpenter WT Jr. Postpsychotic depression in schizophrenia. Archives of General Psychiatry 1976 33:231–239.

McGrath JA, Nestadt G, Liang KY, et al. Five latent factors underlying schizophrenia: analysis and relationship to illnesses in relatives. Schizophrenia Bulletin 2004 30(4):855-73.

Mellor CS. First rank symptoms of schizophrenia. British Journal of Psychiatry 1970 117:15–23.

Messinger JW, Tremeau F, Antonius D, et al. Avolition and expressive deficits capture negative symptom phenomenology: Implications for DSM-5 and schizophrenia research. Clinical Psychology Review 2011 31:161–168.

Minas IH, Klimidis S, Stuart GW, et al. Positive and negative symptoms in the psychoses: principal components analysis of items from the Scale for the Assessment of Positive Symptoms and the Scale for the Assessment of Negative Symptoms. Comprehensive Psychiatry 1994 35:135−144.

Morrison PD, Murray RM. From real-world events to psychosis: the emerging neuropharmacology of delusions. Schizophrenia Bulletin 2009 35:668–674.

Mueser KT, Bellack AS, Brady EU. Hallucinations in schizophrenia. Acta Psychiatrica Scandinavia 1990 82:26-29.

Mueser KT, Sayers SL, Schooler NR, et al. A multisite investigation of the reliability of the Scale for the Assessment of Negative Symptoms. The American Journal of Psychiatry 1994 151:1453−1462.

Nakaya M, Ohmori K. A two-factor structure for the Schedule for the Deficit Syndrome in schizophrenia. Psychiatry Research 2008 158:256−259.

Niehaus DJH. The role of sibpairs in limiting clinical heterogeneity in a Xhosa schizophrenia population. DMed Thesis, University of Stellenbosch, 2008.

Olincy A, Young DA, Freedman R. Increased levels of the nicotine metabolite cotinine in schizophrenic smokers compared to other smokers. Biological Psychiatry 1997 42(1):1–5.

Oorschot M, Lataster T, Thewissen V, et al. Temporal dynamics of visual and auditory hallucinations in psychosis. Schizophrenia Research 2012 140:77–82.

Oosthuizen P, Emsley R, Niehaus D, et al. The relationships between depression and remission in first-episode psychosis. World Psychiatry 2006 5:172–176.

Pallanti S, Quercioli L, Hollander E. Social anxiety in outpatients with schizophreniaa relevant cause of disability. American Journal of Psychiatry 2004 161:53–58.

Peralta V, Cuesta MJ. The Nosology of Psychotic Disorders. Schizophrenia Bulletin 2003 29(3):413-425.

Peralta V, Cuesta MJ. Dimensional structure of psychotic symptoms: an item-level analysis of SAPS and SANS symptoms in psychotic disorders. Schizophrenia Research 1999 38:13−26.

Peralta V, Cuesta MJ. Factor Structure and Clinical Validity of Competing Models of Positive Symptoms in Schizophrenia. Biological Psychiatry 1998 44:107−114.

Petersen DC, Libiger O, Tindall EA, et al. Complex patterns of genomic admixture within Southern Africa. PLOS Genetics 2013 9:eprint.

Salyers MP, Mueser KT. Social functioning, psychopathology, and medication side effects in relation to substance use and abuse in schizophrenia. Schizophrenia Research 2001 48:109-123.

Sayers SL, Curran PJ, Mueser KT. Factor structure and construct validity of the scale for the assessment of negative symptoms. Psychological Assessment 1996 8:269−280.

Schultz SK, Miller DD, Oliver SE, et al. The life course of schizophrenia: age and symptom dimensions. Schizophrenia Research 1997 23:15-23.

Schultze-Lutter F, Ruhrmann S, Picker H, et al. Basic symptoms in early psychotic and depressive disorders. British Journal of Psychiatry 2007 191:s31–s37.

Sham PC, MacLean CJ, Kendler KS. A typological model of schizophrenia based on age at onset, sex and familial morbidity. Acta Psychiatrica Scandinavia 1994 89:135-141.

Siris SG. Diagnosis of secondary depression in schizophrenia: implications for DSM-IV. Schizophrenia Bulletin 1991 17:75–98.

Siris SG. Depression in schizophrenia: perspectives in the era of "atypical" antipsychotic agents. American Journal of Psychiatry 2000 157:1379–1389.

Small IF, Small JG, Andersen JM. Clinical characteristics of hallucinations of schizophrenia. Diorders of the Nervous System 1996 27:349-353.

Smit J, Swartz L, Kilian S, et al. Mediating words, mediating worlds: Interpreting as hidden work in a South African psychiatric institution. Transcultural Psychiatry 2013 50:493-514.

Strakowski SM, Flaum M, Amador X, et al. Racial differences in the diagnosis of psychosis. Schizophrenia Research 1996 21(2):117-24.

Strakowski SM, McElroy SL, Keck PE Jr, et al. Racial influence on diagnosis in psychotic mania. Journal of Affective Disorders 1996 39(2):157-62.

Strand JE, Nyback H. Tobacco use in schizophrenia: a study of cotinine concentrations in the saliva of patients and controls. European Psychiatry 2005 20(1):50–54.

Swartz L. Culture and Mental Health: A South African View. Oxford University Press, Cape Town, South Africa, 1998.

Tandon R, Nasrallah HA, Keshavan MS. Schizophrenia, "just the facts" 4. Clinical features and conceptualization. Schizophrenia Research 2009 110(1-3):1-23.

Tandon R, Nasrallah HA, Keshavan MS. Schizophrenia, "just the facts" 5. Treatment and prevention. Past, present, and future. Schizophrenia Research 2010 122(1-3):1-23.

Tibbo P, Swainson J, Chue P, et al. Prevalence and relationship to delusions and hallucinations of anxiety disorders in schizophrenia. Depression and Anxiety 2003 17:65–72.

Toomey R, Kremen WS, Simpson JC, et al. Revisiting the factor structure for positive and negative symptoms: evidence from a large heterogeneous group of psychiatric patients. The American Journal of Psychiatry 1997 154:371−377.

Tremeau F, Antonius D, Cacioppo JT, et al. In support of Bleuler: Objective evidence for increased affective ambivalence in schizophrenia based upon evocative testing. Schizophrenia Research 2009 107:223−231.

Uck A, Polat A, Bozkurt O, et al. Cigarette smoking among patients with schizophrenia and bipolar disorders. Psychiatry and Clinical Neuroscience 2004 58(4):434–437.

Upthegrove R, Birchwood M, Ross K, et al. The evolution of depression and suicidality in first episode psychosis. Acta Psychiatrica Scandinavia 2010 122 (3):211–218.

Van Os J. From schizophrenia metafacts to non-schizophrenia facts. Schizophrenia Research 2011 172:16-17.

Van Os J, Hanssen M, Bijl RV, et al. Strauss (1969) revisited: a psychosis continuum in the general population? Schizophrenia Research 2000 45(1–2):11–20.

Van Os J, Bak M, Hanssen M, et al. Cannabis use and psychosis: a longitudinal population-based study. American Journal of Epidemiology 2002 156:319–327.

Van 't Wout M, Aleman A, Kessels RP, et al. Emotional processing in a non-clinical psychosis-prone sample. Schizophrenia Research 2004 68(2–3):271–281.

Weisman AQ, Lopez SR, Ventura J, et al. A Comparison of Psychiatric Symptoms Between Anglo-Americans and Mexican- Americans With Schizophrenia. Schizophrenia Bulletin 2000 26(4):817-824.

Zammit S, Allebeck P, Andreasson S, et al. Self reported cannabis use as a risk factor for schizophrenia in Swedish conscripts of 1969: historical cohort study. British Medical Journal 2002 325:1199.

Zisook S, McAdams LA, Kuck J, et al. Depressive symptoms in schizophrenia. American Journal of Psychiatry 1999 156:1736–1743.

Appendix 5.1: Results of SANS/SAPS Interviews: Criteria for diagnosing the presence of each sign or symptom are given as direct quotes from Andreassen et al (1984)

Persecutory Delusions

"People suffering from persecutory delusions believe that they are being conspired against or persecuted in some way. Common manifestations include the belief that one is being followed, that one's mail is being opened, that one's room or office is bugged, that the telephone is tapped, or that police, government officials, neighbors, or fellow workers are harassing the subject. Persecutory delusions are sometimes relatively isolated or fragmented, but sometimes the subject has a complex set of delusions involving both a wide range of forms of persecution and a belief that there is a well-designed conspiracy behind them. For example, a subject may believe that his house is bugged and that he is being followed because the government wrongly considers him a secret agent for a foreign government; this delusion may be so complex that it explains almost everything that happens to him. The ratings of severity should be based on duration and complexity."

Persecutory delusions: SAPS item scores

SAPS Score	Frequency	Percentage
0	370	50.2
1	18	2.4
2	87	11.8
3	133	18.0
4	93	12.6
5	16	2.2
Total	717	97.3
Missing data	20	2.7
Total	737	100.0

Delusions of Jealousy

"The subject believes that his/her mate is having an affair with someone. Miscellaneous bits of information are construed as "evidence". The person usually goes to great effort to prove the existence of the affair, searching for hair in the bedclothes, the odor of shaving lotion or smoke on clothing, or receipts or checks indicating a gift has been bought for the lover. Elaborate plans are often made in order to trap the two together." Andreasen 1984

SAPS Score	Frequency	Percentage
0	0	683
1	1	7
2	2	9
3	3	4
4	0	0
5	0	0
Total	703	95.4
Missing data	20	34
Total	737	100.0

Delusions of Sin or Guilt

"The subject believes that he has committed some terrible sin or done something unforgivable. Sometimes the subject is excessively or inappropriately preoccupied with things he did wrong as a child, such as masturbating. Sometimes the subject feels responsible for causing some disastrous event, such as a fire or accident, with which he in fact has no connection. Sometimes these delusions may have a religious flavor, involving the belief that the sin is unpardonable and that the subject will suffer eternal punishment from God. Sometimes the subject simply believes that he deserves punishment by society. The subject may spend a good deal of time confessing these sins to whomever will listen." Andreasen 1984

SAPS Score	Frequency	Percentage
0	672	91.2
1	7	.9
2	16	2.2

3	9	1.2
4	2	.3
Total	706	95.8
Missing data	31	4.2
Total	737	100.0

Grandiose Delusions

"The subject believes that he has special powers or abilities. He may think he is actually some famous personage, such as a rock star, Napoleon, or Christ. He may believe he is writing some definitive book, composing a great piece of music, or developing some wonderful new invention. The subject is often suspicious that someone is trying to steal his ideas, and he may become quite irritable if his ideas are doubted." Andreasen 1984

SAPS Score	Frequency	Percentage
0	567	76.9
1	9	1.2
2	44	6.0
3	56	7.6
4	26	3.5
5	5	.7
Total	707	95.9
Missing data	30	4.1
Total	737	100.0

Religious Delusions

"The subject is preoccupied with false beliefs of a religious nature. Sometimes these exist within the context of a conventional religious system, such as beliefs about the Second Coming, the Antichrist, or possession by the Devil. At other times, they may involve an entirely new religious system or a pastiche of beliefs from a variety of religions, particularly Eastern religions, such as ideas about reincarnation or Nirvana. Religious delusions may be combined with grandiose delusions (if the subject considers himself a religious leader), delusions of guilt, or delusions of being controlled. Religious delusions must be outside the range considered normal for the subject's cultural and religious background." Andreasen 1984

SAPS Score	Frequency	Percentage
0	594	80.6
1	15	2.0
2	44	6.0
3	37	5.0
4	15	2.0
5	3	.4
Total	708	96.1
Missing data	29	3.9
Total	737	100.0

Somatic Delusions

"The subject believes that somehow his body is diseased, abnormal, or changed. For example, he may believe that his stomach or brain is rotting, that his hands or penis have become enlarged, or that his facial features are unusual (dysmorphophobia). Sometimes somatic delusions are accompanied by tactile or other hallucinations, and when this occurs, both should be rated. (For example, the subject believes that he has ballbearings rolling around in his head, placed there by a dentist who filled his teeth, and can actually hear them clanking against one another.)" Andreasen 1984

SAPS Score	Frequency	Percentage
0	633	85.9
1	7	9
2	29	3.9
3	23	3.1
4	12	1.6
5	2	.3
Total	706	95.8
Missing data	31	4.2
Total	737	100.0

Ideas and Delusions of Reference

"The subject believes that insignificant remarks, statements, or events refer to him or have some special meaning for him. For example, the subject walks into a room, sees people laughing, and suspects that they were just talking about him and laughing at him. Sometimes items read in the paper, heard on the radio, or seen on television are considered to be special messages to the subject. In the case of ideas of reference, the subject is suspicious, but recognizes his idea is erroneous. When the subject actually believes that the statements or events refer to him, then this is considered a delusion of reference." Andreasen 1984

SAPS Score	Frequency	Percentage
0	505	68.5
1	27	3.7
2	64	8.7
3	82	11.1
4	28	3.8
5	3	.4
Total	709	96.2
Missing data	28	3.8
Total	737	100.0

Delusions of Being Controlled

"The subject has a subjective experience that his feelings or actions are controlled by some outside force. The central requirement for this type of delusion is an actual strong subjective experience of being controlled. It does not include simple beliefs or ideas, such as that the subject is acting as an agent of God or that friends or parents are trying to coerce him to do something. Rather, the subject must describe, for example, that his body has been occupied by some alien force that is making it move in peculiar ways, or that messages are being sent to his brain by radio waves and causing him to experience particular feelings that he recognizes are not his own." Andreasen 1984

SAPS Score	Frequency	Percentage
0	619	84.0
1	10	1.4
2	25	3.4
3	30	4.1
4	17	2.3
5	2	.3
Total	703	95.4
Missing data	34	4.6
Total	737	100.0

Delusions of Mind Reading

"The subject believes that people can read his mind or know his thoughts. This is different than thought broadcasting (see below) in that it is a belief without a percept. That is, the subject subjectively experiences and recognizes that others know his thoughts, but he does not think that they can be heard out loud."

Andreasen 1984

SAPS Score	Frequency	Percentage
0	585	79.4
1	12	1.6
2	57	7.7
3	36	4.9
4	13	1.8
5	1	.1
Total	704	95.5
Missing data	33	4.5
Total	737	100.0

Thought Broadcasting

"The subject believes that his thoughts are broadcast so that he or others can hear them. Sometimes the subject experiences his thoughts as a voice outside his head; this is an auditory hallucination as well as a delusion. Sometimes the subject feels his thoughts are being broadcast although he cannot hear them himself. Sometimes he believes that his thoughts are picked up by a microphone and broadcast on the radio or television." Andreasen 1984

SAPS Score	Frequency	Percentage
0	617	83.7
1	6	.8
2	40	5.4
3	30	4.1
4	9	1.2
5	1	.1
Total	703	95.4
Missing data	34	4.6
Total	737	100.0

Thought Insertion

"The subject believes that thoughts that are not his own have been inserted into his mind. For example, the subject may believe that a neighbor is practicing voodoo and planting alien sexual thoughts in his mind. This symptom should not be confused with experiencing unpleasant thoughts that the subject recognizes as his own, such as delusions of persecution or guilt." Andreasen 1984

SAPS Score	Frequency	Percentage
0	638	86.6
1	4	.5
2	34	4.6
3	23	3.1
4	5	.7
5	1	.1
Total	705	95.7

Missing data	32	4.3
Total	737	100.0

Thought Withdrawal

"The subject believes that thoughts have been taken away from his mind. He is able to describe a subjective experience of beginning a thought and then suddenly having it removed by some outside force. This symptom does not include the mere subjective recognition of alogia." Andreassen 1984

SAPS Score	Frequency	Percentage
0	640	86.8
1	5	.7
2	31	4.2
3	20	2.7
4	6	.8
Total	702	95.3
Missing data	35	4.7
Total	737	100.0

Global Delusion Score

SAPS Score	Frequency	Percentage
0	321	43.6
1	13	1.8
2	118	16.0
3	141	19.1
4	103	14.0
5	16	2.2
Total	712	96.6
Missing data	25	3.4
Total	737	100.0

Auditory Hallucinations

"The subject has reported voices, noises, or sounds. The commonest auditory hallucinations involve hearing voices speaking to the subject or calling him names. The voices may be male or female, familiar or unfamiliar, and critical or complimentary. Typically, subjects suffering from schizophrenia experience the voices as unpleasant and negative. Hallucinations involving sounds rather than voices, such as noises or music, should be considered less characteristic and less severe." Andreasen 1984

SAPS Score	Frequency	Percentage
0	355	48.2
1	9	1.2
2	69	9.4
3	146	19.8
4	112	15.2
5	26	3.5
Total	717	97.3
Missing data	20	2.7
Total	737	100.0

Voices Commenting

"Voices commenting are a particular type of auditory hallucination which phenomenologists as Kurt Schneider consider to be pathognomonic of schizophrenia, although some recent evidence contradicts this. These hallucinations involve hearing a voice that makes a running commentary on the subject's behavior or thought as it occurs. If this is the only type of auditory hallucination that the subject hears, it should be scored instead of auditory hallucinations (No. 1 above). Usually, however, voices commenting will occur in addition to other types of auditory hallucinations." Andreasen 1984

SAPS Score	Frequency	Percentage
0	524	71.1
1	9	1.2
2	47	6.4

3	63	8.5
4	55	7.5
5	5	.7
Total	703	95.4
Missing data	34	4.6
Total	737	100.0

Voices Conversing

"Like voices commenting, voices conversing are considered a Schneiderian first-rank symptom. They involve hearing two or more voices talking with one another, usually discussing something about the subject. As in the case of voices commenting, they should be scored independently of other auditory hallucinations." Andreasen 1984

SAPS Score	Frequency	Percentage
0	484	65.7
1	6	.8
2	52	7.1
3	81	11.0
4	70	9.5
5	12	1.6
Total	705	95.7
Missing data	32	4.3
Total	737	100.0

Somatic or Tactile Hallucinations

"These hallucinations involve experiencing peculiar physical sensations in the body. They include burning sensations, tingling, and perceptions that the body has changed in shape or size." Andreasen 1984

SAPS Score	Frequency	Percentage
0	613	83.2
1	11	1.5

2	54	7.3
3	19	2.6
4	8	1.1
5	5	.7
Total	710	96.3
Missing data	27	3.7
Total	737	100.0

Olfactory Hallucinations

"The subject experiences unusual smells which are typically quite unpleasant. Sometimes the subject may

believe that he himself smells. This belief should be scored here if the subject can actually smell the odor

himself, but should be scored among delusions if he only believes that others can smell the odor."

Andreasen 1984

SAPS Score	Frequency	Percentage
0	617	83.7
1	8	1.1
2	62	8.4
3	15	2.0
4	5	.7
Total	707	95.9
Missing data	30	4.1
Total	737	100.0

Visual Hallucinations

"The subject sees shapes or people that are not actually present. Sometimes these are shapes or colors, but

most typically they are figures of people or human-like objects. They may also be characters of a religious

nature, such as the Devil or Christ. As always, visual hallucinations involving religious themes should be

judged within the context of the subject's cultural background. Hypnogogic and hypnopompic visual

hallucinations (which are relatively common) should be excluded, as should visual hallucinations

occurring when the subject has been taking hallucinogenic drugs.'' Andreasen 1984

SAPS Score	Frequency	Percentage
0	565	76.7
1	15	2.0
2	63	8.5
3	46	6.2
4	18	2.4
5	1	.1
Total	708	96.1
Missing data	29	3.9
Total	737	100.0

Global Hallucinations Score

SAPS Score	Frequency	Percentage
0	345	46.8
1	8	1.1
2	91	12.3
3	150	20.4
4	101	13.7
5	18	2.4
Total	713	96.7
Missing data	24	3.3
Total	737	100.0

Behavioral features

Clothing and Appearance

"The subject dresses in an unusual manner or does other strange things to alter his appearance. For

example, he may shave off all his hair or paint parts of his body

different colors. His clothing may be quite unusual; for example, he may choose to wear some outfit that

appears generally inappropriate and unacceptable, such as a baseball cap backwards with rubber galoshes

and long underwear covered by denim overalls. He may dress in a fantastic costume representing some

historical personage or a man from outer space. He may wear clothing completely inappropriate to the

climatic conditions, such as heavy wools in the midst of summer." Andreasen 1984

SAPS Score	Frequency	Percentage
0	526	71.4
1	38	5.2
2	83	11.3
3	44	6.0
4	22	3.0
5	6	.8
Total	719	97.6
Missing data	18	2.4
Total	737	100.0

Social and Sexual Behavior

"The subject may do things that are considered inappropriate according to usual social norms. For

example, he may masturbate in public, urinate or defecate in inappropriate receptacles, or exhibit his sex

organs inappropriately. He may walk along the street muttering to himself, or he may begin talking to

people whom he has never met about his personal life (as when riding on a subway or standing in some

public place). He may drop to his knees praying and shouting in the midst of a crowd of people, or he may

suddenly sit in a yoga position while in the midst of a crowd. He may make inappropriate sexual overtures

or remarks to strangers." Andreasen 1984

SAPS Score	Frequency	Percentage
0	509	69.1

1	15	2.0
2	78	10.6
3	73	9.9
4	38	5.2
5	7	.9
Total	720	97.7
Missing data	17	2.3
Total	737	100.0

Aggressive and Agitated Behavior

"The subject may behave in an aggressive, agitated manner, often quite unpredictably. He may start arguments inappropriately with friends or members of his family, or he may accost strangers on the street and begin haranguing them angrily. He may write letters of a threatening or angry nature to government officials or others with whom he has some quarrel. Occasionally, subjects may perform violent acts such as injuring or tormenting animals, or attempting to injure or kill human beings." Andreasen 1984

SAPS Score	Frequency	Percentage
0	488	66.2
1	18	2.4
2	61	8.3
3	74	10.0
4	67	9.1
5	12	1.6
Total	720	97.7
Missing data	17	2.3
Total	737	100.0

Repetitive or Stereotyped Behavior

"The subject may develop a set of repetitive actions or rituals that he must perform over and over. Frequently, he will attribute some symbolic significance to these actions and believe that they are either influencing others or preventing himself from being influenced. For example, he may eat jelly beans every night for dessert, assuming that different consequences will occur depending on the color of the jelly beans.

He may have to eat foods in a particular order, wear particular clothes, or put them on in a certain order.

He may have to write messages to himself or to others over and over; sometimes this will be in an unusual

or occult language." Andreasen 1984

SAPS Score	Frequency	Percentage
0	664	90.1
1	8	1.1
2	22	3.0
3	21	2.8
4	4	.5
5	3	.4
Total	722	98.0
Missing data	15	2.0
Total	737	100.0

Global Behaviour Changes Score

SAPS Score	Frequency	Percentage
0	408	55.4
1	30	4.1
2	97	13.2
3	99	13.4
4	73	9.9
5	11	1.5
Total	718	97.4
Missing data	19	2.6
Total	737	100.0

Derailment (Loose Associations)

"A pattern of spontaneous speech in which the ideas slip off one track onto another which is clearly but

obliquely related, or onto one which is completely unrelated. Things may be said in juxaposition which lack

a meaningful relationship, or the subject may shift idiosyncratically from one frame of reference to another.

At times there may be a vague connection between the ideas, and at others none will be apparent. This

pattern of speech is often characterized as sounding "disjointed." Perhaps the commonest manifestation of this disorder is a slow, steady slippage, with no single derailment being particularly severe, so that the speaker gets farther and farther off the track with each derailment without showing any awareness that his reply no longer has any connection with the question which was asked. This abnormality is often characterized by lack of cohesion between clauses and sentences and by unclear pronoun references." Andreasen 1984

SAPS Score	Frequency	Percentage
0	493	66.9
1	13	1.8
2	109	14.8
3	66	9.0
4	34	4.6
5	3	.4
Total	718	97.4
Missing data	19	2.6
Total	737	100.0

Tangentiality

"Replying to a question in an oblique, tangential or even irrelevant manner. The reply may be related to the question in some distant way. Or the reply may be unrelated and seem totally irrelevant. In the past tangentiality has sometimes been used as roughly equivalent to loose associations or derailment. The concept of tangentiality has been partially redefined so that it refers only to answers to questions and not to transitions in spontaneous speech." Andreasen 1984

SAPS Score	Frequency	Percentage
0	503	68.2
1	12	1.6
2	103	14.0
3	65	8.8
4	32	4.3
5	3	.4

Total	718	97.4
Missing data	19	2.6
Total	737	100.0

Incoherence (Word Salad, Schizophasia)

"A pattern of speech which is essentially incomprehensible at times. Incoherence is often accompanied by derailment. It differs from derailment in that in incoherence the abnormality occurs within the level of the sentence or clause, which contains words or phrases that are joined incoherently. The abnormality in derailment involves unclear or confusing connections between larger units, such as sentences or clauses. This type of language disorder is relatively rare. When it occurs, it tends to be severe or extreme, and mild forms are quite uncommon. It may sound quite similar to Wernicke's aphasia or jargon aphasia, and in these cases the disorder should only be called incoherence when history and laboratory data exclude the possibility of a past stroke, and formal testing for aphasia is negative.

Exclusions: Mildly ungrammatical constructions or idiomatic usages characteristic of particular regional or ethnic backgrounds, lack of education, or low intelligence." Andreasen 1984

SAPS Score	Frequency	Percentage
0	529	71.8
1	25	3.4
2	78	10.6
3	43	5.8
4	34	4.6
5	9	1.2
Total	718	97.4
Missing data	19	2.6
Total	737	100.0

Illogicality

"A pattern of speech in which conclusions are reached which do not follow logically. This may take the form of non-sequiturs (= it does not follow), in which the subject makes a logical inference between two

clauses which is unwarranted or illogical. It may take the form of faulty inductive inferences. It may also take the form of reaching conclusions based on faulty premises without any actual delusional thinking. Exclusions: Illogicality may either lead to or result from delusional beliefs. When illogical thinking occurs within the context of a delusional system, it should be subsumed under the concept of delusions and not considered a separate phenomenon representing a different type of thinking disorder. Illogical thinking which is clearly due to cultural or religious values or to intellectual deficit should also be excluded.''

Andreasen 1984

SAPS Score	Frequency	Percentage
0	549	74.5
1	24	3.3
2	79	10.7
3	38	5.2
4	24	3.3
5	4	.5
Total	718	97.4
Missing data	19	2.6
Total	737	100.0

Circumstantiality

"A pattern of speech which is very indirect and delayed in reaching its goal idea. In the process of explaining something, the speaker brings in many tedious details and sometimes makes parenthetical remarks. Circumstantial replies or statements may last for many minutes if the speaker is not interrupted and urged to get to the point. Interviewers will often recognize circumstantiality on the basis of needing to interrupt the speaker in order to complete the process of history taking within an allotted time. When not called circumstantial, these people are often referred to as "long-winded." Andreasen 1984

SAPS Score	Frequency	Percentage
0	541	73.4
1	14	1.9
2	82	11.1
3	48	6.5
4	27	3.7

5	6	.8
Total	718	97.4
Missing data	19	2.6
Total	737	100.0

Pressure of Speech

"An increase in the amount of spontaneous speech as compared to what is considered ordinary or socially customary. The subject talks rapidly and is difficult to interrupt. Some sentences may be left uncompleted because of eagerness to get on to a new idea. Simple questions which could be answered in only a few words or sentences are answered at great length so that the answer takes minutes rather than seconds and indeed may not stop at all if the speaker is not interrupted. Even when interrupted, the speaker often continues to talk. Speech tends to be loud and emphatic. Sometimes speakers with severe pressure will talk without any social stimulation and talk even though no one is listening. When subjects are receiving phenothiazines or lithium, their speech is often slowed down by medication, and

then it can be judged only on the basis of amount, volume, and social appropriateness. If a quantitative measure is applied to the rate of speech, then a rate greater than 150 words per minute is usually considered rapid or pressured. This disorder may be accompanied by derailment, tangentiality, or incoherence, but it is distinct from them." Andreasen 1984

SAPS Score	Frequency	Percentage
0	663	90.0
1	20	2.7
2	19	2.6
3	14	1.9
4	2	.3
Total	718	97.4
Missing data	19	2.6
Total	737	100.0

Distractible Speech

"During the course of a discussion or interview, the subject stops talking in the middle of a sentence or idea and changes the subject in response to a nearby stimulus, such as an object on a desk, the interviewer's clothing or appearance, etc." Andreasen 1984

SAPS Score	Frequency	Percentage
0	663	90.0
1	20	2.7
2	19	2.6
3	14	1.9
4	2	.3
Total	718	97.4
Missing data	19	2.6
Total	737	100.0

Clanging

"A pattern of speech in which sounds rather than meaningful relationships apear to govern word choice, so that the intelligibility of the speech is impaired and redundant words are introduced. In addition to rhyming relationships, this pattern of speech may also include punning associations, so that a word similar in sound brings in a new thought." Andreasen 1984

SAPS Score	Frequency	Percentage
0	710	96.3
2	5	.7
3	3	.4
Total	718	97.4
Missing data	19	2.6
Total	737	100.0

Global Thought Disorder Score

SAPS Score	Frequency	Percentage
0	420	57.0
1	21	2.8

2	139	18.9
3	84	11.4
4	45	6.1
5	9	1.2
Total	718	97.4
Missing data	19	2.6
Total	737	100.0

Negative Symptoms

SAPS Score	Frequency	Percentage
0	88	11.9
1	43	5.8
2	310	42.1
3	176	23.9
4	96	13.0
5	17	2.3
Total	730	99.1
Missing data	7	.9
Total	737	100.0

The individual items of the affective flattening component of the SANS are reported below (definition of each item is provided)

Unchanging Facial Expression

"The subject's face appears wooden, mechanical, frozen. It does not change expression, or changes less than normally expected, as the emotional content of discourse changes. Since phenothiazines may partially mimic this effect, the interviewer should be careful to note whether or not the subject is on medication, but should not try to "correct" the rating accordingly." Andreasen 1984

SAPS Score	Frequency	Percentage
0	110	14.9
1	43	5.8

2	271	36.8
3	198	26.9
4	83	11.3
5	10	1.4
Total	715	97.0
Missing data	22	3.0
Total	737	100.0

Decreased Spontaneous Movements

"The subject sits quietly throughout the interview and shows few or no spontaneous movements. He does not shift position, move his legs, move his hands, etc., or does so less than normally expected." Andreasen 1984

SAPS Score	Frequency	Percentage
0	279	37.9
1	36	4.9
2	233	31.6
3	134	18.2
4	40	5.4
5	8	1.1
Total	730	99.1
Missing data	7	.9
Total	737	100.0

Paucity of Expressive Gestures

"The subject does not use his body as an aid in expressing his ideas, through such means as hand gestures, sitting forward in his chair when intent on a subject, leaning back when relaxed, etc. This may occur in addition to decreased spontaneous movements." Andreasen 1984

SAPS Score	Frequency	Percentage
0	222	30.1
1	44	6.0

2	240	32.6
3	150	20.4
4	62	8.4
5	11	1.5
Total	729	98.9
Missing data	8	1.1
Total	737	100.0

Poor Eye Contact

"The subject avoids looking at others or using his eyes as an aid in expression. He appears to be staring into space even when he is talking." Andreasen 1984

SAPS Score	Frequency	Percentage
0	282	38.3
1	49	6.6
2	252	34.2
3	94	12.8
4	42	5.7
5	10	1.4
Total	729	98.9
Missing data	8	1.1
Total	737	100.0

Affective Nonresponsivity

"Failure to smile or laugh when prompted may be tested by smiling or joking in a way which would usually elicit a smile from a normal individual. The examiner may also ask, "Have you forgotten how to smile?" while smiling himself." Andreasen 1984

SAPS Score	Frequency	Percentage
0	251	34.1
1	49	6.6

2	237	32.2
3	122	16.6
4	58	7.9
5	12	1.6
Total	729	98.9
Missing data	8	1.1
Total	737	100.0

Inappropriate Affect

"Affect expressed is inappropriate or incongruous, not simply flat or blunted. Most

typically, this manifestation of affective disturbance takes the form of smiling or

assuming a silly facial expression while talking about a serious or sad subject. (Occasionally subjects may

smile or laugh when talking about a serious subject which they find uncomfortable or embarrassing.

Although their smiling may seem inappropriate, it is due to anxiety and therefore should not be rated as

inappropriate affect.) Do not rate affective flattening or blunting as inappropriate." Andreasen 1984

SAPS Score	Frequency	Percentage
0	559	75.8
1	35	4.7
2	68	9.2
3	38	5.2
4	25	3.4
5	3	.4
Total	728	98.8
Missing data	9	1.2
Total	737	100.0

Page | 247

Lack of Vocal Inflections

"While speaking the subject fails to show normal vocal emphasis patterns. Speech has a monotonic quality, and important words are not emphasized through changes in pitch or volume. Subject also may fail to change volume with changes of subject so that he does not drop his voice when discussing private topics nor raise it as he discusses things which are exciting or for which louder speech might be appropriate."

Andreasen 1984

SAPS Score	Frequency	Percentage
0	223	30.3
1	52	7.1
2	271	36.8
3	124	16.8
4	45	6.1
5	10	1.4
Total	725	98.4
Missing data	12	1.6
Total	737	100.0

Alogia

Alogia was noted in 66.5% and most individuals scored as mildly affected on this item.

SAPS Score	Frequency	Percentage
0	247	33.5
1	36	4.9
2	219	29.7
3	118	16.0
4	82	11.1
5	27	3.7
Total	729	98.9
Missing data	8	1.1
Total	737	100.0

Poverty of Speech

"Restriction in the amount of spontaneous speech, so that replies to questions tend to be brief, concrete, and unelaborated. Unprompted additional information is rarely

provided. Replies may be monosyllabic, and some questions may be left unanswered

altogether. When confronted with this speech pattern, the interviewer may find himself frequently prompting

the subject in order to encourage elaboration of replies. To elicit this finding, the examiner must allow the

subject adequate time to answer and to elaborate his answer." Andreasen 1984

SAPS Score	Frequency	Percentage
0	336	45.6
1	36	4.9
2	190	25.8
3	91	12.3
4	58	7.9
5	18	2.4
Total	729	98.9
Missing data	8	1.1
Total	737	100.0

Poverty of Content of Speech

"Although replies are long enough so that speech is adequate in amount, it conveys little information. Language tends to be vague, often over-abstract or over-concrete, repetitive, and stereotyped. The interviewer may recognize this finding by observing that the subject has spoken at some length but has not given adequate information to answer the question. Alternatively, the subject may provide enough information, but require many words to do so, so that a lengthy reply can be summarized in a sentence or two. Sometimes the interviewer may characterize the speech as "empty

philosophizing." Exclusions: This finding differs from circumstantiality in that the circumstantial subject tends to provide a wealth of detail." Andreasen 1984

SAPS Score	Frequency	Percentage
0	261	35.4
1	42	5.7
2	160	21.7
3	144	19.5
4	89	12.1
5	31	4.2
Total	727	98.6
Missing data	10	1.4
Total	737	100.0

Thought Blocking

"Interruption of a train of speech before a thought or idea has been completed. After a period of silence which may last from a few seconds to minutes, the person indicates that she/he cannot recall what he had been saying or meant to say. Blocking should only be judged to be present if a person voluntarily describes losing his thought or

if, upon questioning by the interviewer, the person indicates that that was the reason for pausing."

Andreasen 1984

SAPS Score	Frequency	Percentage
0	689	93.5
1	14	1.9
2	12	1.6
3	2	.3
4	4	.5
5	1	.1
Total	722	98.0
Missing data	15	2.0
Total	737	100.0

Increased Latency of Response

"The subject takes a longer time to reply to questions than is usually considered normal. He may seem "distant" and sometimes the examiner may wonder if he has even heard the question. Prompting usually indicates that the subject is aware of the question, but has been having difficulty in formulating his thoughts in order to make an appropriate reply." Andreasen 1984

SAPS Score	Frequency	Percentage
0	577	78.3
1	33	4.5
2	65	8.8
3	26	3.5
4	18	2.4
5	2	.3
Total	721	97.8
Missing data	16	2.2
Total	737	100.0

Grooming and Hygiene

"The subject displays less attention to grooming and hygiene than normal. Clothing may appear sloppy, outdated, or soiled. The subject may bathe infrequently and not care for hair, nails, or teeth--leading to such manifestations as greasy or uncombed hair, dirty hands, body odor, or unclean teeth and bad breath. Overall, the appearance is dilapidated and disheveled. In extreme cases, the subject may even have poor toilet habits." Andreasen 1984

SAPS Score	Frequency	Percentage
0	577	78.3
1	33	4.5
2	65	8.8
3	26	3.5
4	18	2.4
5	2	.3
Total	721	97.8

Missing data	16	2.2
Total	737	100.0

Impersistence at Work or School

"The subject has had difficulty in seeking or maintaining employment (or schoolwork) as appropriate for his or her age and sex. If a student, he/she does not do homework and may even fail to attend class. Grades will tend to reflect this. If a college student, there may be a pattern of registering for courses, but having to drop several or all of them before the semester is completed. If of working age, the subject may have found it difficult to work at a job because of inability to persist in completing tasks and apparent irresponsibility. He may go to work irregularly, wander away early, complete them in a disorganized manner. He may simply sit around the house and not seek any employment or seek it only in an infrequent and desultory manner. If a housewife or retired person, the subject may fail to complete chores, such as shopping or cleaning, or complete them in an apparently careless and half-hearted way." Andreasen 1984

SAPS Score	Frequency	Percentage
0	97	13.2
1	24	3.3
2	217	29.4
3	220	29.9
4	135	18.3
5	32	4.3
Total	725	98.4
Missing data	12	1.6
Total	737	100.0

Physical Anergia

"The subject tends to be physically inert. He may sit in a chair for hours at a time and not initiate any spontaneous activity. If encouraged to become involved in an activity, he may participate only briefly and then wander away or disengage himself and return to sitting alone. He may spend large amounts of time in

some relatively mindless and physically inactive task such as watching TV or playing solitaire. His family

may report that he spends most of his time at home "doing nothing except sitting around". Either at home

or in an inpatient setting he may spend much of his time sitting in his room." Andreasen 1984

SAPS Score	Frequency	Percentage
0	182	24.7
1	59	8.0
2	208	28.2
3	175	23.7
4	89	12.1
5	14	1.9
Total	727	98.6
Missing data	10	1.4
Total	737	100.0

Global Avolition Score

SAPS Score	Frequency	Percentage
0	69	9.4
1	32	4.3
2	274	37.2
3	223	30.3
4	110	14.9
5	22	3.0
Total	730	99.1
Missing data	7	.9
Total	737	100.0

Recreational Interests and Activities

"The subject may have few or no interests, activities, or hobbies. Although this symptom may begin

insidiously or slowly, there will usually be some obvious decline from an earlier level of interest and

activity. Subjects with relatively milder loss of interest will engage in some activities which are passive or

non-demanding, such as watching TV, or will show only occasional or sporadic interest. Subjects with the

Page | 253

most extreme loss will appear to have a complete and intractible inability to become involved in or enjoy activities. The rating in this area should take both the quality and quantity of recreational interests into account." Andreasen 1984

SAPS Score	Frequency	Percentage
0	66	9.0
1	24	3.3
2	233	31.6
3	226	30.7
4	151	20.5
5	20	2.7
Total	720	97.7
Missing data	17	2.3
Total	737	100.0

Sexual Interest and Activity

"The subject may show a decrement in sexual interest and activity, as judged by what would be normal for the subject's age and marital status. Individuals who are married may manifest disinterest in sex or may engage in intercourse only at the partner's request. In extreme cases, the subject may not engage in any sex at all. Single subjects may go for long periods of time without sexual involvement and make no effort to satisfy this drive. Whether married or single, they may report that they subjectively feel only minimal sex drive or that they take little enjoyment in sexual intercourse or in masturbatory activity even when they engage in it." Andreasen 1984

SAPS Score	Frequency	Percentage
0	91	12.3
1	32	4.3
2	92	12.5
3	137	18.6
4	97	13.2
5	17	2.3
Total	466	63.2

Missing data	271	36.8
Total	737	100.0

Ability to Feel Intimacy and Closeness

"The subject may display an inability to form close and intimate relationships of a type appropriate for his age, sex, and family status. In the case of a younger person, this area should be rated in terms of relationships with the opposite sex and with parents and siblings. In the case of an older person who is married, the relationship with spouse and with children should be evaluated, while older unmarried individuals should be judged in terms of relationships with the opposite sex and any family members who live nearby. Subjects may display few or no feelings of affection to available family members. Or they may have arranged their lives so that they are completely isolated from any intimate relationships, living alone and making no effort to initiate contacts with family or members of the opposite sex."

Andreasen 1984

SAPS Score	Frequency	Percentage
0	89	12.1
1	53	7.2
2	212	28.8
3	203	27.5
4	134	18.2
5	18	2.4
Total	709	96.2
Missing data	28	3.8
Total	737	100.0

Relationships with Friends and Peers

"Subjects may also be relatively restricted in their relationships with friends and peers of either sex. They may have few or no friends, make little or no effort to develop such relationships, and choose to spend all or most of their time alone." Andreasen 1984

SAPS Score	Frequency	Percentage
0	65	8.8
1	37	5.0
2	238	32.3
3	224	30.4
4	136	18.5
5	24	3.3
Total	724	98.2
Missing data	13	1.8
Total	737	100.0

Global Anhedonia Score

SAPS Score	Frequency	Percentage
0	40	5.4
1	22	3.0
2	241	32.7
3	251	34.1
4	148	20.1
5	25	3.4
Total	727	98.6
Missing data	10	1.4
Total	737	100.0

Social Inattentiveness

"While involved in social situations or activities, the subject appears inattentive. He looks away during conversations, does not pick up the topic during a discussion, or appears uninvolved or unengaged. He may abruptly terminate a discussion or a task without any apparent reason. He may seem "spacy" or "out of it". He may seem to have poor concentration when playing games, reading, or watching TV." Andreasen 1984

SAPS Score	Frequency	Percentage
0	539	73.1
1	41	5.6
2	90	12.2
3	37	5.0
4	12	1.6
5	10	1.4
Total	729	98.9
Missing data	8	1.1
Total	737	100.0

Appendix 5.2: Contingency table for 8 variable model; n=116; cells with missing information are also included in the model

Obser-vation	Eye contact	Auditory hallucinations	Global hallucination score	Global delusions score	Grooming	Affective non-responsiveness	Spontaneous Movement	Commenting voices	Count	Percentage
1	.	.	.	.	.	.	.	.	7	.
2	.	.	.	.	2	2	2	.	1	.
3	1	.	.	.	1	1	1	.	2	.
4	1	.	.	.	1	2	2	.	1	.
5	1	.	.	.	2	1	1	.	1	.
6	1	.	.	1	1	1	1	.	1	.
7	1	.	.	1	2	1	2	.	1	.
8	1	.	.	2	2	1	1	.	1	.
9	1	.	1	1	1	1	1	1	1	.
10	1	1	1	1	1	1	1	1	82	11.7816
11	1	1	1	1	1	1	2	1	3	0.4310
12	1	1	1	1	1	2	1	1	6	0.8621
13	1	1	1	1	1	2	2	1	3	0.4310
14	1	1	1	1	2	1	1	1	24	3.4483
15	1	1	1	1	2	1	2	1	7	1.0057
16	1	1	1	1	2	2	1	1	4	0.5747
17	1	1	1	1	2	2	2	1	6	0.8621
18	1	1	1	2	1	1	1	1	19	2.7299
19	1	1	1	2	1	1	2	1	2	0.2874
20	1	1	1	2	1	2	1	1	2	0.2874
21	1	1	1	2	1	2	2	1	1	0.1437
22	1	1	1	2	2	1	1	1	6	0.8621
23	1	1	1	2	2	2	1	1	2	0.2874
24	1	1	2	2	1	1	1	1	3	0.4310
25	1	1	2	2	1	1	2	1	2	0.2874
26	1	1	2	2	2	1	2	1	1	0.1437
27	1	2	.	.	2	1	1	.	1	.
28	1	2	.	.	2	2	2	2	1	.
29	1	2	.	2	2	1	1	2	1	.
30	1	2	2	.	1	1	1	2	1	.
31	1	2	2	1	1	1	1	1	3	0.4310
32	1	2	2	1	1	1	1	2	2	0.2874
33	1	2	2	1	1	1	2	1	1	0.1437
34	1	2	2	1	1	1	2	2	2	0.2874
35	1	2	2	1	1	2	1	1	1	0.1437
36	1	2	2	1	1	2	2	1	1	0.1437
37	1	2	2	1	1	2	2	2	1	0.1437
38	1	2	2	1	2	1	1	1	1	0.1437
39	1	2	2	1	2	1	1	2	4	0.5747
40	1	2	2	1	2	2	1	2	1	0.1437
41	1	2	2	1	2	2	2	2	1	0.1437
42	1	2	2	2	1	1	1	1	9	1.2931
43	1	2	2	2	1	1	1	2	31	4.4540
44	1	2	2	2	1	1	2	1	1	0.1437

Obser-vation	Eye contact	Auditory hallucinations	Global hallucination score	Global delusions score	Grooming	Affective non-responsiveness	Spontaneous Movement	Commenting voices	Count	Percentage
45	1	2	2	2	1	1	2	2	2	0.2874
46	1	2	2	2	1	2	1	1	2	0.2874
47	1	2	2	2	1	2	1	2	1	0.1437
48	1	2	2	2	1	2	2	1	2	0.2874
49	1	2	2	2	1	2	2	2	5	0.7184
50	1	2	2	2	2	1	1	1	8	1.1494
51	1	2	2	2	2	1	1	2	17	2.4425
52	1	2	2	2	2	1	2	2	5	0.7184
53	1	2	2	2	2	2	1	.	2	.
54	1	2	2	2	2	2	1	1	8	1.1494
55	1	2	2	2	2	2	1	2	7	1.0057
56	1	2	2	2	2	2	2	1	12	1.7241
57	1	2	2	2	2	2	2	2	16	2.2989
58	2	.	.	.	1	2	2	.	1	.
59	2	.	.	.	2	1	2	.	1	.
60	2	.	.	.	2	2	2	.	2	.
61	2	1	1	1	1	1	1	1	5	0.7184
62	2	1	1	1	1	1	2	1	3	0.4310
63	2	1	1	1	1	2	1	1	1	0.1437
64	2	1	1	1	1	2	2	.	1	.
65	2	1	1	1	1	2	2	1	40	5.7471
66	2	1	1	1	2	1	1	1	4	0.5747
67	2	1	1	1	2	1	2	1	6	0.8621
68	2	1	1	1	2	2	1	1	4	0.5747
69	2	1	1	1	2	2	2	1	80	11.4943
70	2	1	1	2	1	1	1	1	4	0.5747
71	2	1	1	2	1	1	2	1	2	0.2874
72	2	1	1	2	1	2	2	1	7	1.0057
73	2	1	1	2	2	1	2	1	1	0.1437
74	2	1	1	2	2	2	1	1	4	0.5747
75	2	1	1	2	2	2	2	1	19	2.7299
76	2	1	2	1	1	2	2	1	1	0.1437
77	2	1	2	1	2	1	1	1	1	0.1437
78	2	1	2	1	2	2	2	1	2	0.2874
79	2	1	2	2	1	2	2	1	1	0.1437
80	2	1	2	2	2	2	2	1	5	0.7184
81	2	2	.	.	2	2	2	.	1	.
82	2	2	.	1	1	1	2	2	1	.
83	2	2	1	1	1	1	1	1	1	0.1437
84	2	2	1	2	2	2	1	1	1	0.1437
85	2	2	1	2	2	2	1	2	1	0.1437
86	2	2	1	2	2	2	2	1	1	0.1437
87	2	2	2	.	2	2	1	1	1	.
88	2	2	2	.	2	2	2	.	2	.
89	2	2	2	.	2	2	2	1	2	.
90	2	2	2	1	1	1	1	2	2	0.2874
91	2	2	2	1	1	2	2	1	1	0.1437
92	2	2	2	1	1	2	2	2	2	0.2874

Obser-vation	Eye contact	Auditory hallucinations	Global hallucination score	Global delusions score	Grooming	Affective non-responsiveness	Spontaneous Movement	Commenting voices	Count	Percentage
93	2	2	2	1	2	1	1	2	1	0.1437
94	2	2	2	1	2	1	2	2	2	0.2874
95	2	2	2	1	2	2	1	1	1	0.1437
96	2	2	2	1	2	2	2	.	1	.
97	2	2	2	1	2	2	2	1	10	1.4368
98	2	2	2	1	2	2	2	2	8	1.1494
99	2	2	2	2	1	.	1	2	1	.
100	2	2	2	2	1	1	1	1	1	0.1437
101	2	2	2	2	1	1	1	2	4	0.5747
102	2	2	2	2	1	1	2	2	2	0.2874
103	2	2	2	2	1	2	1	1	2	0.2874
104	2	2	2	2	1	2	1	2	2	0.2874
105	2	2	2	2	1	2	2	1	7	1.0057
106	2	2	2	2	1	2	2	2	11	1.5805
107	2	2	2	2	2	1	1	1	3	0.4310
108	2	2	2	2	2	1	1	2	7	1.0057
109	2	2	2	2	2	1	2	.	1	.
110	2	2	2	2	2	1	2	1	2	0.2874
111	2	2	2	2	2	1	2	2	1	0.1437
112	2	2	2	2	2	2	1	1	4	0.5747
113	2	2	2	2	2	2	1	2	6	0.8621
114	2	2	2	2	2	2	2	.	4	.
115	2	2	2	2	2	2	2	1	40	5.7471
116	2	2	2	2	2	2	2	2	66	9.4828

The cell with 82 observations (11.7%) is the combination with very little/no symptoms.
The cell with 80 observations (11.5%) is the combination with all negative symptoms.
The cell with 66 observations (9.5%) is the combination with all positive symptoms.
The other 114 cells are observations with the other combinations of symptoms.