Mathematical Statistics
Stockholm University

Socio-Economic Risk Factors for Schizophrenia and Bipolar Disorder: a Swedish Register-Based Study

Anna Levinsson

Examensarbete 2010:5

Postal address:

Mathematical Statistics
Dept. of Mathematics
Stockholm University
SE-106 91 Stockholm
Sweden

Internet:

http://www.math.su.se/matstat

Mathematical Statistics
Stockholm University
Examensarbete **2010:5**,
http://www.math.su.se/matstat

Socio-Economic Risk Factors for Schizophrenia and Bipolar Disorder: a Swedish Register-Based Study

Anna Levinsson*

September 2010

Abstract

Previous research has shown that there is an association between social-economical status in childhood and the development of schizophrenia. The pattern for bipolar disorder is less clear. This study aims at investigating socio-economic risk factors for schizophrenia and bipolar disorder and to compare the results to identify common risk factors for the two disorders. Method: We used a study cohort based on individuals in the Swedish Population Registry born in 1973-1985 and identified cases using the National Patient Registry. Data on variables was linked from other registers by using the individualswith the Multi-generations Registry to link biological father and mother to the subject. Data was analysed using logistic regression in SAS and we created univariate models adjusted for gender and birth year as well as multivariate models. Results: In the multivariable models risk factors birth year, gender (male vs. female), housing, urbanicity, immigrant status and parental history of disorder were all significant over-risks for schizophrenia. For bipolar, birth year, housing and one of the patient's parents being born outside of Sweden were found to be significant over-risks while gender (male vs. female) was found to be an under-risk. 2nd generation immigrant status was not significant for bipolar disorder. Conclusion: Due to the results from the analyses of the multivariate models, birth year, gender, housing and one of parents born abroad were significant risk factors common to both disorders. Discussion: More research, especially for bipolar disorder, is called for, preferably in the form of prospective longitudinal studies so as to be able to assign causality. Also, the gene-environment interaction aspects of effect need to be investigated further.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: anna.levinsson@amm.gu.se. Supervisor: Åke Svensson.

Contents

Preface

The following is a 30 credits magister-level thesis, written at the Division of Mathematical Statistics, Department of Mathematics, Stockholm University in cooperation with the Division of Public Health Epidemiology, Department of Public Health Sciences, Karolinska Institutet.

I would like to thank my supervisors; Susanne Wicks from Karolinska Institutet who has provided data material for my study and encouraged me during times of SAS despair…I could not have done it without you. And Åke Svensson at Stockholm University, who has kept an eye on the mathematical statistics purpose of this project, when I have tended to stray too far into the medical parts of psychiatric epidemiology.

Introduction

A research topic that still needs a lot of research is the aetiology of psychiatric disorders. Theories are introduced, supported and discarded as research takes new approaches, new angles and develops new methods. We no longer believe that schizophrenia is caused by evil spirits or that lobotomy is an efficient cure.

In basic terms, schizophrenia and bipolar disorder are two major mental disorders which often have a severe impact on the patient‟s life. *Schizophrenia* is a severe mental disorder, where the patient may suffer from psychoses with symptoms like: episodes of loss of contact with reality, auditory or visual hallucinations or delusions, lack of motivation/drive, and poverty of speech. *Bipolar disorder* is an „affective disorder‟, that is, a mental disorder that involves the patient‟s mood and energy levels, driving her into spells of depression and mania. Bipolar patients can also suffer from psychoses, but not every patient does.

Several etiological theories have been put forward and schizophrenia is most likely a multi-factorial disorder where both genetic and environmental factors may be relevant. Two theories addressing the association between social status and mental disorder are the „social causation theory‟ and the „social drift theory‟. The former argue that low economic status in the neighbourhood, for parents or for patient is the cause of the disorder, while the latter make a case for that “people about to develop schizophrenia lose social status and move into areas of relative socioeconomic deprivation” [25]. These two theories thus represent causality in two opposite directions.

Some environmental risk factors sort under the category socioeconomic risk factors, e.g. social class, living standard, social inequality, birth country, social conditions and clefts, as well as education level, employment, income and social change. There are numerous sublevels to these summarising terms, but we will limit our study to the more manageable number of main categories and try to determine whether our results correspond with the published international results.

In order to investigate etiological factors, the exposure must occur before the outcome. Otherwise, the results will very likely be influenced by the consequences of the disorder. An individual who has been in and out from inpatient psychiatric care since his early twenties is likely to find it more difficult to keep a job, achieve higher education or keep a place to live. The possible effect of social drift thus blurs the validity of findings that lower educational level of subject or parents lead to higher risk. The phenomenon is well known for psychiatric disorders in general and especially for the major ones. An attempt to circumvent this problem is to assess the childhood conditions for the individual by looking at the socioeconomic status of the parents during the child’s early years.

Earlier research demonstrates an association between socio-economic factors in childhood and risk of later developing schizophrenia, leading to admission to inpatient psychiatric care. The pattern is less clear for bipolar disorder. As we have the data for it, we will use the same underlying material for investigating the risk factors (immigrant status, form of residence when aged between 1 and 5, urbanicity (at birth) and parental history of disorder) in both study groups and compare the results.

We will start our study with a literary review to seek an answer to: What results do we have from research on environmental factors, especially socioeconomic risk factors, for schizophrenia and bipolar disorder? And turning to the study questions: What socioeconomic risk factors are significant for schizophrenia and bipolar disorder onset in Sweden? Do the results correspond to the ones from literature? Are they similar for the two disorders?

Literary review

Current research results on risk factors for Bipolar disorder

It was considerably more difficult to find articles on risk factors for bipolar disorder than for schizophrenia. The published research that was found said overall that results for bipolar disorder were inconclusive and called for prospective longitudinal studies. However, several studies received results pointing in a certain direction.

In "Exposure to obstetric complications and subsequent development of bipolar disorder" (Scott et al. 2006), 22 articles are reviewed, resulting in a pooled odds ratio (OR) of 1.15 (95% CI 0.62 – 2.14) for exposure to obstetric complications, that is, a non-significant result. The value is yielded from a comparison of bipolar patients to healthy controls, comparison with schizophrenia patients yields a pooled OR of 0.61 (95% CI 0.39 – 0.95), which is significant, as is low birth weight (<2.5 kg), which increases the risk for BPD by 16%.

Strakowski & DelBello (2000) reports a startling result from a study by the National Institute of Mental Health: over 60% of Bipolar Type I-patients suffers from substance abuse. Substance use syndrome is also higher represented in those patients even at first admission. There is a trend in the other direction as well: patients with substance abuse have a 5 to 8 times elevated rate of BPD. Focusing on the hypothesis that substance abuse causes BPD; the authors note that BPD patients with preceding alcohol abuse developed the disorder later in life and that a possible explanation is that a non-predisposed individual needs several years of alcohol abuse to precipitate illness onset. Also, patients with both alcohol abuse and BPD had BPD in the family to a lesser degree than patients without alcohol abuse. However, if substance abuse causes BPD, sobriety should reverse it. The idea is supported by a study by

Winokur et al. (1995). On the other hand, a study by Strakowski et al. (1998), showed 54% of the patients with alcohol use preceding the onset of BPD to have an affective episode during a sober period, during the 12 months following the hospitalisation for first-episode mania. 39% of patients with other drug abuse did the same. Experiments with rats suggest that once behavioural sensitization has occurred there is a permanent change in the central nervous system. Also, in sensitive individuals the substance may serve as a stressor. There is likely a heterogenic aetiology of the two disorders, but it is noteworthy that demographic data differ between patient types.

Bale (2006) considers the effects of maladaptive coping strategies on the risk of BPD. She states in paragraph two of the introduction that "... females are diagnosed with affective disorders far more often than males ..." [6] but provides no source for this statement. Basically, dysregulations in parts of the brain lead to increased stress sensitivity and found the basis for maladaptive coping strategies. The term „maladaptive coping strategies" denotes that behavioural response is too strong, too far towards either end of the spectrum, which leads to

disorder predisposition. Her conclusion is that stress pathway regulation is critical for the development of mood disorders.

Another article (Post & Leverich 2006) reasons around the concept of exposure to psychosocial stressors at an early age as relevant to BPD onset. There is a probable confounding effect between the childhood stressors of abuse and a positive family history of mood disorders. They also report that loss of a mother, father or sibling by suicide is a significant stressor and that suicide by other relatives and family members are smaller risk factors.

The next article on risk factors for BPD discusses the psychosocial factors: environmental, cognitive and developmental (Alloy et al 2005). The factors stressful life events, social support and expressed emotions (EE) are assigned to factor type „environmental". Stressful life events may have a part in increased risk of onset and relapse of BPD. Especially negative events might not surprisingly trigger depressive episodes and contribute to manic or hypomanic (a milder form of mania, less destructive and without psychosis) episodes. Research on positive events is very limited. There are indications that lack of social support and low EE in the family leads to a more severe course of illness; longer recovery time and higher risk of recurrence. Alloy et al goes on to review research on cognitive styles and its connections to BPD. Above we have summarised an article by Bale (2006) on maladaptive coping strategies and Alloy, one year earlier, discuss the same type of reactions to negative life events, but without stepping into the neuroscientific territory of reasoning. In summary, Alloy et al find that such coping strategies lead to negative appraisal (evaluation) of negative events, resulting in hopelessness and a negative view of the self and thus starting a negative trend. However, bipolar patients may also have a very positive appraisal scheme, which leads to mania. That is, the bipolar patient swings back and forth from overestimating and underestimating oneself, the grey zone in between can be very short, making it difficult for the patient and her surroundings to cooperate continuously.

Tsuchiya et al (2003) performed a systematic review of studies regarding risk factors for bipolar disorder (BPD). They found no difference in risk between the genders but noted inconsistent findings regarding ethnicity and that they may be due to migration. Further for the socioeconomic risk factors, studies pointed towards an association between single marital status, unemployment and lower income to BPD incidence. However, this is the patient"s own situation after diagnosis and the results are likely influenced by social drift. Studies also point towards an increased risk for BPD when suffering from brain injuries before age 10.

In 2004, Tsuchiya et al conducted another study, this time on whether higher economic status of parents can increase the risk for bipolar disorder for the child. They used two Danish registers, chose all individuals born in 1960 or later and those with a first-time contact with Danish psychiatric care or admission to psychiatric inpatient care during 1981-1998. To each case, 50 controls were matched. This resulted in 947 cases and 47"350 controls. The effects of marital status, occupation, education, income and wealth of parents and subjects were estimated by conditional logistic regression. The results showed that for subjects, single marital status, receiving social assistance, pension or sickness payments, unemployment, shorter educational history and lower income were associated with high risk for bipolar

disorders. For parental factors however, higher education and higher level of wealth increased the risk for subjects. After being adjusted for gender, family history of psychiatric diagnoses and other socio-economic variables, the associations were still significant. Their conclusion: the associations between lower socio-economic status in subjects and increased risk may be a consequence of the disorder.

Our last article is by Swinnen & Selten (2007), on the subject of migration and mood disorders. They start from the platform that migration is a risk factor for schizophrenia. Can this be extended to BPD? They used population-based incidence studies of mood disorders in immigrant groups and calculated age and gender adjusted mean relative risks (RR) using Poisson regression analysis. They found the overall result that mean RR of developing BPD was 2.47 (95% CI 1.33-4.59), a significant value, for migrants. But when they excluded a study from the UK on African-Caribbean immigrants, the result was no longer significant. Conclusion: There is no conclusive evidence for increased risk of BPD for migrants.

In summary, authors call for improved methodologies, mainly more prospective longitudinal studies. The risk factors found significant in the articles reviewed above are for the subject herself: single marital status, unemployment and related lower income, giving birth (women developing BPD within 3 months of childbirth), brain injuries before age 10, maladaptive coping strategies (increased stress sensitivity), substance abuse, stressful life events especially psychosocial stressors at an early age and positive family history of BPD. Regarding parental SES factors, studies have shown either no associations or that longer parental education, higher level of wealth, self-employed father, father as student and father as homemaker increases the risk for BPD. The SES factors for the subject have been found significant, but it is not clear how much this is due to social drift. These factors make up a great range of types of risk factors and their interactions remain to be investigated further.

Current research results on risk factors for Schizophrenia

For schizophrenia, the research has come further and the general consensus is that outbreak of schizophrenia is due to both a genetic predisposition and environmental factors. However, to what percent each owns up, has not been determined.

The article “Schizophrenia: Genes and Environment” (Tsuang 2000) investigates the nowadays generally accepted hypothesis that schizophrenia onset depends on both genetic predisposition and environmental factors. First of all, she supports the „nature and nurture‟ hypothesis by the following statement: “identical twins show average concordance of only 50%; rates of 100% would be expected on the basis of genetic equivalence alone.” [26] A theory is the model:

$V_{phenotype} = V_{genetic} + V_{environmental} + V_{error}$ ($V_p = V_g + V_e + V_{error}$) ⇔

$V_p = V_{g\ additive} + V_{g\ dominance} + V_{g\ epistatic} + V_{e\ common}\ V_{e\ unshared} + V_{g.e} + V_{error}$

The ‘V’s above stand for variance and the term heritability is the part of the phenotypic variance that can be attributed to the genetic variance in the population. Thus the model above represents the equation for phenotypic variance. Phenotype is the observed properties of an individual, the genotype is its complete genetic make-up and so what is to show is that the schizophrenia phenotype is, at least partially, caused by „schizophrenia genes‟ in its genetic

make-up. Tsuang also reports that the excess winter-spring births in female schizophrenic offspring correlate with urban birth and that research now works along with the idea of schizophrenia as a neurodevelopmental disorder rather than a neurodegenerative.

Wicks et al (2005) wrote an article on their two-generation national cohort study in Sweden, where they looked at social adversity in childhood as risk factor, indicated by living in rented apartments, low socio-economic status, single-parent households, unemployment and household receiving welfare benefits, as well as at possible confounders. The study population consisted of approximately 2.1 million children born in 1963-1983, whose variable values were collected from the Swedish National Board of Health and Welfare and from Statistics Sweden, then linked by the individuals‟ personal ID-numbers. By using Cox regression on the multivariate models, they reached results in the form of hazard ratios (HR). All social adversity indicators in childhood showed increased HRs and increased with the number of social adversity factors present, indicating a dose-response relationship. The analysis of confounders showed that urbanicity, parental inpatient care for substance abuse and/or psychosis, parental age and foreign-born parents all co-varied with social adversity. Parental inpatient care for schizophrenia had the highest HR: 8.4 (95% CI 7.1-9.9).

The article by Torrey & Yolken (1998) focuses on urbanicity and household crowding as risk factors for both schizophrenia and bipolar disorder. They review five studies on these risk factors. A Swedish study from 1992 is reported to have shown that city-raised individuals are 1.65 times more likely to develop schizophrenia. A study in England the same year showed a 14% increased risk for schizophrenia for urban births. In 1996, an American study reported a more than 5 times increased risk for schizophrenia when raised in an urban area. The fifth article is on a study in the Netherlands and reports a 10-25% increase in risk when born in a city. The article also extends this result to bipolar disorder. The authors of the review article propose a reason for these numbers; they argue that household crowding can be the culprit. Household crowding is more common in urban areas and four of the studies reviewed by Torrey & Yolken support the suggestion.

Kirkbride et al (2008) [18] and Kirkbride et al (2008) [17] investigate the possibility of an association between ethnicity and schizophrenia when adjusted for socio-economic status, using a population in East London, individuals aged 18-64 years, diagnosed at first psychotic episode by the DSM-IV tool. According to the articles, this was the first epidemiological study to simultaneously investigate the effects of age, gender and socio-economic status across ethnic groups in the United Kingdom. Their findings were that for schizophrenia, after adjustment for age, gender and socio-economic status: black Caribbean (IRR 3.1, 95% CI 2.1-4.5) and black African (IRR 2.6, 95% CI 1.8-3.8) as well as for women in groups Bangladeshi (IRR 2.3, 95% CI 1.1-4.7) and Pakistani (IRR 3.1, 95% CI 1.2-8.1), risk was elevated. A definite weakness of their study is that they used *current* socio-economic status, which may likely have been affected downwards from origin by social drift, as mentioned above. The results complement those in a Swedish study, where the analysis was performed on hospital discharge data and SES in childhood, to minimise social drift effect (Hjern et al, 2004).

A study from 2003 (Harrison et al) focused on association between psychotic disorders and urbanicity. They used data from the Swedish Inpatient Discharge register (same register as we used) and individuals born between 1973 and 1980 with linked birth and socio-economic data. They got the results hazard ratio 1.34 and 95% CI 0.91-1.96 for urban compared to rural birthplace in schizophrenia patients. No interaction was shown between sex and birthplace in schizophrenia patients, reported p-value 0.55. For a model adjusted for age, birth weight, birth length, gestational age, season of birth, age of mother, maternal parity, Caesarean section and sex, using „Birthplace rural areas‟ as reference, reported results are Main cities and suburbs HR 1.31 95% CI 0.89-1.93, for Large cities and industry HR 1.04 95% CI 0.72-1.49, both for

schizophrenia. A weakness in their study is that they used maternal education as indicator of childhood socio-economic status, which is not necessarily indicative of the socio-economic status during the child‟s upbringing, e.g. unemployment during the child‟s upbringing is not due is not restricted to mothers with low education.

Cantor-Graae and Selten (2005) reviewed the research on an association between schizophrenia and migration. They concluded that "a family history of migration is an important risk factor for schizophrenia." [9] The synthesis of previously published results were that the relative risk for first generation immigrants was 2.7 (95% CI 2.3-3.2) and for second generation immigrants 4.5 (1.5-13.1). Further, the RR for studies of both first and second generation and studies that grouped them together had RR 2.9 (2.5-3.4) and finally, comparing migrants from developed and undeveloped countries and skin colour of migrant (black vs. white and neither black nor white) yielded RRs 3.3 (2.8-3.9) and 4.8 (3.7-6.2) respectively. The Danish population study by Cantor-Graae et al (2003) on migration reached results pointing in the same direction. That study had a population-based cohort of approximately 2.14 million, who were living in Denmark on their fifteenth birthday. The following results were obtained: First- and second generation immigrants, RRs 2.45 and 1.92 with 95% CIs (2.25-2.67) and (1.74-2.12) respectively, for schizophrenia. Danish residents with a family history of migration had a RR of 1.6 (1.25-2.05). Overall concluding migration to be a significant risk factor.

In 2007, Cantor-Graae did a review article on social adversity, where she reviews a number of studies conducted from 1996 onwards and printed in English-language journals. She found significant evidence for migration being a risk factor, especially second generation immigrant status. Most studies showing these results were conducted in Europe. The overall RR for developing schizophrenia associated with migration was 2.9 (95%CI 2.5-3.4). The material she reviewed reached ambiguous results concerning social causation by urban birth and upbringing. Regarding future research, she anticipates studies on the actual mechanisms that prompt social adversity to result in psychoses. A theory that needs exploration is that of social factors causing dopamine dysregulations or sensitization.

In 2004, Hjern et al also looked at migration as a risk factor and in addition at whether social adversity contributed to the association. Their national cohort study on Swedish individuals gave the results that there is a higher risk for first and second generation immigrants compared to Swedish majority population and that those results point towards an association with social adversity. The model adjusted for age and gender and yielded RR 1.4-3.1 for first generation and 1.0-2.0 for second generation immigrants. When controlling for parental disorder history, the results did not change.

An Irish study by Kelly et al (2010) includes both a prospective study and a literary review, on the subject of urbanicity in Ireland. Their findings were consistent with the literature: being born, growing up and living in an urban area increase the risk for schizophrenia, particularly in men (incidence rate ratio (IRR) 1.92, 95% CI 1.52-2.44). For females, the IRR was 1.34, 1.00-1.80. The associations that lead to this are unknown, but they suspect that the factors are both biological and environmental, naming air pollution, cannabis and social exclusion as possible culprits.

Weiser et al (2008) elaborates on schizophrenia and migration in a study on a cohort of Israeli adolescents, of a total 661‟792, 104‟638 were immigrants. They controlled for socio-economic status obtained from CBS 1995 census data, here represented by a compound number calculated from variables such as number of people per room in household, number of computers in household, number of motor vehicles per household, education and per capita income level, and for gender. First-generation immigrant adolescents had an adjusted hazard

ratio (HR) of 1.62 (95% CI 1.1-2.22). Children of two immigrants had a HR 1.49 (1.12-2.0) and children with one immigrant and one native parent HR 1.41 (1.01-1.95). These numbers were from comparing immigrant to non-immigrant, when differentiating immigrant categories, Soviet and Ethiopia had the highest HRs, 1.55 and 2.95.

Next is a Danish study by Byrne et al published 2004; a national population-based nested case-control study based on Danish longitudinal registers. As cases, they had 7704 first admittances to in-patient care between 1981 and 1998, diagnosed with schizophrenia with tools ICD-8 or ICD-10. 192‟590 population controls were matched individually by age, gender. Socio-economic factors were measured the year before admission, putting it at risk for bias by social drift as illness onset and first admission are not necessarily in close succession.

They found that a number of factors increased the risk for schizophrenia. For the patient herself: unemployment, low educational level, being single, without children, lower wealth status, lower income, birth in urban area, birth outside of Denmark and having more than 2 siblings. Parental SES: parental lower income, parental unemployment and a family history of psychiatric disorders were also significant risk factors.

The same authors as above [5], with the addition of Bennedsen, published an article in 2007 on obstetric complications as risk factors [6]. This was also a national population-based nested case-control study based on Danish longitudinal registers, with 1039 first admissions to inpatient care with ICD-8 or ICD-10 diagnosis schizophrenia and 24‟826 individually matched population controls. The model adjusted for family psychiatric history, socio-economic and demographic factors. The factors found significant were „mother not attending prenatal check-ups‟ IRR 2.08 (95% CI 1.0-4.4), birth at <37 weeks IRR 1.51 (1.0-2.2), maternal influenza IRR 8.2 (1.4-48.8), preeclampsia IRR 2.72 (1.0-7.3), threatened premature delivery IRR 2.39 (1.4-4.1) bleeding during delivery IRR 2.43 (1.1-5.6), manual extraction of the baby IRR 2.15 (1.1-4.4) and maternal sepsis IRR 2.91 (1.1-7.9). The results showed no significant interactions between the obstetric factors and sex or family psychiatric history.

In their article from 2002, Cannon et al review the findings of an association between obstetric complications and schizophrenia. They concluded that three types of complications had a significant association with schizophrenia; pregnancy complications, issues regarding foetal growth and development and delivery complications. They came up with the following list of factors, in order of effect size: diabetes in pregnancy (OR 7.76), placental abruption (OR 4.02 but 95% CI covering 1), birth weight <2000g (OR 3.89), emergency Caesarean section (OR 3.24), congenital malformations (malformations present at birth) OR 2.35, uterine atony (uterus lacking normal muscle tone) (OR 2.29), rhesus baby (a blood disease at birth where the baby‟s red blood cells have been attacked in the uterus by the mother‟s, because the baby have blood group Rh positive and the mother Rh negative) OR 2.00, asphyxia (extreme lack of oxygen in the blood) OR 1.74, bleeding in pregnancy (OR 1.69), birth weight <2500g (OR 1.67) and on the verge of significance preeclampsia (causes increased amounts of protein in the urine and elevated blood pressure) OR 1.36 (0.99-1.85). They found heterogeneity significance for asphyxia and birth weight <2500g (p-values 0.005 and <0.02) with random effects estimates 2.01 95% CI 0.73-5.49 p-value 0.18 (asphyxia) and 1.66, 0.94-2.95 p-value 0.08. P-values were just above significance for abruptioned placenta (damaged placenta), head circumference <32cm and negative association with non-spontaneous delivery onset.

A theory by Kinney et al (2009) is that if the foetus is exposed to adverse environments during sensitive stages of growth, the immune system might be compromised and with the backing of earlier findings such as immune system abnormalities in schizophrenia individuals, the authors believe that it is born with a latent vulnerability.

Mortensen et al (2010) discuss what disorders in the family psychiatric history are risk factors for schizophrenia. They followed all 1‟745‟970 people born in Denmark between 1955 and 1991 from their 15th birthday until 2006 for a total of 28 million person years. They selected individuals with no history of suicide or psychiatric care among parents or siblings as reference category. The relative risk associated with maternal psychoses is 4.79 and for mother diagnosed with schizophrenia 8.97. A family history of general psychiatric illness was a confounder of the association between schizophrenia and urbanicity. The results suggest that there may be shared environmental factors pooling in families.

In his editorial for American Journal of Psychiatry in January 2010, Gilmore summarises what is considered known about schizophrenia aetiology and where the research is going. He reports that the most important risk factor for schizophrenia is to have a close relative with the diagnosis, but that most schizophrenic patients do not have an afflicted relative. Gene studies have shown an increase to at most 1.5/100 compared to the population rate of 1/100. Prenatal exposure to infection increases the risk to 2-4/100 and the combination of maternal depression and a parent with psychosis increased the rate to 2.6/100. Finally, having a parent with schizophrenia increases the risk to more than 9/100. He raises the issue of the effects of stress on brain development and function, stating that schizophrenia is likely "the result of an abnormal developmental trajectory of synapse and circuit formation that ultimately leads to a miss-wired brain and clinical symptoms" [12] and asks for research on three questions: when do these abnormalities form? How do the genetic and environmental risk factors alter brain development? And can we direct these trajectories at some point during brain development? As Gilmore puts it "this approach signifies a bit of a paradigm shift." [12]

To summarise the above, the risk factors for schizophrenia found significant for the subject were: history of maternal psychosis, urban birth, ethnicity different from country of residence, migration (1st and 2nd generation immigrant, immigration from undeveloped countries), pre-natal exposure to infection, obstetric complications (diabetes in pregnancy, emergency caesarean, placental abruption, birth weight lower <2000 g, birth weight lower than 2500 g, congenital malformation, uterine atony, rhesus baby, asphyxia, bleeding in pregnancy, birth at <37 weeks, preeclampsia and maternal sepsis), unemployment, low educational level, being single, being without children, lower wealth status, lower income, having more than 2 siblings. Parental SES factors: lower income and parental unemployment and family history of schizophrenia.

Risk factors common for both schizophrenia and bipolar disorder

„Common genetic determinants of schizophrenia and bipolar disorder ...‟ by Lichtenstein et al (2009) calculated relative risks and co-morbidity for different degree relatives (biological/adoptive parents, full siblings and offspring of patients with one of the disorders), regarding both schizophrenia and bipolar disorder. They used the Swedish Multi-generation register, providing information on all children and adults in Sweden, and the hospital discharge register, which contains all admissions to public psychiatric inpatient care. They caught 9‟009‟202 unique individuals in >2 million nuclear families. The analysis method used was multivariate generalised linear mixed model. The risk for schizophrenia in biological relatives of cases with BPD was consistently increased. They observed that heritability for schizophrenia and bipolar disorder were 64% and 59%, the shared environmental effects were 4.5% (4.4%-7.4%) for schizophrenia and 3.4% (2.3-6.2%) for BPD. The co-morbidity mainly due to additive genetic effects common to both disorders was 63%. Conclusion: the disorders partly share a common genetic cause.

Methods

Dataset characteristics and preparation

The final dataset was constructed by running several national registries together and organizing the data by the person‟s individual ID-number. The base dataset stems from the Swedish Population Registry and contains all individuals in Sweden born in 1973-1985. Those who died or emigrated before an age of 20 were excluded, since the study is on adult individuals. This is because the disorder most often manifests after puberty, when the subject is in her twenties. Through the Multi-generations Registry we linked ID-numbers for the individual‟s biological mother and father to the individual‟s. Due to the data on urbanicity from the Swedish Birth Registry only being available for children born in Sweden, all individuals born elsewhere were excluded.

We started out by adding the variable „diagnosed with schizophrenia‟, with data from the National Patient Registry, on all discharges from inpatient care for individuals born in 1973-1985. The first registrations in the register are from 1973 and the last from 2006, when the youngest patients are 21 years old. This set contains individual identification numbers of the individuals, which are also found in the base set. For every registered inpatient care episode (the same individual may have been admitted several times), we had the date of admission, the main diagnosis and the 7 bi-diagnoses, if applicable. This set was weeded out in SAS by choosing all entries containing the code for respective disorder. The codes in the set were from ICD-8 (used 1969-1986), ICD-9 (used 1987-1996) and ICD-10 (used from 1997 onwards). For schizophrenia the used codes were 295-29599, excluding 29550 and 29570 (ICD8), 295 without a letter, 295A-E, G, W, X (ICD9) and F20-F209 (ICD10).

The resulting set was then reworked further, selecting the first admission entry for each individual and deleting the rest. Finally, a new variable „schizo‟ was introduced and set to „1‟ for these individuals.

The procedure for bipolar disorder was analogous. The codes for bipolar disorder were 29610-29630, 296A, C-E and F30-F319. The new variable was called „bip‟ and set to „1‟ for these individuals.

Next, the two sets of patients were merged with the base dataset by the patients‟ ID- numbers, resulting in a set with individuals both with and without inpatient episodes for either of the two disorders, all born between 1973 and 1985. The values for variables „schizo‟ and „bip‟ were kept and the values for the individuals not from the patient sets set to „0‟. To each subject ID-number, we linked the ID-numbers of the biological parents (when available) and introduced the variables „biological mother/father diagnosed with schizophrenia/bipolar disorder‟. To this combined dataset were added one by one the variables „form of residence‟, „urbanicity‟ and „immigrant status‟, by puzzling together different datasets piece by piece, carefully avoiding creating double entries. The variables birth country of biological mother and of biological father were used to determine immigrant status of subject. For variables subject diagnosis, birth father diagnosis and birth mother diagnosis, the same registry datasets were used for the individual as for the parents, which meant extra steps in the procedure, for renaming variables and keeping the entries for mother and father connected to the right child. We were careful to add only variable data that related to the subjects‟ exposure at birth or early in childhood (i.e. before disorder onset). The housing data from the Swedish Population and Housing Census was linked to the subjects as the „form of residence‟ where the individual lived when he/she was between 1 and 5 years old. For example, those born in 1973-1974 were linked to their form of residence in 1975, those born in 1975-1979 to their form of residence in 1980 and births in 1981-1984 to the census in 1985. Finally, the births in 1985 were linked to the census in 1990. Gender and immigration status is constant over time

and age of subject at time of biological parent inpatient care for schizophrenia or bipolar disorder is not considered in the analyses. The two variables „diagnosis of father / mother" are combined to one, for the reason that we are more likely to have individuals lacking a registered biological father than a biological mother, which can bias the results.

The organisation of the dataset used SAS procedures including, but not limited to, WHERE and MERGE IN=xx / BY IF =xx to select specific entries. PROC SQL allows us to choose specific variables from two or more datasets and put them in a new dataset, mostly also using a WHERE statement. Please refer to Appendix A for examples of SAS code.

Variable coding

- The variables in the dataset used were coded as follows:
- Subject diagnosed with schizophrenia: 1=yes; 0=no
- Subject diagnosed with bipolar disorder: 1=yes 0=no
- Year of birth: 0=1973-1976; 1=1977-1980; 2=1981-1985
- Gender: 1=male; 2=female
- Form of residence in childhood: 1=house; 2=owned apartment; 3=rented apartment
- Urbanicity: 1=born in the municipality of Stockholm, Gothenburg or Malmoe; 0=elsewhere
- Immigrant status : 0= the individual and both parents born in Sweden, 1=the individual and one parent born in Sweden, 2=individual born in Sweden, both parents born abroad
- For patients diagnosed with schizophrenia: biological mother or father diagnosed with schizophrenia: 1=yes 0=no
- For patients diagnosed with bipolar disorder: biological mother or father diagnosed with bipolar disorder: 1=yes 0=no

Data analysis

The data set was analysed using logistic regression. This means that we used the logistic regression model, that is

$$\pi(x) = \exp(\alpha + \beta x) / (1 + \exp(\alpha + \beta x)) \quad \text{as } x \rightarrow \infty \quad [1]$$

$\pi(x)$ decreases towards 0 when $\beta<0$ and $\pi(x)$ increases towards 1 when $\beta>0$

The odds for [1] are

$$\pi(x) / (1 - \pi(x)) = \exp(\alpha + \beta x) \quad [2]$$

And thus the log odds is linear

$$\log \pi(x) / (1 - \pi(x)) = \alpha + \beta x \quad [3]$$

This is called the log-odds transformation, or the *logit*; logistic regression models are also called logit models. $\pi(x)$ must fall in the interval (0,1) but the logit can be any real number, which are the range for linear predictors.

PROC LOGISTIC in SAS. First we created univariate models for each risk factor, for each disorder, adjusted for year of birth and gender by including them in every model, along with one other variable at a time. Second we created two multivariate models for each disorder, the

first including all risk factors listed above, except for biological parent diagnosed with disorder. The second multivariate model contained all risk factors listed above. We interpreted the results in terms of odds ratios (OR) where H0: no effect on risk for disorder OR = 1 and calculated 95% confidence intervals for the OR.

The adjustment for birth year and gender was done without using interaction terms, for the reason that the data material was not extensive enough to approximate interaction terms.

Results

Table 1 below gives the characteristics of the dataset used for analysis. It gives the total number of individuals for each variable and category as well as the number of cases and what those numbers mean when interpreted as estimated number of cases per 1000 population individuals. Our cohort consisted of 1.199 million individuals, of which 616‟821 (51.4 %) were men and 582‟615 (48.6 %) were women. The estimated cumulative incidence (CuI)for schizophrenia for women is 1 in a thousand but for men this is almost doubled, at 1.7. For bipolar it is the men who have the estimate 1 in 1000 and the women the higher rate 1.5. For variables gender and birth year, we have no missing values. Regarding birth year, the estimated CuI for the oldest category is for schizophrenia 3 times that of the youngest, which is not mirrored in the values for bipolar disorder. Neither does bipolar disorder have any such marked difference regarding immigrant status, but schizophrenia also has a high estimated

CI for category „both parents born abroad‟ of the immigrant status variable, this CuI is 3.1 compared to the variable population's 1.3. Finally, looking at the parent diagnosis variable, we see that both disorders have IRs much higher for positive family history than for the variable population; CuI 9.2 for schizophrenia and 11.1 for BPD, compared to 1.3 and 1.2 respectively.

	Total	**Schizophrenia**		**Bipolar disorder**	
	Number | Missing	**Number**	**Per 1000**	**Number**	**Per 1000**
Gender	1'199'436 | 0	1623	1.4	1446	1.2
Men	616'821	1065	1.7	589	1.0
Women	582'615	558	1.0	857	1.5
Year of birth	1'199'436 | 0	1623	1.4	1446	1.2
1973-1976	395'851	814	2.1	553	1.4
1977-1980	358'489	488	1.4	463	1.3
1981-1985	445'096	321	0.7	430	1.0
Form of residence	1'001'885	1334	1.3	1199	1.2
House	506'370	489	1.0	503	1.0
Owned apartment	78'787	87	1.1	112	1.4
Rented apartment	416'728	758	1.8	584	1.4
Urbanicity	1'186'406 | 13'030	1598	1.3	1420	1.2
Major city	168'453	317	1.9	225	1.3
Other	1'017'953	1281	1.3	1195	1.2
Immigrant status	1'175'634 | 23'802	1571	1.3	1417	1.2
Native Swedish	1'010'164	1186	1.2	1169	1.2
One parent born in Sweden	110'844	215	1.9	179	1.6
Both parents born	54'626	170	3.1	69	1.3

abroad					
Biological parent diagnosed with same disorder	1’175’985 | 23451	1578	1.3	1419	1.2
Yes	14’745	136	9.2	163	11.1
No	1’161’240	1442	1.2	1256	1.1

Table 1. Data set characteristics

Univariate models

The models predict the probability that schizophrenia / bipolar disorder = „yes“ when fitting a model consisting of variables birth year, gender and one of the other variables at a time. The univariate analyses do not take interactions into account and the Ors are likely to change when we combine them all in the multivariate models. However, they point us in the direction we might expect to see in the multivariate analyses.

In the univariate models for schizophrenia, all variables are over-risks significant at confidence level 95%. The lower bound of the 95% CI for owned apartment vs. house is however very near the l.0–limit at the value 1.003. As we suspected from viewing the data in Table 1, the largest OR estimate is for both disorders that of parental diagnosis. For bipolar disorder and schizophrenia, the ORs are 13.162 and 14.601 and indicate a strong association. A difference between disorders is that for BPD, urbanicity and 2nd generation immigrant (both parents born outside of Sweden) are not significant, their ORs are 1.145 and 1.107; both CIs cover 1.0. Please refer to Table 2 for all ORs and CIs.

	Schizophrenia	**Bipolar Disorder**
Model	**Univariate* Odds Ratio (95% CI)**	**Univariate Odds Ratio (95% CI)**
Variable		
Form of residence		
House	**Reference**	**Reference**
Owned apartment	1.261 (1.003-1.585)	1.471 (1.198-1.806)
Rented apartment	1.563 (1.390-1.758)	1.336 (1.182-1.511)
Urbanicity (major city: yes vs no)	1.521 (1.345-1.721)	1.145 (0.993-1.321)
Immigrant status		
Native Swedish	**Reference**	**Reference**
One parent born in Sweden	1.696 (1.466-1.962)	1.406 (1.201-1.645)
Both parents born outside of Sweden	2.769 (2.357-3.253)	1.107 (0.868-1.412)
Mother or father diagnosed with same disorder (yes vs. no)	14.601 (11.863-17.972)	13.162 (11.071-15.648)

Table 2.
Results significant at confidence level 95% are underlined
***** Univariate model is adjusted for age and gender

Multivariate models

The models predict the probability that schizophrenia / bipolar disorder = „yes“

When we construct multivariate models, we note that all risk factors, except housing category „owned apartment“, are still significant for schizophrenia. Starting with Multivariate 1 in Table 3 below, all OR“s show over-risks and interesting to note is that the OR for having one Swedish parent is 1.562 compared to 2.351 for neither parent being Swedish-born. This means that if one parent is Swedish-born, you have 1.6 times the risk of a native Swede to develop

schizophrenia. If neither parent is born in Sweden, the risk increases to 2.4 times that of a native Swede. When comparing housing „owned apartment‟ with „house‟, the OR is 1.141 but not significant. The odds of schizophrenia development are almost 3 times higher if you are born in 1973-76 compared to in 1981-85. This probably has to do with that the last entries in the hospital registry which we used are from 2006. Then those born in 1981-85 were 21-25 years old and as mentioned earlier, schizophrenia onset mostly occur when the patient is in his 20‟s. They have thus had less chance of becoming ill and to enter the data registry. Looking at the values for M.v. 2, we can see that the estimated ORs have not essentially changed significance. It is however the variable with the largest estimated OR, at 10.52 (95% CI 8.2-13.5)

Schizophrenia		
Variable / Model	**Multivariate 1^ Odds Ratio (95% CI)**	**M.v. 2' Odds Ratio (95% CI)**
Gender (male vs female)	1.770 (1.577-1.987)	1.768 (1.575-1.984)
Year of birth		
1973-1976	2.997 (2.591-3.467)	2.979 (2.575-3.446)
1977-1980	1.872 (1.612-2.173)	1.870 (1.611-2.171)
1981-1985	**Reference**	**Reference**
Housing		
House	**Reference**	**Reference**
Owned apartment	1.141 (0.903-1.442)	1.121 (0.887-1.417)
Rented apartment	1.316 (1.163-1.489)	1.277 (1.128-1.445)
Urbanicity (Major city: yes vs no)	1.304 (1.136-1.498)	1.297 (1.30-1.489)
Immigrant status		
Native Swedish	**Reference**	**Reference**
One parent born in Sweden	1.562 (1.329-1.837)	1.512 (1.286-1.779)
Both parents born outside of Sweden	2.351 (1.967-2.810)	2.248 (1.880-2.689)
Mother or father diagnosed with same disorder (yes vs. no)	-	10.516 (8.215-13.461)

Table 3.
Significant OR's are underlined
^ Multivariate model 1 contains variables Age, Gender, Housing, Urbanicity, Immigration status (Native, one parent born abroad, both parents born abroad)
' Multivariate model 2 contains variables of Multivariate model 1 plus variable Disorder diagnosis of biological mother or biological father.

Turning to Table 4, we see that the results are similar for bipolar disorder, but the over-risk for men has changed to an under-risk. This means that our results say that men compared to women have an increased risk of developing schizophrenia, but for bipolar disorder, the risk for men is lower than the risk for women. We also see that as suggested by the univariate model, urbanicity and immigrant status category 'both parents born abroad' are insignificant. The estimates, in both M.v.1 and M.v.2, for 'both parents born abroad' are in fact under-risks, which means that had they been significant, they had indicated that the risk for developing BPD is lower if your parents are born outside Sweden while you are not. However, having one parent born in Sweden is a significant factor, with OR~1.3. Urbanicity has an estimated

OR~1 in both models. Positive parental diagnosis remains the largest effect at OR-estimate 12.75 (10.5-15.5).

Bipolar Disorder		
Variable / Model	**Multivariate 1^ Odds Ratio (95% CI)**	**M.v. 2' Odds Ratio (95% CI)**
Gender (male vs female)	0.683 (0.608-0.768)	0.681 (0.606-0.766)
Year of birth		
1973-1976	1.422 (1.224-1.651)	1.399 (1.205-1.624)
1977-1980	1.348 (1.174-1.548)	1.337 (1.164-1.535)
1981-1985	**Reference**	**Reference**
Housing		
House	**Reference**	**Reference**
Owned apartment	1.456 (1.180-1.796)	1.426 (1.156-1.760)
Rented apartment	1.312 (1.153-1.492)	1.281 (1.126-1.457)
Urbanicity (major city: yes vs no)	1.050 (0.896-1.230)	1.037 (0.885-1.215)
Immigrant status		
Native Swedish	**Reference**	**Reference**
One parent born in Sweden	1.311 (1.100-1.562)	1.282 (1.076-1.528)
Both parents born outside of Sweden	0.960 (0.733-1.256)	0.980 (0.749-1.282)
Mother or father diagnosed with same disorder (yes vs. no)	-	12.750 (10.499-15.484)

Table 4.
Significant OR's are underlined
^ Multivariate model 1 contains variables Birth year, Gender, Housing, Urbanicity, Immigration status (Native, one parent born abroad, 2 parents born abroad)
' Multivariate model 2 contains variables of Multivariate model 1 plus variable Disorder diagnosis of biological mother or biological father.

Conclusions

When reviewing the multivariate analyses" results, we see that gender which is an over-risk for men for schizophrenia is an under-risk for men for bipolar disorder. What the birth year variable really tells us is the „register effect": having had enough time to be caught in the system. This too appears stronger for schizophrenia than for BPD.

Housing („rented apartment"), immigrant status („one parent born in Sweden") and positive parental diagnosis, are over-risks for both schizophrenia and bipolar disorder. Urbanicity and immigrant status „both parents born abroad" are clear over-risks for schizophrenia but not for BPD. „Owned apartment" is an over-risk for BPD but just barely not for schizophrenia.

Thus it seems that the factors common to the both diseases are biological parents" disorders and housing. Gender, urbanicity and 2nd generation immigrant (both parents born outside of Sweden) are not, and year of birth is more of a control variable than an indicator.

Looking at results from previous research, we see that the risk factors we found significant are among those then found significant and that they are in the same direction. For BPD, there are no conclusive results in the literature, due to that the research so far has not been extensive enough, but as our results are similar to those for schizophrenia, we find it likely that future research will find these to be risk factors for BPD as well.

Discussion

Statistical epidemiological methodologies and problems

To cite Alloy (2005): The review of developmental risk factor studies arrive at the now familiar conclusion: results are mixed and more research on methodological issues are called for. [1]

We basically have two study types to choose from: the cohort study, which is prospective, and the case-control study, which is retrospective. There are advantages and disadvantages to both, but it is noteworthy that most studies on bipolar disorder so far are retrospective and that the review articles call for prospective longitudinal ones. In 2006, Alloy et al pointed out the problem of assigning the causality of the association as A □B or B □A, usually all that can be said is that there is an association between A and B. This is for case-control studies, cohort studies can give a better idea and the description below of the two methods will tell why.

The population sample used in the study can be selected in different ways. There is the cohort method, the case-control method and the matched case-control method to name some. The study type also determines what effect measure is used to interpret the results.

In a cohort (prospective and longitudinal) the participants are followed up over time to see whether they become sick (or die). Here we have the opportunity to select participants by random sampling, which helps to avoid bias. Sometimes it is also possible to sample (almost) the whole population, which minimizes the variance. The individuals entering the study are ideally non-afflicted by the disease being investigated. At baseline (the starting time point), it is common to divide the selected participants into two groups; one with individuals with the suspected risk characteristic and the other without it. The two major advantages of cohort studies are that 1) they demonstrate causality because we can see the order and timeline of incidences and 2) we can study a number of diseases simultaneously. One of the problems with cohort studies is that they are time-consuming and expensive. If the disease in question has a very long latency, the study must be planned to go on for many years of follow up and that is not realistic. Also, cohort studies are not suitable for rare diseases, because in order to sample a sufficient number of cases, the total number of sample individuals would be very large. Confounders are another problem and they are uncontrolled due to the self-assigning of participants to risk factor status. Finally, when a study goes on for a period of time, there is a risk of dropouts (people leaving the study before experiencing an event or before the study period has ended).

There is an analysis method called the person-years method, which Mortensen et al (2010) used. The person-years incidence rate *p* is estimated by

$$p=e/y$$

e=number of events during follow-up, y=sum of years individuals spent in the study.

As soon as an individual has an event or is censored, his or her follow-up is terminated. If an individual has no event during follow-up, he or she contributes the whole follow-up time to the sum. P is interpreted as number of events per person year, and can be scaled to for example no. of events per 1000 per person year. Because the selections of participants are random, cohort studies can estimate the population incidence rate. This is done by using number of participants in the denominator in the equation above. This too can be scaled up by calculating events per 1000 people. From cohort studies we can also estimate relative risk (the ratio between the risks of disease for exposed and unexposed subjects)

A case-control study is a retrospective study, where a number of cases and healthy controls are selected and their background is investigated with the hope of discovering what has caused the disease. The case-control study is much cheaper than the cohort study; the individuals are interviewed and/or examined once and there is no waiting time. They can also be asked questions on several suspected risk factors and rare diseases are not a problem here. Finally, they require smaller sample sizes and confounding can be estimated more precisely than for cohort studies. However, because of the lack of a time sequence, we cannot demonstrate causality. Which came first, the risk factor or the disease onset? Which caused which, if at all? .

In the matched case-control study, cases are identified and then an equal number of controls are matched to each case, matched by age, gender and other relevant factors. Concordant pairs are pairs where the diseased and the healthy have had the same exposure, in a discordant pair, the case and the control have different exposures.

Case-control studies can estimate the measure odds ratio, but because the sample in a case-control is by its design not at all random, it cannot estimate population risk, relative risk (it is not a valid estimate of its population equivalent) or odds.

To investigate the aetiology of diseases, the cohort study is a lot more preferable to the case-control type. As mentioned above, the cohort study follows the at outset disease-free participants over time and registers the order of events. It provides the possibility of assigning causality.

The article by Alloy et al (2006) begins the methods section with an interesting notion: what conditions must a risk factor for a disorder meet? They propose: 1) consistent association with the disorder 2) must temporally precede affective episodes and 3) must be to some degree stable independently of disorder symptoms. Since their article is focussed on bipolar disorder, the second condition says „precede affective episodes", but for non-affective disorders this can be changed. An example: in the case of schizophrenia, we can alter it to „episodes of psychoses". These criteria lead to the same conclusion as above, that retrospective studies are not sufficient for determining the stability of psychosocial variables independent of symptoms, but they can be used to identify possible risk factors. Instead, prospective longitudinal studies are needed as the possibility of assigning causality is greater, but not even they are safe from confounders, in agreement with our discussion above.

Withdrawals, or drop-outs, create a big problem for the analysis. If we ignore their data we will tend to overestimate the risk (we removed follow-up time free from event) and if we include them with those who stayed in the study until its end and had no event, we underestimate the risk, as we cannot be sure that an event would not have occurred during the

time left of the study. Neither is ideal. To include this person among the event-positive is naturally not an option. If reason for withdrawal is connected to the risk factor but not the outcome, such as in a study of lung disease, where many people leave the studied work environment because of severe skin rashes, the analysis is not invalidated but the outcome variable may be insubstantial.

For all study designs, it is also important that the size of the study population is not too small; the statistical power diminishes and only greater effects can be detected.

We conclude that a cohort study is preferable, as it can better assign causality than the case-control study which cannot allocate it at all.

Strengths and weaknesses

An advantage of our study is that we used the cohort method, where we tried to include as much of the study population (people born in Sweden) and all hospitalised cases of the disorders. Also, we have used variables from the subjects‟ childhood, which gives us control over the order of exposure and disorder onset.

A problem that we are faced with throughout the project is that in the underlying data, patients have been diagnosed with schizophrenia and bipolar disorder according to several diagnostic tools as mentioned in the methods section, namely International Classification of Diseases – eighth, ninth and tenth revision (ICD 8 – 10). In the literary review section some of the discussed studies outside of Sweden have also used the system developed by the American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders (DSM) in various editions, mostly the fourth (DSM-IV) published in 1994.

The number of variables used in our project is limited, due to a limit of 20 weeks working time, including the time needed to learn new SAS syntax necessary for the data set construction and to write the project report.

Missing values is usually another problem, but for children born in Sweden during 1973-1985, we have access to very good register data. The variable with most missing values is parental diagnosis, where circa 2% of data is missing.

Variables have also been divided into fairly wide categories, an example is urbanicity, which has only the categories „born in Stockholm, Göteborg or Malmö‟ or not. A finer organisation could for example be that the values are assigned to one of a range of population density categories. The indicator variables themselves are also somewhat rough, a weakness of the register study.

Further directions

A natural extension of this project is to add more socio-economic variables to the data for regression analysis. Variables that are of interest to add to the models are for example parental education level, family income during subject‟s childhood, employment status and type for the subject‟s parents. Variables can also be created on a finer scale, for example investigating immigrants by country of origin and in addition to simply „immigrant status‟.

References

1. Agresti, A. (2002). Categorical Data Analysis. *John Wiley and Sons,* Hoboken NJ, USA, 122-123.

2. Alloy, L.B., Abramson, L.Y., Urosevic, S., Walshaw, P.D., Nusslock, R., & Neeren, A.M. (2005). The psychological context of bipolar disorder: Environmental, cognitive, and developmental risk factors. *Clinical Psychology Review, 25,* 1043-1075.

3. Alloy, L.B., Abramson, L.Y., Smith, J.M., Gibb, B.E., &Neeren, A.M. (2006). Role of Parenting and Maltreatment Histories in Unipolar and Bipolar Mood Disorders: Mediation by Cognitive Vulnerability to Depression. *Clinical Child and Family Psychology Review, Vol. 9, No. 1,* 23-60.

4. Bale, T.L. (2006). Stress sensitivity and the development of affective disorders. *Hormones and Behavior, 50,* 529-533.

5. Buehlmann, E., Berger, G.E., Aston, J., Gschwandtner, U., Plueger, M.O., Borgwardt, S.J., Radue, E-W., & Riecher-Rössler, A. (2009). Hippocampus abnormalities in at risk mental states for psychosis? A cross-sectional high resolution region of interest magnetic resonance imaging study. *Journal of Psychiatric Research, 44:(7),* 447-453.

6. Byrne, M., Agerbo, E., Eaton, W.W., & Mortensen, P.B. (2004). Parental socioeconomic status and risk of first admission with schizophrenia: a Danish national register based study. *Social Psychiatry Psychiatric Epidemiology, 39,* 87 – 96.

7. Byrne, M., Agerbo, E., Bennedsen, B., Eaton, W.W., & Mortensen, P.B. (2007). Obstetric conditions and risk of first admission with schizophrenia: a Danish national register based study. *Schizophrenia Research, 97(1-3),* 51 – 59.

8. Cannon, M., Jones, P.B., & Murray, R.M. (2002). Obstetric Complications and Schizophrenia: Historical and Meta-Analytic Review. *American Journal of Psychiatry, 159;* 1080 – 1092.

9. Cantor-Graae, E., Pedersen, C.B., McNeil, T.F., & Mortensen, P.B. (2003). Migration as a risk factor for schizophrenia: a Danish population-based cohort study. *British Journal of Psychiatry, 182,* 117 – 122.

10. Cantor-Graae, E. (2007). The Contribution of Social Factors to the Development of Schizophrenia: A Review of Recent Findings. *Canadian Journal of Psychiatry, 52,* 277 – 286.

11. Cantor-Graae, E., & Selten, J.-P. (2005). Schizophrenia and Migration: A Meta-Analysis and Review. *American Journal of Psychiatry, 162:1,* 12-24.

12. Christensen, M.V., & Kessing L.V. (2006). Do personality traits predict first onset in depressive and bipolar disorder? *Nord J Psychiatry, 60,* 79-88.

13. Craddock, N., O‟Donovan, M.C., & Owen, M.J. (2005). The genetics of schizophrenia and bipolar disorder: dissecting psychosis. *Journal of Medical Genetics, 42(3),* 193-204.

14. Gilmore, J.H. (2010). Understanding What Causes Schizophrenia: A Developmental Perspective. *American Journal of Psychiatry, 167:1,* 8 – 10.

15. Harrison, G., Fouskakis, D., Rasmussen, F., Tynelius, P., Sipos, A., & Gunnell, D. (2003). Association between psychotic disorder and urban place of birth is not mediated by obstetric complications or childhood socioeconomic position: a cohort study. *Psychological Medicine, 33,* 723-731.

16. Hjern, A., Wicks, S., & Dalman, C (2004). Social adversity contributes to high morbidity in psychoses in immigrants – a national cohort study in two generations of Swedish residents. *Psychological medicine, 34,* 1025 – 1033.

17. Howes, O.D., McDonald, C., Cannon, M., Arsenault, L., Boydell, J., & Murray, R.M. (2004). Pathways to schizophrenia: the impact of environmental factors. *International Journal of Neuropsychology, 7 (Supplement 1),* 7 – 13.

18. Kelly, B.D., O‟Callaghan, E., Waddington, J.L., Feeney, L., Browne, S., Scully, P.J., Clarke, M., Quinn, J.F., McTigue, O., Morgan, M.G., Kinsella, A., & Larkin, C. (2010). Schizophrenia and the city: A review of literature and prospective study of psychosis and urbanicity in Ireland. *Schizophrenia Research, 116,* 75 – 89.

19. Kinney, D.K., Hintz, K., Shearer, E.M., Barch, D.H., Riffin, C., Witley, K., & Butler, R. (2010). A unifying hypothesis of schizophrenia: Abnormal immune system development may help explain role of prenatal hazards, post-pubertal onset, stress, genes, climate, infections, and brain dysfunction. *Medical Hypotheses, 74 (3),* 555 – 563.

20. Kirkbride, J.B., Boydell, J., Ploubidia, G.B., Morgan, C., Dazzan, P., McKenzie, K., Murray, R.M., & Jones, P.B. (2008). Testing the association between the incidence of schizophrenia and social capital in an urban area. *Psychological Medicine, 38,* 1083 – 1094.

21. Kirkbride, J.B., Barker, D., Cowden, F., Stamps, R., Yang, M., Jones, P.B., & Coid, J.W. (2008). Psychosis, ethnicity and socioeconomic status. *British Journal of Psychiatry, 193 (1),* 18 - 24.

22. Leask, S.J., Vermunt, J.K., Done, D.J., Crow, T.J., Blows, M., & Boks, M.P. (2009). Beyond symptom dimensions: Schizophrenia risk factors for patient groups derived by latent class analysis. *Schizophrenia Research, 115,* 346 – 350.

23. Lichtenstein, P., Yip, H.P., Björk, C., Pawitan, y., Cannon, T.C., Sullivan, P.F, & Hultman, C.F. (2009). Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *Lancet, 373,* 234-239.

24. McGrath, J., Saha, S., Wellham, J., El Saadi, O., MacCauley, C., & Chant, D. (2004). A systematic review of the incidence of schizophrenia: the distribution of rates and the influence of sex, urbanicity, migrant status and methodology. *BMC Medicine, 2:13*. http://www.biomedcentral.com/1741-7015/2/13

25. Mortensen, P.B., Pedersen, M.G., & Pedersen, C.B. (2010). Psychiatric family history and schizophrenia risk in Denmark: which mental disorders are relevant? *Psychological Medicine, 40,* 201 – 210.

26. Post, R.M., & Leverich, G.S. (2006). The role of psychosocial stress in the onset and progression of bipolar disorder and its comorbities: The need for earlier and alternative modes of therapeutic intervention. *Development and Psychopathology, 18,* 1181-1211.

27. Prince, Martin (1998). An overview of psychiatric epidemiology. *International Review of Psychiatry, 10,* 258-263.

28. Scott, J., McNeill, Y., Cavanagh, J., Cannon, M., & Murray, R. (2006). Exposure to obstetric complications and subsequent development of bipolar disorder: Systematic Review. *British Journal of Psychiatry, 189,* 3-11.

29. Strakowski, S.M., & DelBello, M.P. (2000). The co-occurence of bipolar and substance use disorders. *Clinical Psychology Review, Vol. 29, No. 2,* 191-206.

30. Swinnen, S.G.H.A., & Selten J-P. (2007). Mood disorders and migration – meta analysis. *British Journal of Psychiatry, 190,* 6-10.

31. Torrey, E.F., & Yolken, R.H. (1998). At Issue: Is Household Crowding a Risk Factor for Schizophrenia and Bipolar Disorder? *Schizophrenia Bulletin, 24,* 321-324.

32. Tsuang, M. (2000). Schizophrenia: Genes and Environment. *Society of Biological Psychiatry, 47,* 210 – 220.

33. Tsuchiya, K.J., Agerbo, E., Byrne, M., & Mortensen, P.B. (2004). Higher socio-economic status of parents may increase the risk for bipolar disorder in the offspring. *Psychological Medicine, 34,* 787 – 793.

34. Tsuchiya, K.J., Byrne, M., & Mortensen, P.B. (2003). Risk factors in relation to an emergence of bipolar disorder: a systematic review. *Bipolar Disorder, 5,* 231-242.

35. Weiser, M., Werbeloff, N., Vishna, T., Yoffe, R., Lubin, G., Shmushkevitch; M., & Davidson; M. (2007). Elaboration on immigration and risk for schizophrenia. *Psychological Medicine, 38,* 1113 – 1119.

36. Wicks, S., Hjern, A., Gunnell, D., Lewis, G., & Dalman, C. (2005). Social Adversity in Childhood and the Risk of Developing Psychosis: A National Cohort Study. *American Journal of Psychiatry, 162,* 1652–1657.

37. Winokur, G., Coryell, W., Akiskal, H. S., Maser, J. D., Keller, M. B., Endicott, J., & Mueller, T. (1995). Alcoholism in manic-depressive (bipolar) illness: Familial illness, course of illness, and the primary-secondary distinction. *American Journal of Psychiatry, 152,* 365 – 372.

38. Woodward, M. (1999). Epidemiology, study design and data analysis. *Chapman and Hall,* Boca Raton FL, USA, 191-284.

PubMed and online links to references (numbering as above)

1. ISBN 978 047 136 093 3

2. http://www.ncbi.nlm.nih.gov/pubmed/16140445

3. http://www.ncbi.nlm.nih.gov/pubmed/16718583

4. http://www.ncbi.nlm.nih.gov/pubmed/16901485

5. http://www.ncbi.nlm.nih.gov/pubmed/19939408

6. http://www.ncbi.nlm.nih.gov/pubmed/15052389

7. http://www.ncbi.nlm.nih.gov/pubmed/17764905

8. http://www.ncbi.nlm.nih.gov/pubmed/12091183

9. http://www.ncbi.nlm.nih.gov/pubmed/12562738

10. http://www.ncbi.nlm.nih.gov/pubmed/17542378

11. http://www.ncbi.nlm.nih.gov/pubmed/15625195

12. http://www.ncbi.nlm.nih.gov/pubmed/16635925

13. http://www.ncbi.nlm.nih.gov/pubmed/15744031

14. http://www.ncbi.nlm.nih.gov/pubmed/20068120

15. http://www.ncbi.nlm.nih.gov/pubmed/12785474

16. http://www.ncbi.nlm.nih.gov/pubmed/15551573

17. http://www.ncbi.nlm.nih.gov/pubmed/14972079

18. http://www.ncbi.nlm.nih.gov/pubmed/19897342

19. http://www.ncbi.nlm.nih.gov/pubmed/19836903

20. http://www.ncbi.nlm.nih.gov/pubmed/17988420

21. http://www.ncbi.nlm.nih.gov/pubmed/18700213

22. http://www.ncbi.nlm.nih.gov/pubmed/19840897

23. http://www.ncbi.nlm.nih.gov/pubmed/19150704

24. http://www.ncbi.nlm.nih.gov/pubmed/15115547

25. http://www.ncbi.nlm.nih.gov/pubmed/19607751

26. http://www.ncbi.nlm.nih.gov/pubmed/17064434

27.http://pdfserve.informaworld.com/103354_731377804_713673622.pdf

28. http://www.ncbi.nlm.nih.gov/pubmed/16816299

29. http://www.ncbi.nlm.nih.gov/pubmed/10721497

30. http://www.ncbi.nlm.nih.gov/pubmed/1719765

31. http://www.ncbi.nlm.nih.gov/pubmed/9718626

32. http://www.ncbi.nlm.nih.gov/pubmed/10682218

33. http://www.ncbi.nlm.nih.gov/pubmed/15500299

34. http://www.ncbi.nlm.nih.gov/pubmed/12895201

35. http://www.ncbi.nlm.nih.gov/pubmed/17988415

36.http://www.ncbi.nlm.nih.gov/pubmed/16135624

37. http://www.ncbi.nlm.nih.gov/pubmed/7864261

38. ISBN 158 488 009 0

APPENDIX A

Choosing entries using 'WHERE'

```
Data chosen;

Set base;

Where variable name=: A OR variable name between 'B' and 'C' AND variable name ^=D;

Run;
```

This creates a new dataset *chosen*, containing all entries of *variable name* from the dataset *base* which value contains „A" or is between „B" and „C" (this interval is inclusive) but does not equal D.

How to merge datasets while controlling for duplicates in the new set

```
Data herbs;

Merge forest field (in=aa);

By plant; if aa;

Run;
```

We are merging the datasets *forest* and *field* by their common variable *plant*. The in= / if= statement ensures that only entries of *plant* that exist in *field* are included in the new dataset. Entries of *plant* that only exist in *forest* will thus not be included in the new set *herbs*.

Selecting to include only certain variables when merging two or more datasets using 'PROC SQL'

```
Proc sql;

Create table auswahl as

Select a.wohnung, b.hamster

From Work.Wohnen as a, Work.Tiere as b

Where a.jemand = b.jemand;
```

This creates a new dataset *auswahl* made up of the variable *wohnung* from set *Wohnen* and variable *hamster* from set *Tiere*, both sets in SAS library „Work". Both sets contain the variable *jemand*. The procedure chooses only entries from the two sets where the entry for *wohnung* in *Wohnen* has the same value for *jemand* as does the entry for *hamster* in *Tiere*. Proc sql does not need a run statement at the end.