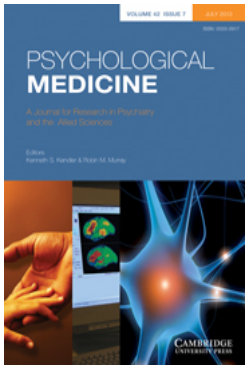
Risk of schizophrenia in relation to parental origin and genome-wide divergence

Pedersen, C. B.; Demontis, D.; Pedersen, M. S.; Agerbo, E.; Mortensen, P. B.; Borglum, A. D.; Hougaard, D. M.; Hollegaard, M. V.; Mors, O.; Cantor-Graae, Elizabeth

Published in:
Psychological Medicine

DOI:
10.1017/S0033291711002376

2012



Risk of schizophrenia in relation to parental origin and genome-wide divergence

C. B. Pedersen, D. Demontis, M. S. Pedersen, E. Agerbo, P. B. Mortensen, A. D. Børglum, D. M. Hougaard, M. V. Hollegaard, O. Mors and E. Cantor-Graae

Psychological Medicine / Volume 42 / Issue 07 / July 2012, pp 1515 - 1521
DOI: 10.1017/S0033291711002376, Published online: 09 November 2011

Psychological Medicine (2012), **42**, 1515–1521. © Cambridge University Press 2011
doi:10.1017/S0033291711002376

ORIGINAL ARTICLE

Risk of schizophrenia in relation to parental origin and genome-wide divergence

C. B. Pedersen[1]*, D. Demontis[2], M. S. Pedersen[1], E. Agerbo[1], P. B. Mortensen[1], A. D. Børglum[2,4], D. M. Hougaard[3], M. V. Hollegaard[3], O. Mors[4] and E. Cantor-Graae[5]

[1] *National Centre for Register-based-Research, Aarhus University, Aarhus, Denmark*
[2] *Department of Human Genetics, Aarhus University, Aarhus, Denmark*
[3] *Section of Neonatal Screening and Hormones, State Serum Institut, Copenhagen, Denmark*
[4] *Centre for Psychiatric Research, University Hospital, Risskov, Denmark*
[5] *Department of Clinical Sciences, Lund University, University Hospital Malmö, Malmö, Sweden*

Background. Second-generation immigrants have an increased risk of schizophrenia, a finding that still lacks a satisfactory explanation. Various operational definitions of second-generation immigrants have been used, including foreign parental country of birth. However, with increasing global migration, it is not clear that parental country of birth necessarily is informative with regard to ethnicity. We compare two independently collected measures of parental foreign ethnicity, parental foreign country of birth *versus* genetic divergence, based on genome-wide genotypic data, to access which measure most efficiently captures the increased risk of schizophrenia among second-generation immigrants residing in Denmark.

Method. A case–control study covering all children born in Denmark since 1981 included 892 cases of schizophrenia and 883 matched controls. Genetic divergence was assessed using principal component analyses of the genotypic data. Independently, parental foreign country of birth was assessed using information recorded prospectively in the Danish Civil Registration System. We compared incidence rate ratios of schizophrenia associated with these two independently collected measures of parental foreign ethnicity.

Results. People with foreign-born parents had a significantly increased risk of schizophrenia [relative risk (RR) 1.94 (95% confidence intervals (CI) 1.41–2.65)]. Genetically divergent persons also had a significant increased risk [RR 2.43 (95% CI 1.55–3.82)]. Mutual adjustment of parental foreign country of birth and genetic divergence showed no difference between these measures with regard to their potential impact on the results.

Conclusions. In terms of RR of schizophrenia, genetic divergence and parental foreign country of birth are interchangeable entities, and both entities have validity with regard to identifying second-generation immigrants.

Received 7 June 2011; Revised 22 September 2011; Accepted 26 September 2011; First published online 9 November 2011

Key words: Genome-wide association, immigrants, principal component, schizophrenia.

Introduction

Studies have found that second-generation immigrants have an increased risk of schizophrenia (Cantor-Graae & Selten, 2005; Cantor-Graae & Pedersen, 2007). However, the factor or factors responsible for this association are unknown. Hypothetical explanations include environmental factors, such as discrimination (Janssen *et al*. 2003; Veling *et al*. 2008), social defeat (Selten & Cantor-Graae, 2007), exposure to viral agents (Sharpley *et al*. 2001; Mortensen *et al*. 2007, 2010), low prenatal vitamin D levels (McGrath, 1999), parental or offspring stress related to migration and post-migration (Fearon & Morgan, 2006; Weiser *et al*. 2008), parental selective migration (Pedersen *et al*. 2011) and parental malnutrition in the country of origin (Susser *et al*. 1996; Hulshoff Pol *et al*. 2000).

Various operational definitions have been used to identify second-generation immigrants. In a study conducted in Israel, Corcoran *et al*. (2009) defined second-generation immigrants as persons having at least one parent born abroad and used parental birth certificates to ascertain the birthplace of the parent. Veling *et al*. (2008) used information obtained from municipal authorities in the Hague to determine ethnicity of the background population data, also defined as at least one foreign-born parent, and also face-to-face patient interviews. Coid *et al*. (2008), in their population-based survey, used a multi-ethnic panel of researchers to compile all available information on ethnicity, including self-ascription, parental

* Address for correspondence: Dr C. B. Pedersen, National Centre for Register-based Research, Aarhus University, Taasingegade 1, 8000 Aarhus C, Denmark.
(Email: cbp@ncrr.dk)

birthplace, with second-generation defined as 'UK-born'. Cantor-Graae & Pedersen (2007) defined second-generation immigrants prospectively as people born in Denmark having at least one parent born abroad, using information from the Danish Civil Registration System. However, with increasing global migration, it is not clear that parental country of birth necessarily is informative with regard to ethnicity. Therefore, to access the validity of the finding of the increased risk of schizophrenia among second-generation immigrants residing in Denmark, we compared two independent definitions of parental foreign ethnicity, foreign parental country of birth *versus* genetic divergence based on genome-wide genotypic data. Genetic divergence refers to allele frequency differences between individuals, reflecting population stratification (allele frequency differences between individuals due to ancestry differences).

Materials and method

The study was based on a linkage between information in the Danish Psychiatric Central Register (Munk-Jorgensen & Mortensen, 1997), the Danish Civil Registration System (Pedersen *et al*. 2006) and genetic information obtained from dried blood spot samples from the Danish Neonatal Screening Biobank (Norgaard-Pedersen & Hougaard, 2007). These blood spots have been systematically stored for individuals born in Denmark since 1 May 1981. We identified all singletons who had been diagnosed with schizophrenia (ICD-10: F20) during the period from 1994 to 2006 and who were born in Denmark on 1 May 1981 or later ($n=1510$). From these singletons, the sample was restricted to cases who had biological material located in the Danish Neonatal Screening Biobank ($n=967$). For each case we randomly selected one control based on the following matching criteria: (*a*) sex; (*b*) exact date of birth; (*c*) born in Denmark; (*d*) alive and with no history of schizophrenia on the date of first diagnosis of schizophrenia of the matched case. We subsequently restricted the study to case–control pairs, where both samples contained sufficient biological material to allow at least four discs to be sampled. Each disc was 3.2 mm in diameter, which corresponds approximately to 3 μl of whole blood. Using this procedure, we identified 915 cases and their 915 individually time-matched controls.

DNA was extracted from the dried blood spots using Extract-N-Amp Blood PCR kit (Sigma Aldrich, USA) and subsequently whole-genome-amplified in triplicate using the REPLIg kit (Qiagen, Germany) (Hollegaard *et al*. 2009). The three separate reactions were pooled before genotyping. Genotyping was performed using the Illumina Human 610-quad beadchip (Illumina Inc., USA). Genotyping was successful for 1775 individuals (892 cases, 883 controls), all individuals had a call rate >0.96 and it was performed consecutively on a matched-pair basis, thereby eliminating potential bias due to differential genotyping procedures for cases and controls.

Among the 582 549 bi-allelic single nucleotide polymorphisms (SNPs) included on the chip, 541 148 SNPs passed quality control, all having a call-rate >0.99, minor allele frequency >0.0015 and no genome-wide significant deviation from the Hardy Weinberg Equilibrium in the controls ($p>0.0001$). The autosomal SNPs that passed quality control were pruned for linkage disequilibrium ($r^2<0.3$) using PLINK (Purcell *et al*. 2007), leaving 128 890 markers for further analyses. We used principal component analyses (PCA) based on the remaining 128 890 markers to quantify genetic divergence, which outputs each individual's coordinates along the principal axes of variation. Principal components were calculated using SAS (McVean, 2009; SAS Institute Inc., 2009) according to the procedure used by Eigenstrat (Patterson *et al*. 2006; Price *et al*. 2006). Missing alleles were sampled at random from a Bernoulli distribution with a parameter that equalled the minor allele frequency. PCA is a standard tool that has been used to investigate population structure in genetics for decades (for a review, see Price *et al*. 2010). There exist several limitations to detect population structure. The top principal components do not necessary reflect population structure. They may instead reflect family relatedness, long-range linkage disequilibrium, or assay artefacts. It is therefore crucial to perform stringent quality control of the data prior to PCA to ensure correct interpretation of the results. In this study, we have applied a very stringent quality control to our samples, which should eliminate the above-mentioned artefacts. Identity by state analysis was performed in order to identify related individuals, PCA was performed on a subset of SNPs pruned for linkage disequilibrium (thereby excluding the effect of long-range linkage disequilibrium) and only high quality SNPs and samples were analysed.

Among the originally identified 915 cases and their 915 individually time-matched controls, we successfully obtained genetic information for both the case and the control among 863 case–control pairs. Since cases without individually matched controls ($n=29$) and controls without individually matched cases ($n=20$) did not contribute information when estimating relative risks, we randomly allocated these individuals to one of the 863 case–control pairs according to almost identical matching criteria to those used previously, i.e. (*a*) sex; (*b*) born within 30 days of the matched case; (*c*) alive and with no

history of schizophrenia on the date of first diagnosis of the matched case. We performed sensitivity analyses showing that this choice had no influence on our results.

Incidence rate ratios were estimated using conditional logistic regression (Andersen *et al*. 1997; King & Zeng, 2002; SAS Institute Inc., 2008) and, due to the matching scheme where each case was compared individually to its matched control, only all incidence rate ratios were controlled for age, sex and date of birth. Confidence intervals and p values were two-sided and based on likelihood ratio tests (Clayton & Hills, 1993). The nested time-matched design has the advantage of estimating incidence rate ratios as opposed to odds ratios (Andersen *et al*. 1997; King & Zeng, 2002). However, incidence rate ratios will be termed relative risks, as these have equivalent interpretations.

The associations between the first two principal components and the relative risk of schizophrenia were estimated, both by including each principal component classified according to deciles of their distribution and as a continuous variable using second degree fractional polynomials (Royston *et al*. 1999). Based on the findings of Novembre *et al*. (2008), showing that the first two principal components mirror geography within Europe, we *a priori* decided to focus on these two principal components.

Using information on parental country of birth recorded in the Danish Civil Registration System, and independently of the genetic data, individuals were classified as both parents born in Denmark or parental foreign country of birth (one or both parents born abroad). Parental foreign country of birth was further subdivided according to the foreign parent's geographical region of birth, i.e. Africa, America, Asia, Europe, Greenland, Middle East and Scandinavia, as previously (Cantor-Graae & Pedersen, 2007). Due to limited power, individuals having both parents born abroad were classified according to the most distant foreign parent's region of birth. Most of those with parental foreign country of birth included in the study had one Danish-born parent (69 Danish-born mothers, 63-Danish born fathers, 70 both parents foreign-born, 34 one or both parents with unknown parental country of birth).

As described above, 40% of cases were not genotyped due to the lack of the necessary amount of blood in the Neonatal Screening Biobank. Participants born early in the study period were less likely to have sufficient biological material for inclusion in the study (results not shown). However, as date of birth is a matching criterion in this study, this cannot bias our results. Furthermore, to evaluate whether our sample was subject to selection bias, we systematically compared the influence of parental history of mental illness (Pedersen & Mortensen, 2001), parental foreign country of birth (Cantor-Graae & Pedersen, 2007), age of parents (Byrne *et al*. 2003), urbanization of place of birth (Pedersen, 2006) in the 863 case–control pairs included in the study *versus* the original risk pairs (1510 cases and 1510 individually matched controls). We found very similar magnitude and direction of all associations when comparing these samples in separate analyses (results not shown), thus ensuring that the obtained sample was not biased with regard to any of these factors.

The identity of the individuals in the study was blinded to the investigators and the study did not involve contact with individual patients. The study therefore did not need approval from the ethics committee according to Danish laws, but the project was approved by the Danish Data Protection Agency and the Steering Committee of the Danish Neonatal Screening Biobank.

Results

The principal component scatterplot shows a clear correspondence between the first two principal components based on the genome-wide SNP genotypic data and parental country of birth as registered in the Danish Civil Registration System (Fig. 1). It should be remembered that only people born in Denmark were included in the study. Also, Danish-born people who had parents born in Scandinavian countries clustered together with Danish-born people with Danish-born parents, as expected.

Relative risks of schizophrenia according to the first principal component

Fig. 2 shows the relative risk of schizophrenia (left-hand y-axis, black lines) according to deciles of the first principal component. Individuals in the lowest decile of the distribution were chosen as the reference category. Individuals in the lowest 60% of the distribution of the first principal component had identical relative risks of schizophrenia (Fig. 2, black lines). Among the remaining individuals, the risk of schizophrenia increased with increasing genetic divergence ($p = 0.0002$, Fig. 2). The association between the first principal component and the risk of schizophrenia was not reducible to a trend including the first principal component in the model as one continuous variable ($p = 0.004$), but it was reducible to a fractional polynomial ($p = 0.85$, Fig. 2, red line), which was also highly significant ($p < 0.0001$). In addition, there was a weak significant association between the risk of schizophrenia and the second principal components ($p = 0.03$). However, this association was explained

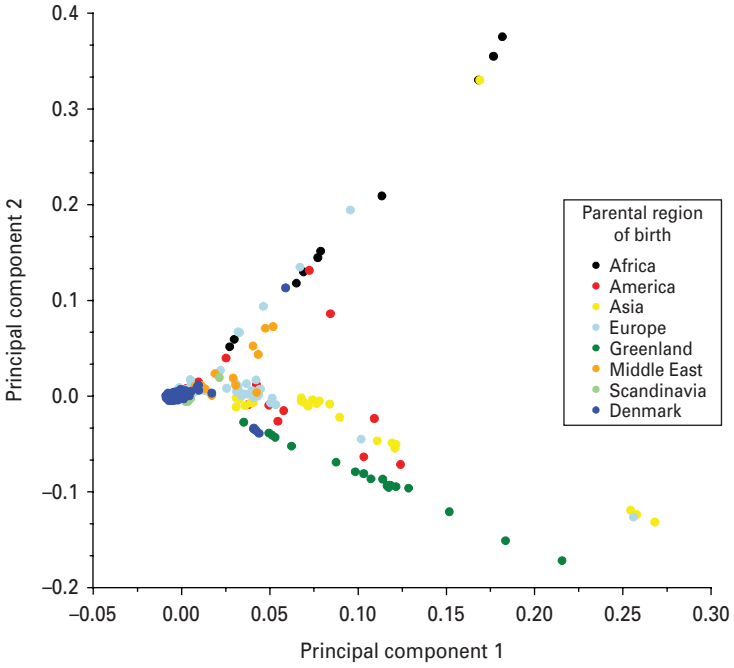


Fig. 1. Scatterplot of the first two principal components according to parental region of birth.

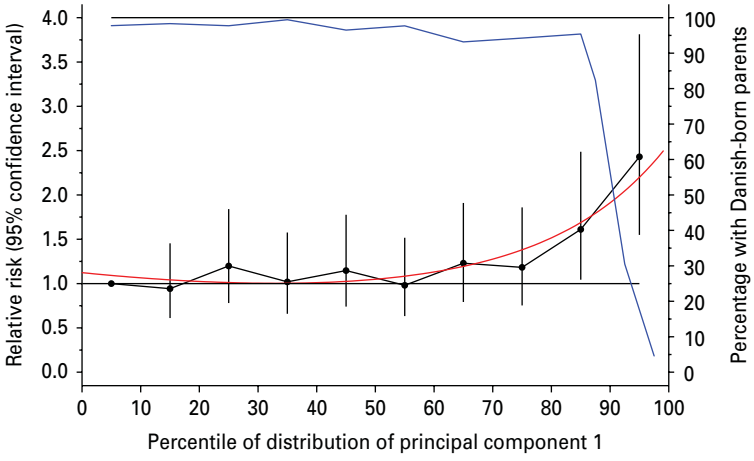


Fig. 2. The left-hand y-axis shows the relative risk of schizophrenia according to the first principal component (black and red lines). Vertical lines indicate 95% confidence intervals. Individuals in the lowest decile of the distribution were chosen as the reference category for the categorical analyses (black lines). The red line shows the fitted second-degree fractional polynomial. Both for the categorical (black line) and the continuous model (red line), there was a significant association between the relative risk of schizophrenia and the first principal component ($p=0.0002$ *v.* $p<0.0001$). The right-hand y-axis shows the percentage of Danish-born people with Danish-born parents (blue line) according to the first principal component.

entirely when adjusting for either parental foreign country of birth ($p=0.40$) or the first principal component ($p=0.58$).

There was a very strong association between the first principal component and parental foreign country of birth (Fig. 2, right-hand y-axis, blue line). Among people in the lowest decile of the distribution of the first principal component, 97.7% had Danish-born parents. This fraction decreased only slightly until the 85th percentile of the distribution. People with higher values of the first principal component were most often of foreign-born parentage.

Relative risk according to parental foreign country of birth

People of foreign-born parents had a significantly 1.92-fold increased risk of schizophrenia as compared with people of Danish-born parents ($p<0.0001$, Table 1). Table 1 also shows the median of the first

Table 1. *Relative risks (RR) of schizophrenia according to parental country of birth*

Parental country of birth	Cases	Controls	Median PC1[a]	RR[b]	95% CI
Total	892	883	−5.19		
Overall[c]					
Denmark	739	800	−5.41	1.00	Ref.
Any foreign country	129	73	25.64	1.92	1.41–2.63
Unknown	24	10	−4.27	2.64	1.24–5.59
					$p<0.0001$
Detailed classification[c]					
Denmark	739	800	−5.41	1.00	Ref.
Scandinavia	15	13	−4.59	1.26	0.59–2.67
Europe	53	30	5.55	1.90	1.20–3.02
Africa	7	5	73.10	1.37	0.43–4.36
Australia	1	0	−0.74	–	
North and South America	14	6	31.67	2.67	1.01–7.05
Middle East	6	6	29.91	1.17	0.37–3.71
Asia	16	9	74.19	1.97	0.85–4.57
Greenland	17	4	103.22	4.88	1.62–14.7
Unknown	24	10	−4.27	2.64	1.24–5.59
					$p=0.0003$

CI, Confidence interval.

Estimates were based on people born in Denmark 1981–1995 who developed schizophrenia 1994–2006.

[a] Median of the first principal component (PC1) ($\times 1000$) based on the genome-wide association data.

[b] Estimates of RR were controlled for age, sex and calendar period using the individually time-matched design.

[c] Individuals with unknown father (i.e. unknown paternal place of birth) were treated in a separate category in the analyses (24 cases and 10 controls). The p values do not include the effect of unknown father.

principal component according to parental region of birth. People with parents born in Denmark and Scandinavia had almost identical values, whereas people with parents born in Greenland, Asia and Africa were the most genetically divergent.

To assess which measure of parental foreign ethnicity best captured the increased risk associated with foreign paternal ethnicity, the effects of genetic divergence and foreign parental country of birth were adjusted mutually. After mutual adjustment, the effects of both measures of foreign parental ethnicity were reduced to a level below statistical significance ($p=0.40$ *v.* $p=0.28$). As a consequence, in terms of the relative risks obtained, these two independently collected measures of parental foreign ethnicity are interchangeable.

We performed additional analyses that showed that our results were robust to the normalization of the number of variants used by Eigenstrat and to using all 541 148 markers to calculate principal components as opposed to the 128 890 markers pruned for linkage disequilibrium.

Discussion

To investigate which measure of paternal foreign ethnicity most accurately captures the increased risk of schizophrenia among second-generation immigrants, we used two independent sources of information: (1) parental place of birth as recorded prospectively in the Danish Civil Registration System; (2) genetic divergence based on genome wide genotypic data. We found that the increased risk associated with foreign parental birth was reduced to a level well below significance when adjusting for genetic divergence, as quantified by the first principal component (Table 1). But we also found that the increased risk associated with extreme genetic divergence was reduced to a level well below statistical significance when adjusting for foreign parental country of birth. As a consequence, these two independently collected measures of parental foreign ethnicity are interchangeable. This is the first study to compare two competing measures of foreign parentage and their influence on the risk of schizophrenia.

Previously, PCA have been used with success in order to assign an individual to a geographic birth location based on genetic information, demonstrating that genetic correlation between pairs of individuals decays with distance in Europe (Novembre *et al*. 2008). As expected for this study, Danish-born people with foreign-born parents differ genetically from Danish-born people with Danish-born parents. Our data also support the findings of Novembre *et al*. (2008), with

individuals with a parent born in Scandinavia clustering with people having Danish-born parents and that individuals with a parent born in Asia, Greenland or Africa were among the most distantly clustering individuals (Table 1, Fig. 1). Our results therefore demonstrate that genetic divergence and parental foreign country of birth are closely related entities and that both entities seem to have validity with regard to the ability to distinguish genetically heterogeneous groups of people. Correlation between genetics and measures of ethnicity based on oral or written information has also been found in a study of people in Qatar, where surname origin correlated with genetic ancestry (Hunter-Zinck *et al*. 2010).

In a larger population-based study including the total Danish population, we observed an increased risk of schizophrenia among Danish-born people with foreign-born parents. This increased risk was virtually independent of parental foreign country of origin (Cantor-Graae & Pedersen, 2007). In this study, the risk of schizophrenia increased with increasing genetic divergence. These findings taken together suggest that genetic divergence and foreign parental birth are both poor proxy variables of the true risk-increasing mechanism responsible for the increased risk of schizophrenia among Danish second-generation immigrants. Supporting the above statement, it should be noted that Danish-born people with a parent born in a Scandinavian country have an increased risk of schizophrenia while genetically they are almost identical to native Danes (Table 1, Figs 1, 2).

Strength and limitations

Our study was based on people admitted or in outpatient contact with a diagnosis of schizophrenia (F20), as registered in the Danish Psychiatric Central Register, which has shown a high diagnostic validity (Jakobsen *et al*. 2005). In addition to genetic information, we also had information on other potential confounders, such as parental country of birth. Although only 60% of the individuals with schizophrenia had sufficient biological material to be included in the study, we demonstrated that, whatever the cause of this high drop-out, it is very unlikely to have biased our findings. As opposed to most other genome-wide genotypic data generated in relation to schizophrenia, all our blood samples were stored, treated and analysed consecutively on a matched-pair basis. As a consequence, it is extremely unlikely that our findings are caused by bias from systematic differences in the genetic information on cases and controls. A limitation of the study is that, except for Danish-born people with a parent born in Europe or Scandinavia, we only had limited power to perform more detailed analyses of exact parental country of birth.

Information on parental country of birth originates from the Danish Civil Registration System, which has recorded this information for all Danish residents prospectively from 1968 onwards. There exist no studies evaluating the quality of the information recorded, which is generally believed to be of high quality (Pedersen *et al*. 2006).

In conclusion, Danish-born people with foreign-born parents and genetically divergent persons had a significant increased risk. In terms of relative risk of schizophrenia, genetic divergence and parental foreign country of birth are interchangeable entities and both entities have validity with regard to identifying second-generation immigrants.

Acknowledgements

This study was supported by The Stanley Medical Research Institute and the Danish Strategic Research Council. The sponsors had no role in the design of the study, in analysis and interpretation of the data or in the writing of the manuscript.

Declaration of Interest

None.

References

Andersen PK, Borgan Ø, Gill RD, Keiding N (1997). *Statistical Models Based on Counting Processes*. Springer: New York.

Byrne M, Agerbo E, Ewald H, Eaton WW, Mortensen PB (2003). Parental age and risk of schizophrenia: a case–control study. *Archives of General Psychiatry* **60**, 673–678.

Cantor-Graae E, Pedersen CB (2007). Risk of schizophrenia in second-generation immigrants: a Danish population-based cohort study. *Psychological Medicine* **37**, 485–494.

Cantor-Graae E, Selten JP (2005). Schizophrenia and migration: a meta-analysis and review. *American Journal of Psychiatry* **162**, 12–24.

Clayton D, Hills M (1993). *Statistical Models in Epidemiology*. Oxford University Press: Oxford, New York, Tokyo.

Coid JW, Kirkbride JB, Barker D, Cowden F, Stamps R, Yang M, Jones PB (2008). Raised incidence rates of all psychoses among migrant groups: findings from the East London first episode psychosis study. *Archives of General Psychiatry* **65**, 1250–1258.

Corcoran C, Perrin M, Harlap S, Deutsch L, Fennig S, Manor O, Nahon D, Kimhy D, Malaspina D, Susser E (2009). Incidence of schizophrenia among second-generation immigrants in the Jerusalem Perinatal Cohort. *Schizophrenia Bulletin* **35**, 596–602.

Fearon P, Morgan C (2006). Environmental factors in schizophrenia: the role of migrant studies. *Schizophrenia Bulletin* **32**, 405–408.

Hollegaard MV, Grauholm J, Borglum A, Nyegaard M, Norgaard-Pedersen B, Orntoft T, Mortensen PB, Wiuf C, Mors O, Didriksen M, Thorsen P, Hougaard DM (2009). Genome-wide scans using archived neonatal dried blood spot samples. *BMC Genomics* **10**, 297.

Hulshoff Pol HE, Hoek HW, Susser E, Brown AS, Dingemans A, Schnack HG, van Haren NE, Pereira Ramos LM, Gispen-de Wied CC, Kahn RS (2000). Prenatal exposure to famine and brain morphology in schizophrenia. *American Journal of Psychiatry* **157**, 1170–1172.

Hunter-Zinck H, Musharoff S, Salit J, Al-Ali KA, Chouchane L, Gohar A, Matthews R, Butler MW, Fuller J, Hackett NR, Crystal RG, Clark AG (2010). Population genetic structure of the people of Qatar. *American Journal of Human Genetics* **87**, 17–25.

Jakobsen KD, Frederiksen JN, Hansen T, Jansson LB, Parnas J, Werge T (2005). Reliability of clinical ICD-10 schizophrenia diagnoses. *Nordic Journal of Psychiatry* **59**, 209–212.

Janssen I, Hanssen M, Bak M, Bijl RV, de Graaf R, Vollebergh W, McKenzie K, van Os J (2003). Discrimination and delusional ideation. *British Journal of Psychiatry* **182**, 71–76.

King G, Zeng L (2002). Estimating risk and rate levels, ratios and differences in case-control studies. *Statistics in Medicine* **21**, 1409–1427.

McGrath J (1999). Hypothesis: is low prenatal vitamin D a risk-modifying factor for schizophrenia? *Schizophrenia Research* **40**, 173–177.

McVean G (2009). A genealogical interpretation of principal components analysis. *PLoS Genetics* **5**, e1000686.

Mortensen PB, Norgaard-Pedersen B, Waltoft BL, Sorensen TL, Hougaard D, Torrey EF, Yolken RH (2007). Toxoplasma gondii as a risk factor for early-onset schizophrenia: analysis of filter paper blood samples obtained at birth. *Biological Psychiatry* **61**, 688–693.

Mortensen PB, Pedersen CB, Hougaard DM, Norgaard-Petersen B, Mors O, Borglum AD, Yolken RH (2010). A Danish National Birth Cohort study of maternal HSV-2 antibodies as a risk factor for schizophrenia in their offspring. *Schizophrenia Research* **122**, 257–263.

Munk-Jorgensen P, Mortensen PB (1997). The Danish Psychiatric Central Register. *Danish Medical Bulletin* **44**, 82–84.

Norgaard-Pedersen B, Hougaard DM (2007). Storage policies and use of the Danish Newborn Screening Biobank. *Journal of Inherited Metabolic Disease* **30**, 530–536.

Novembre J, Johnson T, Bryc K, Kutalik Z, Boyko AR, Auton A, Indap A, King KS, Bergmann S, Nelson MR, Stephens M, Bustamante CD (2008). Genes mirror geography within Europe. *Nature* **456**, 98–101.

Patterson N, Price AL, Reich D (2006). Population structure and eigenanalysis. *PLoS Genetics* **2**, e190.

Pedersen CB (2006). No evidence of time trends in the urban-rural differences in schizophrenia risk among five million people born in Denmark from 1910 to 1986. *Psychological Medicine* **36**, 211–219.

Pedersen CB, Gotzsche H, Moller JO, Mortensen PB (2006). The Danish Civil Registration System. A cohort of eight million persons. *Danish Medical Bulletin* **53**, 441–449.

Pedersen CB, Mortensen PB (2001). Family history, place and season of birth as risk factors for schizophrenia in Denmark: a replication and reanalysis. *British Journal of Psychiatry* **179**, 46–52.

Pedersen CB, Mortensen PB, Cantor-Graae E (2011). Do risk factors for schizophrenia predispose to emigration? *Schizophrenia Research* **127**, 229–234.

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics* **38**, 904–909.

Price AL, Zaitlen NA, Reich D, Patterson N (2010). New approaches to population stratification in genome-wide association studies. *Nature Review Genetics* **11**, 459–463.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* **81**, 559–575.

Royston P, Ambler G, Sauerbrei W (1999). The use of fractional polynomials to model continuous risk variables in epidemiology. *International Journal of Epidemiology* **28**, 964–974.

SAS Institute Inc. (2008). The PhReg procedure. In *SAS/STAT 9.2 User's Guide*. SAS Institute Inc.: Cary, NC.

SAS Institute Inc. (2009). *SAS 9.2 Language Reference: Dictionary*, 2nd edn. SAS Institute Inc.: Cary, NC.

Selten JP, Cantor-Graae E (2007). Hypothesis: social defeat is a risk factor for schizophrenia? *British Journal of Psychiatry* (Supplement) **51**, s9–s12.

Sharpley M, Hutchinson G, McKenzie K, Murray RM (2001). Understanding the excess of psychosis among the African-Caribbean population in England. Review of current hypotheses. *British Journal of Psychiatry Supplement* **40**, s60–s68.

Susser E, Neugebauer R, Hoek HW, Brown AS, Lin S, Labovitz D, Gorman JM (1996). Schizophrenia after prenatal famine. Further evidence. *Archives of Geneneral Psychiatry* **53**, 25–31.

Veling W, Susser E, van Os J, Mackenbach JP, Selten JP, Hoek HW (2008). Ethnic density of neighborhoods and incidence of psychotic disorders among immigrants. *American Journal of Psychiatry* **165**, 66–73.

Weiser M, Werbeloff N, Vishna T, Yoffe R, Lubin G, Shmushkevitch M, Davidson M (2008). Elaboration on immigration and risk for schizophrenia. *Psychological Medicine* **38**, 1113–1119.