



VCU

Virginia Commonwealth University
VCU Scholars Compass

Theses and Dissertations

Graduate School

2008

Season of Birth and Risk for Schizophrenia

Seth Roberts

Virginia Commonwealth University

Follow this and additional works at: <https://scholarscompass.vcu.edu/etd>



Part of the [Epidemiology Commons](#)

© The Author

Downloaded from

<https://scholarscompass.vcu.edu/etd/1633>

This Thesis is brought to you for free and open access by the Graduate School at VCU Scholars Compass. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of VCU Scholars Compass. For more information, please contact libcompass@vcu.edu.

Season of Birth and Risk for Schizophrenia

by

Seth Belote Roberts

Gonzalo Bearman, MD, Advisor

Kenneth Kendler, MD, Preceptor

Department of Epidemiology and Community Health
Master of Public Health Program
MPH Research Project: EPID 691

Virginia Commonwealth University
Richmond, Virginia

December 2008

Submission Statement

Master of Public Health Research Project

This MPH Research Project report is submitted in partial fulfillment of the requirements for a Master of Public Health degree from Virginia Commonwealth University's School of Medicine. I agree that this research project report be made available for circulation in accordance with the program's policies and regulations pertaining to documents of this type. I also understand that I must receive approval from my Faculty Advisor in order to copy from or publish this document, or submit to a funding agency. I understand that any copying from or publication of this document for potential financial gain is not allowed unless permission is granted by my Faculty Advisor or (in the absence of my Faculty Advisor) the Director of the MPH Program.

Student Signature

Date

12/4/2008

Master of Public Health
Research Project Agreement Form
Department of Epidemiology and Community Health

Student name: Seth B. Roberts E-mail address: sbroberts@vcu.edu

Street address: 3650 Hermitage Road

Richmond, VA 23227

Home phone: (804)355-0113 Work phone: (804)828-3897 Fax: _____

Number of semester hours (3-6): 3 Semester: Summer/Fall Year: 2008

Please complete the following outline. Do not exceed 2 pages (A-H).

A. PROJECT TITLE:

Season of birth and risk for schizophrenia

B. PURPOSE (state hypothesis/research question):

Season of birth, specifically birth in late winter or early spring, is thought to confer a slight elevation in risk for schizophrenia. The effect, though small in size, has been repeatedly observed over decades of research. Multiple hypotheses have been put forth to explain this phenomenon, for example, that maternal influenza infection during the second trimester of pregnancy impairs normal fetal brain development, which subsequently leads to increased risk for schizophrenia. Our hypothesis is that elevated risk for schizophrenia is conferred by an environmental risk factor indexed by season of birth, and further, that this risk factor alters risk for schizophrenia depending on one's genotype, i.e., that this is an example of gene – environment interaction in conferring risk for a complex human phenotype.

C. SPECIFIC OBJECTIVES (list major aims of the study):

1. To evaluate whether there is evidence for an effect of season of birth on risk for schizophrenia using the Irish Study of High Density Schizophrenia Families
2. To examine, by means of Monte Carlo simulations, the power to detect an association between schizophrenia and genotypes, while accounting for season of birth, under a range of possible genetic architectures
3. To conduct test specific genetic loci for association with schizophrenia, accounting for season of birth, using the Irish Case Control Schizophrenia Sample

D. DESCRIPTION OF METHODS

D.1. Identify source(s) of data (eg, existing data set, data collection plans, etc):

Existing data set – Irish Study of High Density Schizophrenia Families

Existing data set – Irish Case Control Schizophrenia Sample

D.2. State the type of study design (eg, cross-sectional, cohort, case-control, intervention, etc):

Case-control

D.3. Describe the study population and sample size:

The Irish Study of High Density Schizophrenia Families (ISHDSF) was formed by standardized ascertainment of multiplex schizophrenia families from psychiatric facilities covering over 90% of the population in Ireland and Northern Ireland. It is comprised of 277 families with 1,770 individuals, 837 of whom have schizophrenia.

The Irish Case Control Study is a sample of schizophrenic and control individuals who have been genotyped at many loci across the genome. Schizophrenics were ascertained through in- and outpatient psychiatric facilities, had diagnoses verified by an expert, and their birthdates recorded. Controls were selected from several sources, e.g. blood donation centers, and denied any lifetime history of schizophrenia. For each subject in the ICCSS, all four grandparents were born in Ireland or the United Kingdom.

D.4. List variables to be included (If a qualitative study, describe types of information to be collected)

affected status for schizophrenia (schizophrenia, control), month of birth, genotypes at 6 loci

D.5. Describe methods to be used for data analysis (If a qualitative study, describe general approach to compiling the information collected)

Objective 1: chi-square test comparing monthly birth rates in cases and controls

Objective 2: Monte Carlo simulations under a range of possible genetic architectures, analysis of data using methods from Objective 3

Objective 3: chi-square test comparing allele frequencies in controls, cases born in low-risk seasons, and cases born in high-risk seasons

E. ANTICIPATED RESULTS:

Objective 1: There will be evidence for an effect of season of birth in the ISHDSF

Objective 2: Under certain circumstances, power to detect genetic associations with schizophrenia will be increased by dividing cases into low- and high-risk birth seasons for analysis

Objective 3: There will be evidence for genetic loci that increase risk for schizophrenia depending on season of birth

F. SIGNIFICANCE OF PROJECT TO PUBLIC HEALTH:

This study could help further understanding of how environmental factors influence risk for schizophrenia (what genetic pathways are involved), suggest new therapeutic targets, and help to identify individuals at elevated risk for schizophrenia (allowing for early screening).

G. IRB Status:

1) Do you plan to collect data through direct intervention or interaction with human subjects? ___yes ___x___no

2) Will you have access to any existing identifiable private information? ___yes ___x___no

If you answered “no” to both of the questions above, IRB review is not required.
If you answered “yes” to either one of these questions, your proposed study must be reviewed by the VCU Institutional Review Board (IRB). Please contact Dr. Turf or Dr. Buzzard for assistance with this procedure.

Please indicate your IRB status:

- to be submitted (targeted date _____)
 submitted (date of submission _____; VCU IRB # _____)
 IRB exempt review approved (date _____)
 IRB expedited review approved (date _____)
 IRB approval not required

H. PROPOSED SCHEDULE: Start Date: June, 2008 Anticipated End Date: Dec, 2008

I. INDICATE WHICH OF THE FOLLOWING AREAS OF PUBLIC HEALTH KNOWLEDGE WILL BE DEMONSTRATED:

1. Biostatistics – collection, storage, retrieval, analysis and interpretation of health data; design and analysis of health-related surveys and experiments; and concepts and practice of statistical data analysis. _yes _no (if yes, briefly describe):

This study will use Monte Carlo simulations for determining the power of a statistical test under various assumptions.

2. Epidemiology – distributions and determinants of disease, disabilities and death in human populations; the characteristics and dynamics of human populations; and the natural history of disease and the biologic basis of health. _yes _no (if yes, briefly describe):

This is a study of determinants of schizophrenia, both environmental (those indexed by season of birth) and genetic.

3. Environmental Health Sciences – environmental factors including biological, physical and chemical factors which affect the health of a community. _yes _no (if yes, briefly describe):
4. Health Services Administration – planning, organization, administration, management, evaluation and policy analysis of health programs. _yes _no (if yes, briefly describe):
5. Social/Behavioral Sciences – concepts and methods of social and behavioral sciences relevant to the identification and the solution of public health problems. _yes _no (if yes, briefly describe):

SIGNATURE PAGE
Master of Public Health Research Project

Student: Name: Seth Roberts, MD/PhD

Preceptor: Name: Kenneth Kendler, MD Title: Prof. Psychiatry and Human Genetics

Address: VIPBG; PO Box 980126; Biotech One; VCU

E-mail: kendler@vcu.edu Phone: 828-8590

Field of expertise: genetic epidemiology of psychiatric disorders

Faculty


Advisor: Name: Gonzalo Bearman, MD

E-mail: gbearman@mcvh-vcu.edu Phone: 828-6938

SIGNATURES:

Student:  Date: 5/13/08

Preceptor:  Date: 5-13-08

Faculty Advisor:  no, MPH Date: 5/13/08

MPH Program Director:  Date: 9/10/08

MPH Program Coordinator:  Date: 12-5-08

MPH Research Project Approval Form

Schizophrenia and Season of Birth

Submitted to the Graduate Faculty of the
Department of Epidemiology and Community Health
Virginia Commonwealth University

In partial fulfillment of the requirements for the degree of
Master of Public Health

By: Seth Belote Roberts

Comments:

Approval signatures:




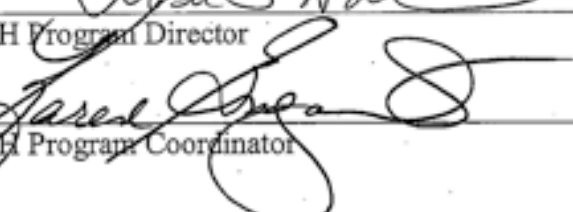
	12/4/2008
MPH Student	Date
 M.D., M.P.H.	12/4/08
MPH Research Project Faculty Advisor	Date
 Lisa S. Aul	12/5/08
MPH Program Director	Date
 Jared S. ...	12/5/08
MPH Program Coordinator	Date

Table of Contents

I. Introduction	1
II. Methods	8
III. Results	20
IV. Discussion	24
V. Conclusion	29
VI. Tables	30
VII. Figures.....	34
VIII. Appendix: Python script	35
IX. References	39

Acknowledgements

The author wishes to thank: 1) Kenneth Kendler, M.D., for generously sharing his time, for providing critical advice, guidance, and encouragement, and for making available the data used in this study, 2) Gonzalo Bearman, M.D., for his enthusiastic support and advocacy on my behalf during the Master of Public Health Program, and 3) my wife Libby, and my beautiful daughters, Eleanor and Emelia, for their patience, love, and forbearance during my studies.

Abstract

Background: Schizophrenia is a chronic, debilitating mental disorder characterized by positive (e.g., hallucinations, delusions) and negative (e.g., catatonia, flat affect) signs and symptoms. Many studies suggest that individuals born in winter or spring months are at increased risk for schizophrenia.

Study Objectives: 1) To determine whether season of birth affects risk for schizophrenia in the Irish Study of High Density Schizophrenia Families (ISHDSF). 2) To examine, by computer simulation, power to detect genetic associations with schizophrenia under a variety of conditions and using different analytic strategies. 3) To test whether specific genes are associated with schizophrenia in the Irish Case Control Schizophrenia Study (ICCSS), using different analytic strategies to account for season of birth.

Methods: A case-control design was used to examine the relationship between schizophrenia and season of birth. Cases were individuals from the ISHDSF diagnosed with schizophrenia. Controls were the general population of Ireland, with data provided by Ireland's Central Statistics Office (CSO). The birth frequencies for each month or quarter were compared in the two groups by a chi square test. Computer simulations were conducted to examine power to detect schizophrenia susceptibility loci using either all cases or only cases born in high-risk months, under different conditions and models for how genetic risk and season of birth interact to influence risk for schizophrenia. Data for six genetic markers from the ICCSS were analyzed for evidence of association, using all cases, and then using only cases born in high-risk months.

Setting and Study Participants: ISHDSF families were ascertained through inpatient psychiatric care facilities serving >95% of the Irish population. Diagnoses were established using DSM-III-R criteria, and birthdates were recorded for all individuals. The Irish CSO provided aggregate, population-level data for number of births in Ireland by month for the years 1976-2000 and by quarter for the years 1900-2000. The ICCSS is a sample of schizophrenic and control individuals who have been genotyped at many loci across the genome. Schizophrenics were ascertained through in- and outpatient psychiatric facilities, had diagnoses verified by an expert, and their birthdates recorded. Controls were selected from several sources, e.g. blood donation centers, and denied any lifetime history of schizophrenia. For each subject in the ICCSS, all four grandparents were born in Ireland or the United Kingdom.

Results: Number of births in each month was compared for schizophrenics in the ISHDSF and general population controls, resulting in a chi square of 19.44 (p value ~ 0.054 , 11 df). Simulations revealed that, in some circumstances, power to detect genetic associations was increased by restricting cases to those born in high-risk months. Analysis of data from the ICCSS revealed that restricting cases to high-risk birth months increased the evidence for association for three of six markers tested, two of which were associated with the gene FBXL21.

Conclusions: Birth in the months of March, April, or May appears to be associated with elevated risk for schizophrenia in the ISHDSF. In attempting to find susceptibility loci for schizophrenia, restricting genetic association analyses to schizophrenics born in high-risk months may result in increased power to detect genetic association in some circumstances.

Introduction

Background and significance:

Schizophrenia is a mental disorder characterized by abnormalities in the perception or expression of reality. According to the Diagnostic and Statistical Manual[1], schizophrenia includes characteristic symptoms, such as delusions, hallucinations, and/or grossly disorganized behavior; social and/or occupational dysfunction below the level achieved prior to onset of symptoms; and a duration of at least six months (see Box 1). Although the course of schizophrenia may be marked by varying degrees of recovery, complete cures are uncommon. The average duration that an individual lives with schizophrenia is approximately 30 years[2]. The toll of schizophrenia on patients and their families is considerable. Affected individuals have a 10 year reduction in life expectancy[3], mostly an effect of increased rates of suicide. Among the leading causes of disability-adjusted life years, schizophrenia ranks 8th in 15-44 year olds[4]. Schizophrenia is a relatively common disease. The point prevalence has been estimated at 4.6 per 1000 persons (80% confidence intervals 1.9-10)[5]. The incidence has been estimated as 15.2 per 100,000 persons per year (80% confidence intervals 8-43)[6]. Lifetime risk of schizophrenia is estimated to be 0.7%[5].

Research to identify the causes of schizophrenia has been an active area of investigation for decades. Although much has been learned, a recent survey of findings over the last 20 years concluded, "what we can confidently assert is essentially the same -- both genetic and environmental factors are important, but exactly which specific exposures and exactly how they cause schizophrenia is still unknown..."[2]. Family history of schizophrenia is a risk factor. The

relative risk associated with a positive family history depends on how closely related the affected relative is to the person at risk. Having a monozygotic twin with schizophrenia equates to a relative risk (RR) of 50-70, a first degree relative (e.g., a sibling) to a RR of 9-18, and a second degree relative to a RR of 3-6[2]. At least some of the familial risk is a consequence of specific genetic variants; dozens of loci within the human genome have been identified as putative schizophrenia susceptibility loci, although the number of replicated findings is much smaller[7]. The RR associated with any specific single gene variant is 1.1-1.5[2]. Non-genetic risk factors for schizophrenia include living in an urban environment (RR 2-3), history of migration (RR 2-3), and use of certain illicit drugs (RR 2-3)[2].

Another putative non-genetic risk factor for schizophrenia is season of birth. Tramer first made the observation that among 3100 patients diagnosed with “psychosis”, there were 15% excess births in winter-spring months (cited in [8]). Since then, many researchers have attempted to replicate this finding. The most extensive single review of all these studies was that conducted by Torrey et al, in 1997[8]. These authors reviewed > 250 articles, spanning 34 countries in both the Northern and Southern Hemispheres and nearly 80 years of accumulated research. Even in the face of methodological inconsistencies and other shortcomings, the authors concluded that the studies were remarkably consistent in demonstrating a small, but significant, winter-spring birth excess among schizophrenics. A more recent systematic review and meta-analysis of Northern Hemisphere season of birth studies found a significant excess of winter/spring births compared to summer/autumn (pooled odds ratio 1.07)[9]. More recent articles from Japan[10], Poland[11], and Puerto Rico[12] have provided additional evidence to support an excess of schizophrenic births in winter-spring.

There are several prior studies that have addressed whether season of birth is a risk factor for schizophrenia using samples from Ireland. O'Hare et al.[13] found 11% excess births for the second quarter of the year in 4855 inpatients born between 1921 and 1955. O'Callaghan et al.[14] examined 3253 persons with schizophrenia ascertained by case registers and found a 5% excess of births from the first quarter, although the results were non-significant. Significant results were found in this same study when subjects were restricted to individuals living in urban environments. O'Callaghan et al.[15] examined 616 inpatients from 1983-1988 and found a 9% excess of schizophrenia births for the first quarter, which was not statistically significant. This last study further examined results from two patient subsets: those with a first-degree relative affected by schizophrenia and those without. Season of birth and schizophrenia were found to be significantly associated only in those individuals without a family history. Similar findings have been reported in other, non-Irish samples[16], although other researchers have found evidence for a season of birth effect even in individuals with a family history[17]. Hettema et al. found no relationship between degree of familial vulnerability to schizophrenia and season of birth in an Irish family study[18]. King et al.[19] studied a combination of 184 schizophrenic patients and 60 with affective disorder from Northern Ireland and found excess schizophrenia births in March, April, and May.

On balance then, one could reasonably conclude that 1) in most studies, season of birth does account for a small but significant risk for schizophrenia, 2) two of four studies using Irish samples found a significant association between season of birth and schizophrenia, 3) all but one[19] of the studies using Irish samples were limited to examining schizophrenia and quarter

of birth, as opposed to month of birth, and 3) there is conflicting evidence as to whether association of season of birth and schizophrenia is limited to those without a family history.

If season of birth is associated with risk for schizophrenia, what exactly does this mean? How are the two related? The most probable explanation is that season of birth is a proxy variable for some unmeasured, seasonally varying environmental factor that influences risk in prenatal or early neonatal life; possibilities include meteorological variables such as temperature and light, maternal infections during pregnancy (e.g., influenza, rubella, etc.), and malnutrition[20]. The action of such a risk factor would disrupt the normal maturation of the brain in prenatal and early neonatal life, a concept known as the neurodevelopmental hypothesis[21]. Seasonal variation in exposure to this unmeasured environmental risk factor combined with vulnerability only during prenatal or early neonatal life would translate into the observed excess of schizophrenic births at certain times of the year.

Individual human genetic variation is thought to be an important determinant of the response to various environmental exposures, including infections[22] and nutrition[23]. Thus, it is not unreasonable to expect that genetic and environmental factors might also interact to produce risk for psychiatric disorders, including schizophrenia. This concept is supported by the observation that schizophrenia is known to be associated with various environmental factors (see above), but there is apparently considerable variation in individual response to these environments[24]. In fact, indirect evidence that gene-environment interactions influence risk for schizophrenia has already been adduced, using proxy genetic variables (e.g., positive family history) and environments (reviewed in [24]). In the case of season of birth and schizophrenia, this concept

translates to the hypothesis that different individuals react differently to the exposure indexed by season of birth, depending on their genotype. This particular type of gene-environment interaction is also known as “genetic control of sensitivity to the environment[25].” For example, in the case of maternal infections, this hypothesis would state that certain individuals have genotypes that make them quite vulnerable to the effect of maternal infections during their *in utero* development, while other individuals with different genotypes are relatively invulnerable. In this scenario, the combination of the “wrong” genotype with a “bad” season of birth--a proxy for the likelihood of maternal infection during a critical period of pregnancy--would increase risk for schizophrenia beyond that expected from simply adding the risks of the two factors.

If gene-environment interaction of the type discussed above does underlie the apparent relationship of schizophrenia and season of birth, might this suggest an alternative approach to detecting susceptibility loci? In gene association studies, one typically compares genotype frequencies in a set of cases and controls. If differences are found, this is taken as evidence that gene variants at that locus may confer risk for the phenotype under study. Alternatively, one could restrict the set of schizophrenics to those exposed to the environmental risk factor, i.e., those born in high-risk months. Such an approach would necessarily result in a reduced sample size, since only the exposed cases would be included. However, the loss of power associated with a smaller sample could potentially be more than offset by a gain in power due to increased effect size in a more homogenous subgroup, as would be the case if genotype and phenotype were more strongly associated in exposed individuals. This leads to the following question: under what specific conditions, i.e., combination of sample size, effect size, etc., does use of exposed cases only result in increased overall power to detect genetic associations? Further,

which variables are most important in determining power? Finally, could such an approach actually result in increased evidence for genetic association when applied to real data?

Objectives

1) The Irish Study of High Density Schizophrenia Families (ISHDSF) is a study of high-density schizophrenia families from Ireland and Northern Ireland. This sample has not previously been examined for evidence of excess schizophrenic births in winter-spring. Dates of birth are recorded for nearly all individuals in the sample. Because of the inclusion criteria, schizophrenic individuals in this sample represent familial, as opposed to sporadic, cases of schizophrenia. Given the findings from the literature above, it is of interest to examine whether there is an excess of schizophrenic births in winter-spring months in this sample, especially in light of the controversy as to whether the schizophrenia-season of birth relationship is confined to non-familial cases. Also, monthly birth figures for the general population of Ireland were obtained for use as controls. It is of interest to analyze the data by month, since most previous studies of Irish samples categorized birth by quarter. For these reasons, the first objective of this work is to determine whether season of birth differs between schizophrenics from the ISHDSF and general population controls by month and also by quarter.

2) As discussed above, the most powerful analytic strategy to detecting genetic associations with schizophrenia, assuming gene-environment interaction, is unclear. However, the most powerful strategy will likely depend on specific parameters such as sample size, gene effect size, etc. Computer simulations, incorporating pre-specified event probabilities, can be extremely valuable in uncovering general trends and important factors that determine the power of a given statistical

test. For these reasons, the second objective of this study is to examine, by computer simulation, power to detect the association of a susceptibility locus and schizophrenia under a variety of assumptions, including different models for the interaction of genetic risk and season of birth in determining schizophrenia. It is of particular interest to compare analyses using all cases, and analyses using cases only born in high-risk months.

3) The Irish Case Control Schizophrenia Study (ICCSS) is a case-control study designed to identify genetic loci associated with schizophrenia. Genotypes at multiple marker loci are available for individuals in this sample. Birthdates are available for schizophrenics in the sample. The third objective of this study is to examine six markers for evidence of association to schizophrenia using the two analytic strategies explored in the simulations above, i.e., with all cases and then with cases born in high-risk months. This is meant primarily to illustrate the analytic strategies employed in the simulations using actual, real-world data.

Methods

Season of birth and risk for schizophrenia in Ireland

Design: This case-control study was conducted using the Irish Study of High Density Schizophrenia Families (ISHDSF) to identify cases, and the general population of Ireland, as provided by Ireland's Central Statistics Office (CSO), as controls. It was assumed that the prevalence of schizophrenia in the entire population of Ireland is low enough so that, any relationship that may exist between schizophrenia and season of birth would have little or no effect on statistics calculated using the general population as controls. General population data have been used previously as controls in studies of season of birth and schizophrenia[26].

The study hypothesis was that an excess of schizophrenic births exists in winter-spring months. This study tested this hypothesis by comparing birth frequencies, by month and by quarter, in schizophrenic cases to expected values based on general population controls.

Sample and Data Collection: The source of cases, the ISHDSF, has been described in detail elsewhere[27]. Fieldwork for this study was completed between April, 1987 and November, 1992. The sample was initially conceived for the purposes of genome-wide searches for schizophrenia susceptibility loci. Prior research had indicated that, while genetic factors are a major part of an individual's risk for schizophrenia, their mode of transmission is complex[28]. The genetic architecture of schizophrenia is consistent with either 1) different genes of major effect across different families or 2) multiple, relatively common genes, each with a smaller

effect on vulnerability[27]. Because of this, detection of schizophrenia susceptibility loci was expected to require large numbers of families, ascertained according to standardized procedures. The ISHDSF was designed to accomplish these goals.

Families were ascertained through 39 in-patient psychiatric care facilities serving over 95% of the population of Ireland and Northern Ireland. Hospital records and personal interviews with facility staff (including psychiatrists, nurses, and medical record librarians) were used to identify families with at least two individuals who might have psychotic illness. Approximately 1000 families identified in this way were subjected to an initial screening. The field criterion for follow-up and exhaustive study was families with two or more first-, second-, or third- degree relatives who, according to the field psychiatrist, met DSM-III-R criteria for schizophrenia or poor-outcome schizoaffective disorder (PO-SAD). No exclusion criteria were used. This meant for example, a family might be included that met entry criteria, but also had individuals with alcohol dependence or major depression. The decision to include PO-SAD as a possible inclusion criterion was based on results from family and adoption studies suggesting a common genetic basis for this condition and schizophrenia[29, 30].

From families who met the above preliminary screening criteria, an individual was contacted to provide a family pedigree and also to help decide which family members would be interviewed and sampled for DNA. General rules for these decisions included the following: 1) sample all affected individuals, 2) sample all “connecting” individuals situated between affecteds within the family pedigree, 3) sample all first-degree relatives of affecteds ages 16 and above, and 4) if a key individual is unaffected but cannot be sampled, attempt to sample up to 3 siblings.

Attempts were made to personally interview all traceable and cooperative family members residing in Ireland, Northern Ireland, or cities in England. The study authors estimate they interviewed 88% of traceable living probands and 86% of traceable living relatives.

The field team consisted of psychiatrists and social scientists with experience in mental health or surveys. Whenever possible, a psychiatrist interviewed individuals with suspected psychosis, while social scientists generally interviewed unaffected individuals. Thus, interviewers were not blind to the probable history of psychopathology in interviewees.

The assessment instrument was a modified version of the Structured Interview for DSM-III-R Diagnosis (SCID)[31] for selected disorders including major depression, psychosis and mania; and the Structured Interview for Schizotypy (SIS)[32] for schizophrenia spectrum personality disorders. In almost all cases of schizophrenia or schizoaffective disorder (98.6%), psychiatric in-patient records were obtained and abstracted, as well. Two psychiatrists blinded to pedigree and genotype data then reviewed all available diagnostic information. Diagnostic agreement was excellent with a kappa of 0.94.

On the basis of the best available information from twin and adoption studies, several diagnostic categories were defined for this study, which were intended to reflect the underlying relationship of the various conditions[33]. The definitions of affection were: 1) narrow, to capture “core schizophrenia phenotypes”, 2) intermediate, which in addition to those meeting narrow criteria included individuals with schizotypal personality disorder and all other nonaffective psychotic

disorders, 3) broad, including all the previous, plus individuals with psychotic affective disorders and paranoid, avoidant, and/or schizoid personality disorders. Final inclusion criteria for the ISHDSF used the above diagnostic classification and can be stated as follows: include families with two or more affected first-, second-, or third-degree relatives with an intermediate diagnosis, one or more of whom also met criteria for a narrow diagnosis. The birth month and year were recorded for all individuals in the sample. The final sample included 277 pedigrees with 1,770 individuals. Using the intermediate definition of affected, the sample included 837 affected individuals. Approximately 67% were male and the average age at the time of evaluation was 45.9 years. For the purposes of this study, cases were defined by the intermediate set of criteria described above. Cases defined using the other criteria were also examined, but results were essentially the same as those discussed below.

There are at least two methodological limitations of the ISHDSF that may be relevant to the present study. First, the sample is not truly an “epidemiologic” sample, since ascertainment depended partly on the recollections of health care professionals. Second, interviewers were not blind to knowledge of psychopathology in the subject’s relatives. Interviewers were repeatedly cautioned to discount any such knowledge during their clinical assessments, but this is a potential source of bias from the diagnostic phase of the ISHDSF.

The affected individuals from the ISHDSF constituted the sample of cases. Controls were from the general population of Ireland. Information on numbers of births for each month was obtained from the Central Statistics Office (CSO) of Ireland, covering the years 1976-2000. General population births by quarter were obtained for the years 1900-2000. One unavoidable limitation

of the available data is that the range of birth years for the general population was different from that of the cases, when using birth month data. This was the reason for also examining births by quarter.

By using general population numbers to form the control sample, some individuals with schizophrenia were inevitably included as 'controls'. It was assumed that the prevalence of schizophrenia in the general population was low enough so that this effect was negligible to a first approximation.

Testing for an association between season of birth and schizophrenia: Information on cases was taken from the ISHDSF, as described above. Information on controls was taken from the Irish CSO, as described. A chi square test was used to determine whether there was a difference between the monthly birth frequencies in the schizophrenic and control groups. Expected numbers of births for each month were derived by multiplying the observed monthly birth frequencies for control patients by the total number of schizophrenic patients. The list of observed and expected numbers of monthly schizophrenic births were compared using a chi square test with 11 degrees of freedom.

For each month, the average daily birth frequency was calculated for both groups as follows. First, the fraction of the total group born in a given month was calculated as the count of individuals born in that month divided by the number of individuals in the group. Next, the average daily birth frequency was determined by dividing the fraction born in that month by the number of days in the month. For February, the number of days used was 28.25, to

approximately account for the effect of leap years. This daily birth frequency was graphically plotted to directly compare trends across the year for both groups.

Power Simulations

Design: The gene-environment interaction hypothesis discussed above (see Introduction) raises several questions. Under such a hypothesis, is there any justification for altering the standard approach to searching for disease susceptibility loci? Specifically, in a case-control study designed to test for an association between genotype and schizophrenia, is there any justification for restricting the sample of schizophrenics to those born only in "bad" (i.e. winter-spring) months? Such an approach would necessarily lead to a smaller sample size. However, it is conceivable that the loss of power associated with this could be more than offset by a gain in power due to a more homogenous set of cases. What factors are most critical in determining power to detect genetic associations, assuming gene-environment interaction? In order to gain insight into these questions, Monte Carlo computer simulations were conducted to examine power to detect genetic association under a variety of plausible scenarios and using two different analytic approaches.

These computer simulations all followed the same basic steps: 1) simulate a set of cases and controls, including genotypes and month of birth for each individual, 2) using the genotypes and month of birth, simulate phenotype (i.e., schizophrenia or not) for each individual, 3) conduct a statistical test of association using the simulated case-control data, and record the resulting p-value.

For each set of simulation conditions, these three steps were repeated until 1000 tests of association had been conducted. Power to detect association for a given set of conditions was defined as the fraction of the 1000 statistical tests of association with p-values < 0.05 .

All simulations were scripted using Python, version 2.5.1, which includes a pseudo-random number generating module. The effect of several variables on power to detect genetic associations with schizophrenia was evaluated by simulating data sets assuming different sets of conditions. A condition was defined as one particular combination of sample size, gene effect size, minor allele frequency, and phenotype model (each defined below). A total of 8 conditions were simulated (see Table 1) and each was analyzed using two different approaches, giving a total of 16 estimates of power.

Simulation conditions and procedures: Sample size was either 1000 cases and 1000 controls, or 500 cases and 500 controls. Minor allele frequency (MIF), which here refers to the population frequency of the disease allele, was set to either 0.2 or 0.3, consistent with observed MIFs for many human single nucleotide polymorphisms[34]. Genotype for an individual was simulated by randomly choosing two alleles independently, according to the MIF.

As noted above, month of birth was also simulated for all individuals. To simulate month of birth, an equal probability of being born on any day of the year was assumed. A cumulative probability distribution was constructed to reflect the probability of being born in January, February, March, etc. For all months except February, the probability of being born in that

month was simply the number of days in the month divided by 365. For the month of February, the number of days was set equal to 28.25, to include the effect of leap years. Thus, the probability of being born in February was set as 28.25 divided by 365. Month of birth for an individual was simulated by randomly choosing a month according to this probability distribution.

Once month of birth and genotype were simulated for a given individual, these were combined using a probabilistic model, model1, to determine phenotype:

$$\text{probability of disease} = 0.01 + (\text{number of disease alleles}) * (\text{effect size}) + (\text{bad month}) * (\text{number of disease alleles}) * (\text{effect size})$$

Here, “disease” refers to schizophrenia. All individuals were assumed to have a baseline risk of schizophrenia of 0.01 (approximately equal to the lifetime prevalence of schizophrenia[35]), independent of their month of birth and their genotype at the susceptibility locus. In addition, each disease allele conferred a specific amount of risk. This was termed the gene's effect size; effect sizes used for model1 were 0.001 and 0.002. March, April and May were designated as high-risk birth months (see Results, below). A binary variable called "bad month" was created that took a value of 1 for individuals born in March, April, or May and a value of zero for individuals born in other months. All of these variables were then combined using the equation for model1 to produce a probability of disease.

As an example of the use of model1, consider a simulated individual with one disease allele

(genotype Aa, with 'a' standing for the disease allele) and born in the month of December (thus "bad month" = 0) with a gene effect size of 0.002. The probability of this person having schizophrenia by model1 would then be:

$$0.01 + (1)*(0.002) + (0)*(1)*(0.002) = 0.012$$

The phenotype in this case would be determined by randomly drawing a number between 0 and 1 (all numbers equally likely); if the number were less than or equal to 0.012, this individual would have schizophrenia, otherwise not. Such individuals were simulated until the specified numbers of cases and controls were reached.

We also investigated a variation of this model (model2), where the susceptibility locus only conferred risk on those individuals born in high-risk months. This alternate model for phenotype development can be expressed as:

$$\text{probability of disease} = 0.01 + (\text{bad month}) * (\text{number of disease alleles}) * (\text{effect size})$$

Note that, in this model, there is no elevated risk associated with the disease allele in individuals who are born in "good" months. The effect sizes used in this model were 0.002 and 0.004; these numbers were different from those used in model1. This was done so that individuals in the highest risk category (two disease alleles and born in a bad month) would have the same risk in model1 and model2. Note that both model1 and model2 are probabilistic. This means that the simulated phenotype for an individual with two disease alleles could be "normal," and

conversely, that the simulated phenotype for an individual with no disease alleles could be "schizophrenia."

Analysis of simulated data: Once the specified numbers of cases and controls were simulated, the entire simulated data set was analyzed to test for any evidence of association between genotype and phenotype. Two separate chi square tests were conducted for each set of simulated data. In the first, the observed genotype frequencies across all cases were compared to the expected frequencies based on the controls. In the second test, the observed genotype frequencies in the subset of cases born in "bad" months (March, April, or May) were compared to the expected frequencies. This second test represents an alternative approach to searching for susceptibility loci. Again, the primary purpose of these simulations was to determine if there was any rationale for choosing this second approach over, or at least in addition to, the standard approach (using the entire set of cases and controls). That is, are there circumstances when using only a subset of the cases would actually result in a gain in power to detect genetic associations?

Analysis of the Irish Case Control Study of Schizophrenia

Design: As an illustration of the analytic approach used in the power simulations, and to test the hypothesis that gene-season of birth interaction influences risk for schizophrenia, data from the Irish Case Control Study of Schizophrenia were analyzed (ICCSS). The ICCSS is a sample of schizophrenic and control subjects with genotypes available for several loci within the genome. The strategy employed for this study was to 1) select genotyped loci from individuals in the ICCSS, and 2) analyze these loci for evidence of association with schizophrenia using all cases,

and then using only cases born in March, April, or May. This is exactly the analytic strategy employed in the simulations; the main purpose here was to use this strategy with real data.

Sample: The ICCSS is a sample of schizophrenic and control individuals from Ireland and Northern Ireland, i.e., the same geographic areas as the ISHDSF. Cases were selected from in- and outpatient psychiatric facilities. To be eligible for enrollment, cases had to meet DSM-III-R criteria for schizophrenia or schizoaffective disorder with poor outcome. A blinded, expert reviewer verified all diagnoses. Controls were selected from several sources, including blood donation centers. Controls were eligible if they denied a lifetime history of schizophrenia. Potential subjects were included only if all four of their grandparents were born in Ireland or the United Kingdom. The number of cases and controls with complete data, and thus used in these analyses, varied depending on the marker, but there were approximately 700 cases and 600 controls.

Marker selection and genotyping: A set of six previously genotyped markers for cases and controls was selected for this study. All markers were single nucleotide polymorphisms (SNPs) within or near genes one of three genes: IL3, FBXL21, and CSF2RB. The SNPs included in this analysis are shown in Table 2, along with the names and descriptions of genes they occur in or near.

As noted above, all markers selected for analysis in this study were previously genotyped and analyzed, and details of marker genotyping have been described[36]. Briefly, HapMap data and existing assays developed by Applied BioSystems were used to select markers. SNPs were

selected only if they covered haplotypes with population frequencies $> 1\%$. Genotyping was done using the TaqMan method. Genotypes were scored, and all SNPs were checked for deviations from Hardy Weinberg equilibrium and for Mendelian consistency.

Statistical Analyses: A chi square test was used to evaluate whether there was any evidence for genetic association at each of the markers. For each SNP, the genotype frequencies were derived for cases and controls. Expected genotype frequencies were derived using control genotype frequencies. Observed genotype frequencies were defined as those observed for the cases. For each SNP, a chi square test was performed using all cases, and then a second test was performed using only those cases born in March, April, or May. The reason for choosing these specific months was that excess schizophrenic births were observed in these months using the ISHDSF (see results, *Season of birth and schizophrenia in the ISHDSF*, below). All controls were used for each test. All analyses excluded any cases with missing birthdates or genotypes, and any controls missing genotypes.

Results

Season of birth and schizophrenia in the ISHDSF

The first part of this studied aimed to determine whether there is evidence that season of birth influences risk for schizophrenia using cases from the ISHDSF and general population controls. Table 3 shows, for each month of birth, the observed counts of cases and controls, as well as the expected number of cases based on the observed control frequencies. Comparing the observed and expected counts for cases gave a chi square value of 19.44 (p-value 0.054, 11 df). Note that data on month of birth were available for controls only for the years 1976-2000. The years of birth for the cases, however, ranged from 1893-1973.

Table 4 shows, for each quarter of birth (quarter 1 including January, February, or March as the month of birth; quarter 2 including April, May, or June; etc.), the observed counts of cases and controls, as well as the expected number of cases based on the observed control frequencies. Comparing the observed and expected counts for cases gave a chi square value of 7.48 (p-value 0.058, 3 df). Here, data on quarter of birth for controls was available for the years 1900-2000.

In order to examine the data visually for any possible trends, data for cases and controls were plotted together. Figure 1 shows daily birth frequency by month for cases and controls. Daily birth frequency is the fraction of births in a given month divided by the number of days in that month. This measure was used so that cases and controls would be on a comparable scale, and so that the effect of months with differing lengths would be removed. The figure indicates a trend

toward excess of schizophrenic births in the months of March, April, and May, and also an apparent birth deficit in the months of August, September, and October. Figure 2 similarly shows the daily birth frequency by quarter for cases and controls. Note that the quarters, as defined by the calendar, split the months of excess birth (March in quarter 1, April and May in quarter 2). Still, there seems to be an excess of schizophrenic births in the first two quarters and a concomitant deficit in the second two quarters. As before, month of birth is available for controls covering 1976-2000; quarter of birth is available for controls covering 1900-2000.

As noted in Methods, the ISHDSF had several diagnostic categories for schizophrenia, ranging from narrowly-defined schizophrenia to a much broader phenotype including categories such as paranoid, avoidant and schizoid personality disorder; mood incongruent and mood congruent psychotic affective illness; and delusional disorder. The analyses above were repeated with each diagnostic category. The trends observed were essentially the same as those shown, with excess schizophrenic births in March, April, and May. The lowest p-value (0.014) was for a broad definition of affection, which included schizophrenia, poor-outcome schizoaffective disorder, simple schizophrenia, schizotypal personality disorder, schizophreniform disorder, delusional disorder, atypical psychosis, good-outcome schizoaffective disorder, psychotic affective illness, and paranoid, avoidant, and schizoid personality disorders.

Power simulations

The second part of this study was designed to explore the power to detect genetic associations under a variety of plausible hypotheses regarding how genes and season of birth might interact to

produce risk for schizophrenia. Table 5 shows the results of the power simulations using model1, i.e., where the susceptibility locus increases risk for schizophrenia in all months of birth, but more so for March, April, and May. For analyses with all cases and all controls, power ranges from 42% to 96%. The most significant gains in power appear to be associated with increasing gene effect size (e.g. with 500 cases/controls and $MIF = 0.2$, power increases from 42% to 72% as the gene effect size increases from 0.001 to 0.002). For analyses that restrict cases to those born in March, April, and May, power is always less than the corresponding conditions using the entire set of cases. When cases are restricted to "bad" birth months, power to detect genetic association ranges from 25% to 85%. As with analyses using the full set of data, the most marked jumps in power are associated with increased gene effect size.

Table 6 shows the results of the power simulations using model2, i.e., where the susceptibility locus increases risk for schizophrenia only among those born in March, April, or May. The most marked difference from the simulations using model1 (Table 5) is that power to detect genetic association is almost always greater for the analyses using cases restricted to "bad" birth months. For example, with 1000 cases/controls, a gene effect size of 0.004, and $MIF = 0.2$, power is 82% for when cases are restricted, while power is 48% using all cases. For analyses with all cases and all controls, power ranges from 28% to 51%. For analyses with cases restricted to those born in bad months, power ranges from 29% to 87%. When the gene effect size was 0.002, power using the restricted set of cases was approximately equal to the power using all cases with 500 cases/controls. The difference between the two analytic strategies was substantially greater for a gene effect size of 0.004.

Analysis of the Irish Case Control Study of Schizophrenia

Table 7 contains the results of the analysis of six SNPs for evidence of association with schizophrenia using the ICCSS. Significant evidence of association ($p < 0.05$) was found using all cases for markers rs2284031, rs909486, and rs31555. Significant evidence of association using the subset of cases born in March, April, and May was found for only one SNP, rs31555. Restricting cases to the specified birth months resulted in three instances of an increased chi square value (decreased p-value), for SNPs rs1859427, rs31555, and rs3914025. In one of these instances (rs31555), the p-value decreased from ~ 0.038 to ~ 0.011 , even though the number of cases decreased from 692 to 173.

Discussion

This multi-part study sought to explore, from several perspectives, season of birth as a potential risk factor for schizophrenia. First, this study examined whether there was evidence to suggest that months of birth are different in schizophrenics from the ISHDSF vs. the Irish general population using a case-control design. Second, power simulations were conducted to compare different analytic strategies given different parameter values and assumptions about the interaction of genetics and season of birth to produce risk for schizophrenia. Third, actual genetic data from the ICCSS were analyzed using the strategies studied as part of the simulations.

The first part of this study used a case-control design. Cases were individuals diagnosed with schizophrenia from the ISHDSF. Controls were the general population of Ireland. The chi square test of whether months of birth differed between cases and controls was marginally significant (p -value 0.054). Visual inspection of the data revealed a trend toward increased schizophrenic births in March, April, and May. Taken together, these results suggest that season of birth is a risk factor for schizophrenia for cases in the ISHDSF. The chi square test is relatively insensitive to a positive finding for the following reason: it does not take into account the expected month-to-month correlations in birth rates[17]. More specifically, if birth rate truly does change over the course of the year, it would be expected to gradually rise and fall, as opposed to suddenly rising in one month, falling in the next, etc. The insensitivity of the chi square test to these different possibilities can be understood by considering the following: randomly shuffling months, so that an implausibly jagged rise and fall of schizophrenic birth rates were observed, would give the same chi square value. Thus, a better statistical test would be one that accounted for the expected

inter-month correlations of birth rates. Moreover, while it is certainly possible that any individual month with an excess of schizophrenic births could be a random occurrence, the random consecutive occurrence of excess births across three months is much less likely. Thus, it seems reasonable to conclude that season of birth is different for schizophrenics and controls, with an excess of schizophrenic births in the winter-spring months of March, April, and May. It is interesting that such an effect was observed in light of two previous findings: 1) power analyses which suggest that thousands of subjects are necessary for a reasonable chance at detecting a season of birth effect[37], and 2) findings which suggest that excess winter-spring schizophrenic births are limited to cases without a family history[15].

The case-control approach used above is subject to a number of limitations. First, the control data were taken from aggregate population-level data provided by the Irish CSO. The only information available was the total number of individuals born in Ireland by month for each year from 1976 to 2000, and by quarter from 1900 to 2000. Since this is data for the entire population, it is reasonable to suppose that the number of females is approximately equal to the number of males, and obviously, chronological ages can be determined. Other potential schizophrenia covariates are, however, unavailable. As discussed above, increased risk associated with season of birth is thought to be a proxy for maternal infection or some other seasonally varying toxic environmental exposure. A direct measure of any of these putative factors would be desirable, since one could directly test whether the factor accounts for the observed relationship between season of birth and schizophrenia. Studies of this kind have been conducted[38], but are generally difficult because of time and resource constraints.

Second, we had to assume that the individuals represented by the aggregate population data were not schizophrenic. Because the prevalence of schizophrenia is approximately 1%, this is a reasonable first approximation. The effect of including some schizophrenics in the set of controls should, in fact, make detection of an association between season of birth and schizophrenia more difficult, since such a misclassification would tend to make the control and schizophrenic groups more similar. This design, with general population controls, has been used previously to study schizophrenia and season of birth[26].

Third, for the month-to-month analysis in Figure 1, the controls were born over a different range of years as compared to the schizophrenics. It was assumed that monthly birth frequencies would not vary across time, i.e., that the observed monthly birth frequencies would be the same for a control group drawn from the same range of birth years as the schizophrenic group. To more directly test this assumption, quarter of birth and schizophrenia were examined using the same set of ISHDSF cases and CSO Irish controls born between 1900 and 2000. As shown in Figure 2, the ranges of birth years for the cases and controls still do not exactly match, but the ranges are more closely matched than for the analysis by month. The analysis by quarter produced essentially the same result as the first: a marginally significant p-value (0.058) and a trend toward increased schizophrenic births in quarters 1 and 2 (see Figure 2).

The power simulations involved two distinct models for how genes and season of birth could combine to determine risk for schizophrenia. In the first, model1, the susceptibility gene increased risk for all individuals, but more so for those born in months 3-5. In the second, model2, the susceptibility gene only increased risk for individuals born in months 3-5, but not at

all for individuals born in other months. Under model1, there was greater power to detect genetic associations when all cases are used in the analysis. Under model2, there was generally more power to detect genetic associations when the pool of cases is restricted to those individuals born in months 3-5. This is in spite of the fact that such restriction decreased sample size. Such a finding suggests that in actual studies of genetic association, restricting the set of cases may increase the likelihood of detecting true positive associations. Since the only additional data required for such an analysis would be season of birth, it should be relatively easy to implement. As always, the number of tests should influence one's interpretation of the resulting p-values, and statistical correction for multiple testing is advisable[39]. As with any set of computer simulations, the conclusions from these simulations are, strictly speaking, only applicable to the conditions studied. Furthermore, this approach requires assumptions regarding the relationship of genotype and phenotype, MIF, and gene effect size. Care was taken to ensure that the values used for the simulation conditions were plausible[2, 34].

Based on the findings of the power simulations, data on 6 SNP markers from the ICCSS were analyzed using all cases, and then using cases born in months 3-5. In one of six cases, an initially significant result became more significant when cases were restricted to bad months of birth (for SNP rs31555). As shown in Table 2, this SNP is associated with the gene FBXL21, an F-box containing protein that functions in ubiquitin-mediated proteolysis. The results for the other SNP associated with this gene, rs1859427, were nonsignificant, however, the p-value did decrease from 0.342 to 0.175. Thus, restriction of cases was associated with larger chi square values for both SNPs associated with this gene. There was one other marker that showed an increased chi square value with the restricted set of cases, marker rs3914025 associated with the IL3 gene. The

other marker associated with IL3 and the two remaining markers all had decreased chi square values with the restricted set of cases. Findings of genetic association are often plagued by failed attempts at replication. Thus, these results should be interpreted with a great deal of caution and skepticism. However, these tests do serve as a general illustration of the procedure suggested by the simulations above.

In all the tests for genetic association, all controls were used, regardless of whether the cases were restricted by month of birth or not. The validity of this procedure requires the following assumption: genotype frequencies are the same across all months of birth for the control individuals. Unfortunately, it is not possible to directly test this assumption, since the only information available for most controls was their genotype and the fact that they denied any history of schizophrenia. It is conceivable that this assumption is false, i.e., certain genotypes combined with certain months of birth lead to increased overall mortality in the general population, perhaps via maternal infections.

Conclusion

In summary, this examination of season of birth and schizophrenia suggests that birth in the months of March, April, or May is associated with increased risk for schizophrenia, even in individuals with highly familial schizophrenia. The power simulations demonstrate that power to detect genetic associations may be increased by restricting cases of schizophrenia to those born in months of elevated risk. Specifically, this was the case when a susceptibility locus conferred risk for schizophrenia among those born in high-risk months, but no risk to those born in other months. Finally, analysis of six SNPs for genetic association revealed that for the gene FBXL21, evidence for association was stronger when cases were restricted to those born in high-risk months. For future work, investigators conducting genome-wide association studies for schizophrenia susceptibility loci should consider using subject dates of birth as an “environmental” risk factor, to probe for gene-environment interactions. Such an approach could result in substantially increased power to detect susceptibility loci for schizophrenia.

Tables

Box 1. DSM-III-R criteria for schizophrenia.

1. Characteristic symptoms: Two or more of the following, each present for much of the time during a one-month period (or less, if symptoms remitted with treatment).

- Delusions
- Hallucinations
- Disorganized speech
- Grossly disorganized behavior (e.g., dressing inappropriately, crying frequently) or catatonic behavior
- Negative symptoms— affective flattening (lack or decline in emotional response), alogia (lack or decline in speech), or avolition (lack or decline in motivation)

If the delusions are judged to be bizarre, or hallucinations consist of hearing one voice participating in a running commentary of the patient's actions or of hearing two or more voices conversing with each other, only that symptom is required above. The speech disorganization criterion is only met if it is severe enough to substantially impair communication.

2. Social/occupational dysfunction: For a significant portion of the time since the onset of the disturbance, one or more major areas of functioning such as work, interpersonal relations, or self-care, are markedly below the level achieved prior to the onset.

3. Duration: Continuous signs of the disturbance for at least six months. This six-month period must include at least one month of symptoms (or less, if symptoms remitted with treatment).

Table 1. Simulation conditions.

Parameter	Values used in simulations
Number of cases/controls	500/500 1000/1000
Gene effect sizes	For <i>Model1</i> : 0.001, 0.002 For <i>Model2</i> : 0.002, 0.004
Minor allele frequencies (MIFs)	0.2 0.3
Phenotype models	<i>Model1</i> : disease alleles confer risk for persons born in all months, more risk if born in month 3, 4, or 5 <i>Model2</i> : disease alleles only confer risk for persons born in months 3, 4, or 5.

Table 2. Single nucleotide polymorphism (SNP) markers used in this study.

SNP	Associated gene	Description of associated gene
rs1859427	FBXL21	F-box containing protein, functions in ubiquitin-mediated protein breakdown
rs31555	FBXL21	F-box containing protein, functions in ubiquitin-mediated protein breakdown
rs2069803	IL3	interleukin-3
rs3914025	IL3	interleukin-3
rs2284031	CSF2RB	shared subunit of receptors for IL3, CSF2, and IL5; initiates signal transduction after ligand binding
rs909486	CSF2RB	shared subunit of receptors for IL3, CSF2, and IL5; initiates signal transduction after ligand binding

Table 3. Observed and expected schizophrenia cases by month, using data from the ISHDSF and general population controls from Ireland.

	cases observed	controls observed	control frequencies	cases expected
Jan	65	123252	0.082938778	59.71591995
Feb	56	114937	0.077343445	55.68728046
Mar	79	130623	0.087898874	63.28718894
Apr	72	125996	0.084785271	61.04539521
May	78	131482	0.088476912	63.70337671
Jun	56	124114	0.083518835	60.13356123
Jul	66	128671	0.086585333	62.34143978
Aug	46	123970	0.083421935	60.06379285
Sep	54	125600	0.084518795	60.85353216
Oct	43	122764	0.082610393	59.47948266
Nov	50	115511	0.077729701	55.96538498
Dec	55	119140	0.080171729	57.72364507
Total	720	1486060	1	720

Table 4. Observed and expected schizophrenia cases by quarter, using data from the ISHDSF and general population controls from Ireland.

	cases observed	controls observed	control frequencies	cases expected
Quarter 1	200	1726052	0.249793051	179.850997
Quarter 2	206	1822911	0.263810419	189.9435016
Quarter 3	166	1741833	0.252076867	181.4953441
Quarter 4	148	1619132	0.234319663	168.7101573
Total	720	6909928	1	720

Table 5. Results of power simulations using model1, i.e., disease allele confers risk on all persons, but more risk on persons born in months 3, 4, or 5. Results for analyses with all cases and with cases restricted to those born in high-risk months are shown.

number of cases, controls	gene effect size	minor allele frequency	power: all cases	power: only cases born in mo 3-5
500	0.001	0.2	0.424	0.252
		0.3	0.45	0.294
	0.002	0.2	0.723	0.541
		0.3	0.771	0.625
1000	0.001	0.2	0.56	0.36
		0.3	0.626	0.443
	0.002	0.2	0.915	0.807
		0.3	0.957	0.848

Table 6. Results of power simulations using model2, i.e., disease allele confers risk only on persons born in months 3, 4, or 5. Results for analyses with all cases and with cases restricted to those born in high-risk months are shown.

number of cases, controls	gene effect size	minor allele frequency	power: all cases	power: only cases born in mo 3-5
500	0.002	0.2	0.277	0.289
		0.3	0.279	0.276
	0.004	0.2	0.344	0.575
		0.3	0.375	0.617
1000	0.002	0.2	0.308	0.381
		0.3	0.33	0.428
	0.004	0.2	0.479	0.82
		0.3	0.513	0.874

Table 7. Results of genetic association analyses with six single nucleotide polymorphism (SNP) markers, including all cases and restricting cases to those born in March, April, or May.

SNP marker	Associated gene	Using all cases		Using cases born in March, April, or May	
		Chi square	p-value	Chi square	p-value
rs2284031	CSF2RB	7.379289268	0.024980878	1.291983631	0.524142432
rs909486	CSF2RB	9.99247693	0.00676334	3.369596814	0.18548182
rs1859427	FBXL21	2.147428809	0.341736808	3.484086426	0.175162141
rs31555	FBXL21	6.538706984	0.038031006	8.967530259	0.011290822
rs2069803	IL3	3.977176475	0.136888543	2.41160953	0.299450915
rs3914025	IL3	2.607192431	0.27155347	4.287835042	0.117194829

Figures

Figure 1. Daily birth frequency by month for ISHDSF cases and general population controls.

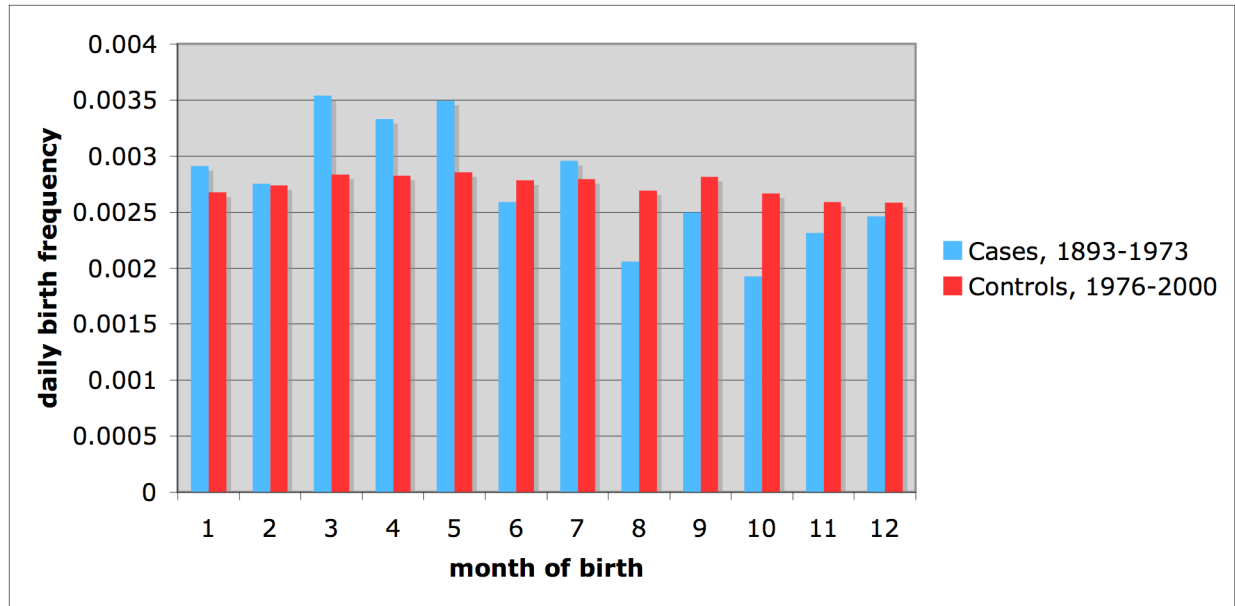
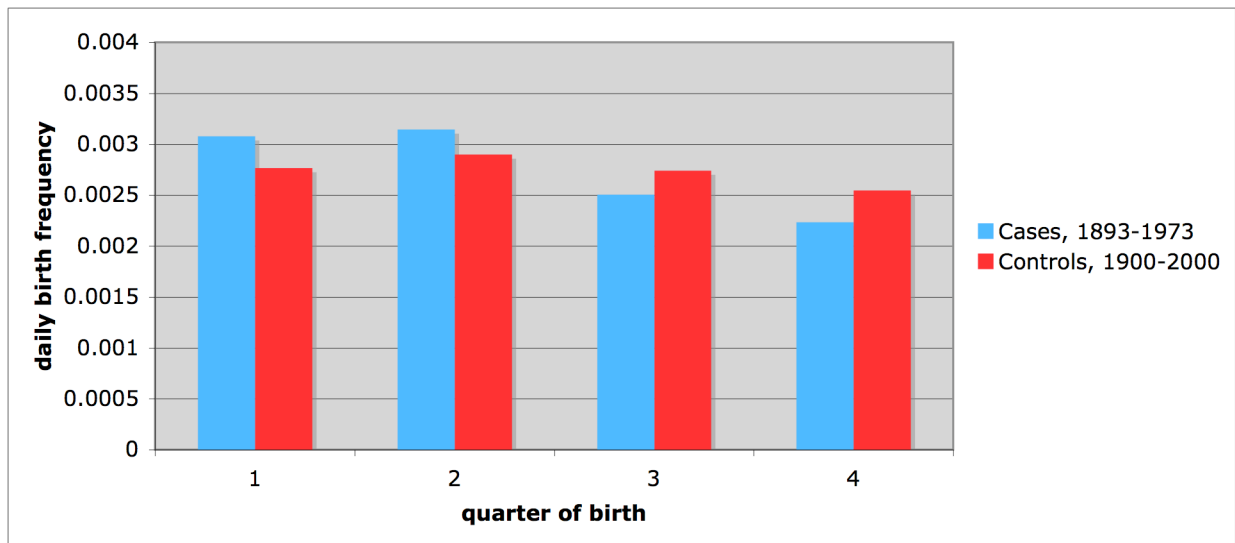


Figure 2. Daily birth frequency by quarter for ISHDSF cases and general population controls.



Appendix. Python script for power simulations.

```
import random, stats

#these are the number of cases (=number of controls)...
sample_sizes = [1000, 500]
#these are effect sizes (see pheno_model below)...
effect_sizes = [0.002, 0.004]
#minor allele frequencies...
mifs = [0.2, 0.3]

#model for phenotype, based on number of disease alleles and season of
birth...
def pheno_model(dz, sob, effect_size):
    return 0.01 + dz*(effect_size) + sob*(dz*(effect_size))

#model for phenotype, based on number of disease alleles and season of
birth...
def pheno_model2(dz, sob, effect_size):
    return 0.01 + sob*(dz*(effect_size))

#bad_months = birth months associated with increased risk of schizophrenia...
bad_months = [3, 4, 5]

daysInYear = 365.0
#days in month; Feb is approximation, based on fact of leap years...
dim = [ 31, 28.25, 31, 30, 31, 30, 31, 31, 30, 31, 30, 31 ]

#create cumulative prob dist for birth month
cp = []
lb, ub = 0.0, 0.0
for m in dim:
    lb = ub
    ub = ub + (m/daysInYear)
    cp.append((lb, min(ub, 1.0)))

#simulates number of disease alleles at a locus, given mif
def sim_genotype(mif):
    r1 = random.random()
    r2 = random.random()
    g1, g2 = 0, 0
    if r1 < mif:
        g1 = 1
    if r2 < mif:
        g2 = 1
    return g1 + g2
```

```

#simulates a birth month according to the cumulative prob distribution
above...
def sim_bd(cp):
    r = random.random()
    for i, (lb, ub) in enumerate(cp):
        if r > lb and r < ub:
            return i + 1
            break

#simulates phenotype, given number of dz alleles and whether person was born
in a 'bad month'...
def sim_pheno(dz, sob):
    pheno = 0
    r = random.random()
    if dz == 0:
        phenocut = 0.01
    else:
        phenocut = pheno_model2(dz, sob, effect_size)
    if r < phenocut:
        pheno = 1
    return pheno

#::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
#simulations start here...

for n in sample_sizes:
    for effect_size in effect_sizes:
        for mif in mifs:

            #make output file...
            outputfilename = 'power3_' + str(n) + '_' + str(effect_size) + '_'
+ str(mif)
            outfile = open(outputfilename, 'w')

            #print sim parameters...
            print >>outfile, 'number of cases (=number controls):', n
            print >>outfile, 'effect size:', effect_size
            print >>outfile, 'minor allele frequency:', mif
            print >>outfile, 'phenotype model: 0.01 + dz*(effect_size) +
sob*(dz*(effect_size))'
            print >>outfile, '\n'

            #starting simulations...
            ctsiggood, ctsigbad, ctsigsep, ctsig, cttotal = 0, 0, 0, 0, 0
            for iter in range(1000):
                #sim controls...
                controls = []
                while len(controls) < n:

```

```

sob = 0
dz = sim_genotype(mif)
mo = sim_bd(cp)
if mo in bad_months:
    sob = 1
pheno = sim_pheno(dz, sob)
if pheno == 0:
    controls.append((dz, mo, sob, pheno))
#sim cases...
cases = []
while len(cases) < n:
    sob = 0
    dz = sim_genotype(mif)
    mo = sim_bd(cp)
    if mo in bad_months:
        sob = 1
    pheno = sim_pheno(dz, sob)
    if pheno == 1:
        cases.append((dz, mo, sob, pheno))

#keys are, for each category, number of bad disease alleles;
values are counts of people in that category with that number of dz alleles
countsCases_goodSOB, countsCases_badSOB, countsCases,
countsControls = {0:0, 1:0, 2:0}, {0:0, 1:0, 2:0}, {0:0, 1:0, 2:0}, {0:0,
1:0, 2:0}

total_badSOB_cases, total_goodSOB_cases = 0, 0
for i, (dz_case, mo_case, sob_case, pheno_case) in
enumerate(cases):
    dz_control, mo_control, sob_control, pheno_control =
controls[i]
    if sob_case == 1:
        countsCases_badSOB[dz_case] =
countsCases_badSOB[dz_case] + 1
        total_badSOB_cases += 1
    elif sob_case == 0:
        countsCases_goodSOB[dz_case] =
countsCases_goodSOB[dz_case] + 1
        total_goodSOB_cases += 1
    countsControls[dz_control] =
countsControls[dz_control] + 1
    countsCases[dz_case] = countsCases[dz_case] + 1

#get expected numbers of cases, based on frequencies of
controls with 0, 1, or 2 disease alleles
expectedCases_goodSOB, expectedCases_badSOB, expectedCases =
[], [], []

for numDzAlleles in [0, 1, 2]:
    freq_controls = countsControls[numDzAlleles] /

```



```

float(n)
                expectedCases_goodSOB.append(freq_controls *
float(total_goodSOB_cases))
                expectedCases_badSOB.append(freq_controls *
float(total_badSOB_cases))
                expectedCases.append(countsControls[numDzAlleles])

                #calculate chisquared statistics...
                #this is for association among just those born in bad months
                chisqbad, pvalbad = stats.lchisquare([countsCases_badSOB[0],
countsCases_badSOB[1], countsCases_badSOB[2]], expectedCases_badSOB)
                #this is for association among just those born in good
months
                chisqgood, pvalgood =
stats.lchisquare([countsCases_goodSOB[0], countsCases_goodSOB[1],
countsCases_goodSOB[2]], expectedCases_goodSOB)
                #this is for association, using all persons, but splitting
into obs/expected counts into good and bad cells
                chisqsep, pvalsep =
stats.lchisquare([countsCases_goodSOB[0], countsCases_goodSOB[1],
countsCases_goodSOB[2], countsCases_badSOB[0], countsCases_badSOB[1],
countsCases_badSOB[2]], expectedCases_goodSOB + expectedCases_badSOB)
                #this is for association, using all persons
                chisqall, pvalall = stats.lchisquare([countsCases[0],
countsCases[1], countsCases[2]], expectedCases)
                if pvalbad < 0.05:
                    ctsigbad += 1
                if pvalgood < 0.05:
                    ctsiggood += 1
                if pvalsep < 0.05:
                    ctsigsep += 1
                if pvalall < 0.05:
                    ctsig += 1
                cttotal += 1
                print >>outfile, [countsCases_badSOB[0],
countsCases_badSOB[1], countsCases_badSOB[2]], [countsCases_goodSOB[0],
countsCases_goodSOB[1], countsCases_goodSOB[2]], [countsControls[0],
countsControls[1], countsControls[2]], pvalbad, pvalgood, pvalsep, pvalall

                #print final tally of significant / nonsignificant tests...
                print >>outfile, ctsiggood, ctsigbad, ctsigsep, ctsig,
cttotal

```

References

1. American Psychiatric Association, American Psychiatric Association. Work Group to Revise DSM-III: *Diagnostic and statistical manual of mental disorders : DSM-III-R*: 3rd , rev ed. Washington, DC: American Psychiatric Association; 1987.
2. Tandon R, Keshavan MS, Nasrallah HA: **Schizophrenia, "just the facts" what we know in 2008. 2. Epidemiology and etiology.** Schizophr Res 2008, **102**(1-3):1-18.
3. Rossler W, Salize HJ, van Os J, Riecher-Rossler A: **Size of burden of schizophrenia and psychotic disorders.** Eur Neuropsychopharmacol 2005, **15**(4):399-409.
4. World Health Organization: *Mental health : new understanding, new hope*: Geneva, Switzerland: World Health Organization; 2001.
5. Saha S, Chant D, Welham J, McGrath J: **A systematic review of the prevalence of schizophrenia.** PLoS Med 2005, **2**(5):e141.
6. McGrath J, Saha S, Welham J, El Saadi O, MacCauley C, Chant D: **A systematic review of the incidence of schizophrenia: the distribution of rates and the influence of sex, urbanicity, migrant status and methodology.** BMC Med 2004, **2**:13.
7. Carter CJ: **Schizophrenia susceptibility genes converge on interlinked pathways related to glutamatergic transmission and long-term potentiation, oxidative stress and oligodendrocyte viability.** Schizophr Res 2006, **86**(1-3):1-14.
8. Torrey EF, Miller J, Rawlings R, Yolken RH: **Seasonality of births in schizophrenia and bipolar disorder: a review of the literature.** Schizophr Res 1997, **28**(1):1-38.
9. Davies G, Welham J, Chant D, Torrey EF, McGrath J: **A systematic review and meta-analysis of Northern Hemisphere season of birth studies in schizophrenia.** Schizophr Bull 2003, **29**(3):587-593.
10. Mino Y, Oshima I: **Seasonality of birth in patients with schizophrenia in Japan.** Psychiatry Clin Neurosci 2006, **60**(2):249-252.
11. Bembenek A, Kociuba L: **Seasonality of birth in schizophrenia patients in Poland.** Schizophr Res 2005, **75**(2-3):447-448.
12. Carrion-Baralt JR, Smith CJ, Rossy-Fullana E, Lewis-Fernandez R, Davis KL, Silverman JM: **Seasonality effects on schizophrenic births in multiplex families in a tropical island.** Psychiatry Res 2006, **142**(1):93-97.
13. O'Hare A, Walsh D, Torrey F: **Seasonality of schizophrenic births in Ireland.** Br J Psychiatry 1980, **137**:74-77.
14. O'Callaghan E, Cotter D, Colgan K, Larkin C, Walsh D, Waddington JL: **Confinement of**

winter birth excess in schizophrenia to the urban-born and its gender specificity. Br J Psychiatry 1995, **166**(1):51-54.

15. O'Callaghan E, Gibson T, Colohan HA, Walshe D, Buckley P, Larkin C, Waddington JL: **Season of birth in schizophrenia. Evidence for confinement of an excess of winter births to patients without a family history of mental disorder.** Br J Psychiatry 1991, **158**:764-769.

16. Roy MA, Flaum MA, Gupta S, Jaramillo L, Andreasen NC: **Epidemiological and clinical correlates of familial and sporadic schizophrenia.** Acta Psychiatr Scand 1994, **89**(5):324-328.

17. Shensky T, Shur E: **Season of birth of schizophrenics: cyclic trends and their analysis.** Br J Psychiatry 1982, **141**:316-317.

18. Hettema JM, Walsh D, Kendler KS: **Testing the effect of season of birth on familial risk for schizophrenia and related disorders.** Br J Psychiatry 1996, **168**(2):205-209.

19. King DJ, Cooper SJ, Earle JA, Martin SJ, McFerran NV, Rima BK, Wisdom GB: **A survey of serum antibodies to eight common viruses in psychiatric patients.** Br J Psychiatry 1985, **147**:137-144.

20. Tochigi M, Okazaki Y, Kato N, Sasaki T: **What causes seasonality of birth in schizophrenia?** Neurosci Res 2004, **48**(1):1-11.

21. Brown AS, Susser ES: **In utero infection and adult schizophrenia.** Ment Retard Dev Disabil Res Rev 2002, **8**(1):51-57.

22. O'Brien RF: **Bacterial vaginosis: many questions--any answers?** Curr Opin Pediatr 2005, **17**(4):473-479.

23. Ordovas JM, Shen J: **Gene-environment interactions and susceptibility to metabolic syndrome and other chronic diseases.** J Periodontol 2008, **79**(8 Suppl):1508-1513.

24. van Os J, Rutten BP, Poulton R: **Gene-environment interactions in schizophrenia: review of epidemiological findings and future directions.** Schizophr Bull 2008, **34**(6):1066-1082.

25. Kendler KS, Eaves LJ: **Models for the joint effect of genotype and environment on liability to psychiatric illness.** Am J Psychiatry 1986, **143**(3):279-289.

26. Modestin J, Ammann R, Wurmle O: **Season of birth: comparison of patients with schizophrenia, affective disorders and alcoholism.** Acta Psychiatr Scand 1995, **91**(2):140-143.

27. Kendler KS, O'Neill FA, Burke J, Murphy B, Duke F, Straub RE, Shinkwin R, Ni Nuallain M, MacLean CJ, Walsh D: **Irish study on high-density schizophrenia families: field methods and power to detect linkage.** Am J Med Genet 1996, **67**(2):179-190.

28. Vogler GP, Gottesman II, McGue MK, Rao DC: **Mixed-model segregation analysis of schizophrenia in the Lindelius Swedish pedigrees.** Behav Genet 1990, **20**(4):461-472.

29. Kendler KS, Gruenberg AM, Tsuang MT: **Subtype stability in schizophrenia.** Am J Psychiatry 1985, **142**(7):827-832.
30. Kendler KS, Gruenberg AM, Kinney DK: **Independent diagnoses of adoptees and relatives as defined by DSM-III in the provincial and national samples of the Danish Adoption Study of Schizophrenia.** Arch Gen Psychiatry 1994, **51**(6):456-468.
31. Spitzer RL, National Institute of Mental Health: *User's guide for the Structured clinical interview for DSM-III-R : SCID:* Washington, DC: American Psychiatric Press; 1990.
32. Kendler KS, Lieberman JA, Walsh D: **The Structured Interview for Schizotypy (SIS): a preliminary report.** Schizophr Bull 1989, **15**(4):559-571.
33. Kendler KS, Diehl SR: **The genetics of schizophrenia: a current, genetic-epidemiologic perspective.** Schizophr Bull 1993, **19**(2):261-285.
34. Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, Mortimore BJ, Willey DL, Hunt SE, Cole CG, Coggill PC, Rice CM, Ning Z, Rogers J, Bentley DR, Kwok PY, Mardis ER, Yeh RT, Schultz B, Cook L, Davenport R, Dante M, Fulton L, Hillier L, Waterston RH, McPherson JD, Gilman B, Schaffner S, Van Etten WJ, Reich D, Higgins J, Daly MJ, Blumenstiel B, Baldwin J, Stange-Thomann N, Zody MC, Linton L, Lander ES, Altshuler D, International SNP Map Working Group: **A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms.** Nature 2001, **409**(6822):928-933.
35. Saha S, Chant D, Welham J, McGrath J: **A systematic review of the prevalence of schizophrenia.** PLoS Med 2005, **2**(5):e141.
36. Chen X, Wang X, Sun C, Chen Q, O'Neill FA, Walsh D, Fanous A, Kendler KS: **FBXL21 association with schizophrenia in Irish family and case-control samples.** Am J Med Genet B Neuropsychiatr Genet 2008, **147B**(7):1231-1237.
37. James WH: **Letter: The power of a test for seasonality of birth with reference to schizophrenia.** Br J Psychiatry 1976, **129**:94-95.
38. Brown AS, Begg MD, Gravenstein S, Schaefer CA, Wyatt RJ, Bresnahan M, Babulas VP, Susser ES: **Serologic evidence of prenatal influenza in the etiology of schizophrenia.** Arch Gen Psychiatry 2004, **61**(8):774-780.
39. Martin LJ, Woo JG, Avery CL, Chen HS, North KE, Au K, Broet P, Dalmaso C, Guedj M, Holmans P, Huang B, Kuo PH, Lam AC, Li H, Manning A, Nikolov I, Sinha R, Shi J, Song K, Tabangin M, Tang R, Yamada R: **Multiple testing in the genomics era: findings from Genetic Analysis Workshop 15, Group 15.** Genet Epidemiol 2007, **31** Suppl 1:S124-31.