Published in final edited form as:

Am J Psychiatry. 2010 July ; 167(7): 854–865. doi:10.1176/appi.ajp.2010.09071050.

The internet-based MGS2 control sample: Self report of mental illness

Alan R. Sanders, M.D.¹, Douglas F. Levinson, M.D.², Jubao Duan, Ph.D.¹, J. Michael Dennis, Ph.D.³, Rick Li, Ph.D.³, Kenneth S. Kendler, M.D.⁴, John P. Rice, Ph.D.⁵, Jianxin Shi, Ph.D.², Bryan J. Mowry, M.D.⁶, Farooq Amin, M.D.⁷, Jeremy M. Silverman, Ph.D.⁸, Nancy G. Buccola, A.P.R.N., B.C.⁹, William F. Byerley, M.D.¹⁰, Donald W. Black, M.D.¹¹, Robert Freedman, M.D.¹², C. Robert Cloninger, M.D.⁵, and Pablo V. Gejman, M.D.¹

¹Center for Psychiatric Genetics, Department of Psychiatry and Behavioral Sciences, NorthShore University HealthSystem, Evanston, Illinois

²Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, California

³Government & Academic Research, Knowledge Networks, Menlo Park, California

⁴Departments of Psychiatry, and Human Genetics, Virginia Commonwealth University, Richmond, Virginia

⁵Department of Psychiatry, Washington University, St. Louis, Missouri

⁶Queensland Centre for Mental Health Research, and Queensland Brain Institute, University of Queensland, Brisbane, Queensland, Australia

⁷Department of Psychiatry and Behavioral Sciences, Atlanta Veterans Affairs Medical Center, and Emory University, Atlanta, Georgia

⁸Department of Psychiatry, Mount Sinai School of Medicine, New York, New York

⁹School of Nursing, Louisiana State University Health Sciences Center, New Orleans, Louisiana

¹⁰Department of Psychiatry, University of California at San Francisco, San Francisco, California

¹¹Mental Health Clinical Research Center, and Department of Psychiatry, University of Iowa Carver College of Medicine, Iowa City, Iowa

¹²Department of Psychiatry, University of Colorado Denver, Aurora, Colorado

Abstract

Objective—The Molecular Genetics of Schizophrenia (MGS2) project recruited an adult control sample of non-Hispanic European-ancestry (EA, 3,364) and African American (AA, 1,301) subjects.

Methods—Subjects gave consent to deposit phenotypic data and blood samples into a repository for general research use, with full anonymization of the sample. We compared the control sample

Address for correspondence and reprints: Dr. Alan R. Sanders, NUH Research Institute, 1001 University Place, Evanston, IL 60201-3137. Telephone: 224-364-7560. Fax: 224-364-7570. asanders@northshore.org.

Disclosures: The authors report no competing interests.

with population census data for demographic data and with previous population surveys for anthropometrics and prevalences of psychiatric disorders as estimated by internet-administered questionnaire.

Results—The full MGS2 control sample includes 4,665 subjects (3,364 EA and 1,301 AA), of whom 3,626 were included in the MGS2 genome-wide association study (GWAS). The sample is generally demographically representative of the U.S. population, except for being older and more female, educated, and affluent, although all strata are represented. Self-reported ancestry was consistent with genotypic and census data. Lifetime prevalences for depressive, anxiety, and substance diagnoses were higher than in previous population-based surveys, probably due to use of an abbreviated self-report instrument. However, patterns such as sex ratios, comorbidity and demographic associations were consistent with previous reports. DNA quality for the internet collected/evaluated control sample was comparable to that of the face-to-face case sample.

Conclusions—The internet-based methods facilitated the rapid collection of large and anonymized EA and AA control samples that have been validated as being generally representative for many aspects of demography, ancestry, and morbidity. Utilization of clinical screening data shared with the scientific community may permit investigators to select appropriate controls for some studies.

INTRODUCTION

MGS2 recruited an adult control sample for genetic research. We now provide the first complete description of the final sample, including clinical and demographic features in relation to the U.S. population, ancestry, and quality control (QC) of genotypic data. We have previously reported on some of these issues for parts of the sample [1, 2].

During 2004–2007, a survey and marketing research company, Knowledge Networks (KN; Menlo Park, CA), internet recruited self-identified non-Hispanic EA or AA participants consenting to deposition of anonymized clinical data, a blood sample, and future genetic data in the NIMH repository. We note that the Wellcome Trust Case Control Consortium (WTCCC, [3]) has successfully used a single control sample for studies of diverse medical and psychiatric disorders, and that analyses gave similar results using population-based and opportunistically-ascertained subsets of the control subjects (see review, [4]).

For psychiatric genetic studies, the recruitment and assessment of control samples present several interesting challenges including representativeness (in relation to the general population and to the case sample), cost-effectiveness, sample size, clinical screening approach (none, brief, full; internet, telephone, in-person; questionnaire, interview), and selection stringency (no psychiatric diagnosis, specific exclusions, unselected) (see reviews, [5–7]). We were particularly interested in providing a way to select control subjects for specific studies. For genetic studies of common disorders (including depression, anxiety, substance use), validity is enhanced when most individuals with the disorder under study can be excluded from the control group [4, 8], especially for prevalence 20% and with only a modest benefit for prevalence 5%. For a rarer disorder like schizophrenia, power is essentially unchanged if 1–2% of control subjects have or develop the disorder [4]. We therefore designed an online assessment to collect information relevant to subject selection.

Since the selection of appropriate controls is in large part dependent on the source population of the cases of a particular association study, we cannot provide a universal algorithm, but do note that the ideal is to draw from the same source population and independently of exposure status (marker genotype) [9, 10].

METHODS

Ascertainment

KN enrolled a total of 4,665 adults (18 years or older), 3,364 EA and 1,301 AA. All EA and 529 AA control subjects were recruited from KN's nationwide panel of survey participants, which is recruited continuously by random digit dialing (RDD) [11, 12] and resembles the population distribution in age, race, Hispanicity, region, employment, marital, and other demographic features (see Table S4 in [1] and Table 1 here). An additional 772 AA controls were recruited by subcontract to Survey Sampling International (SSI; Shelton, CT), using internet recruitment (banner ads on websites). Supplementary Table 1 (ST1) lists the completion rate for the various stages of the study, from invitation to procedure completion. Of the targeted panel members, 15.5% (AA) to 21.7% (EA) of individuals consented and completed the online questionnaire, and gave blood – for comparison, a population-based study requiring blood that used significantly more enrollment effort (2 letters and 3 phone contacts) achieved a 31% rate [13]. SSI completed 2.1% of targeted controls, but this was needed to achieve an adequate AA sample size.

Informed consent

Institutional review board approval was obtained at NorthShore University HealthSystem. Participants consented online to use of their DNA and phenotypic information for the study of any illness or trait, and then signed an identical hard-copy consent at the time of venipuncture. While it is possible to identify a subject in an anonymized dataset with extensive genotypic data, such as this dataset, if one also had access to a large number of genotypes and to subjects' identities in a second dataset [14], there is no known such anonymity violation to date for any genetic study. For genotypic data, NIMH restricts access to qualified investigators and institutions who agree to NIH GWAS data sharing policies (grants.nih.gov/grants/gwas/) which prohibit any attempt to identify individual research subjects or even their presence in a sample.

Assessment

All participants completed an online self-report clinical assessment (nimhgenetics.org, Available Data → Controls → Control Questionnaire), with most items screening for common adult psychiatric disorders. The validated assessment tool in large-scale epidemiological studies of these disorders of widest use is the Composite International Diagnostic Interview (CIDI) [15, 16]. The CIDI was developed for direct interviews, but an abbreviated version, the CIDI-Short Form (CIDI-SF), has been used for self-report (sometimes modified for lifetime assessments, e.g., see [17]) and telephone interviews, and is accurate compared to a full CIDI [18, 19]. We therefore asked control subjects to complete the CIDI-SF, modified to screen for lifetime diagnoses, and also included other questionnaire components assessing various traits and disorders summarized in Table 2.

Phlebotomy

Individuals who completed the questionnaire were then contacted by Examination Management Services Inc. (EMSI; Irving, TX) to arrange blood specimen collection. The phlebotomist confirmed gender and race (EA or AA) and shipped the blood to the Rutgers University Cell and DNA Repository (RUCDR; Piscataway, NJ) for DNA extraction and transformation to lymphoblastic cell lines (99.5% success rate).

Diagnosis

We scored the dichotomous presence/absence of individual disorders according to the CIDI-SF [18] scoring memo [20]. Most participants were past the typical ages of onset for the assessed psychiatric disorders (mean ages, EA=50.0 and AA=45.4) via DSM-IV [21]. Dependence is assessed separately for alcohol, but the drug dependence section asks which substances have been used and then queries dependence criteria for “the year when you used one or more of these drugs on your own the most” rather than for each substance individually. CIDI-SF ignores several exclusion criteria, e.g., whether a major depressive episode (MDE) or anxiety disorder might have been due to a general medical condition or substance use. Omission of full exclusion criteria evaluation should lead to an increased number of these diagnoses (but see discussion below of CIDI-SF dichotomous versus probabilistic methods, and disorder specific criteria discussions).

Ancestry

The questionnaire allowed controls to select 1 ancestry (race/ethnicity) for each grandparent, from fifteen categories (i.e., Figure 1 x-axis labels, plus “other” with a text entry, or “not sure”) used in the NIMH Genetics Initiative’s Diagnostic Interview for Genetic Studies (DIGS) [22]. The “proportion” of each ancestry was not queried. Most “other” entries were the name of a country that was part of a category that had already been selected (e.g., “France” when the category “West Europe” had already been reported). Therefore, we have not translated the “other” entries (<5% of the total) into the fifteen categories for the present analysis.

Quality Control of Genotypic Data

Initially, we genotyped fourteen schizophrenia candidate genes in 2,126 EA control and 1,952 EA case individuals [1]. Extensive sample QC (evaluating genotyping call rates, sex typing discrepancies, cryptic duplicates or relatives, and outliers in an analysis of ancestry-informative markers, AIMs) left 2,002 (94%) control and 1,870 (96%) case samples remaining for analysis, i.e., very similar proportions. Control samples had more unresolved sex-typing discrepancies (23 versus 2, $\chi^2=14.45$, $p=0.00014$), which required excluding these controls because the anonymization procedure prevented further investigation. Conversely, more case samples were excluded because they were outliers to the main EA and AA distributions in analyses of AIMs (63 versus 35, $\chi^2=10.18$, $p=0.0014$), perhaps because controls gave more accurate self-reports and were asked for more detailed information (i.e., ancestry of each grandparent rather than each parent).

Next, for a schizophrenia GWAS [2], we genotyped 3,827 (2,817 EA and 1,010 AA) of the 4,665 collected control samples and 4,196 (2,838 EA and 1,358 AA) case samples with the

Affymetrix 6.0 array; 3,626 (95%) control and 3,967 (95%) case samples passed stringent QC (ST2). Thus, the control and case collections were similar in the quality of DNA and the consistency of genotypes with available phenotypic information.

Sharing of Biomaterials and Data

These materials are available to qualified scientists applying to NIMH, and overlapping subsets of the MGS2 control sample have been genotyped by multiple platforms: Affymetrix 6.0 [2], Affymetrix 500K [23], and Illumina HumanHap550 [24]. Genotypes and phenotypic data for controls used in the MGS2 GWAS are available by application to dbGaP (database of Genotypes and Phenotypes, dbgap.ncbi.nlm.nih.gov, Study Accessions: phs000021.v2.p1 and phs000167.v1.p1), with 177 research projects already (12/07/09) granted access. Multiple investigators have also accessed DNA and/or phenotypic data through the NIMH repository, beginning during the recruitment period when these materials were made available in batches twice yearly.

RESULTS

Questionnaire Completion

The questionnaire includes 69 mandatory or screening items, 45 disease-related and 24 for personality traits. We defined several sets of response-rate outliers for these items: 0.5% answered “yes” to 50 items, 0.5% did not answer 5 items, and 0.4% had software failures (1 item unasked). Users of this control sample may choose to remove these rare questionnaire completion outliers, as we did for MGS2; the current analyses utilize the full control sample.

Demographic Comparability

Table 1 compares demographic data for the control samples (EA, AA-KN, and AA-SSI) with the November 2003, U.S. Current Population Survey (CPS) [25] for non-Hispanic adults reporting a single race (white or black). Differences versus CPS include: more females in the AA-KN subsample; older mean age for EA subjects (50.0 versus 47.5 years old for CPS, t -test $p < 0.0001$); fewer employed EA subjects, perhaps reflecting older age (mean age for EA employed subjects was 43.8 versus 58.1 for non-workers, t -test $p < 0.0001$); EA and particularly the AA control subjects were more educated; more EA subjects in the highest income bracket, and more AA subjects (especially AA-SSI) in several upper brackets; more EA subjects were from the South and fewer from the Northeast and West, while more AA subjects were from the Midwest.

Ancestry

Numerical data for 2 grandparents were reported by 87% of EA and 96% of AA subjects, with “not sure” for most of the remaining grandparents. Counting each entry as an equal proportion for each grandparent, ancestry was distributed similarly in subjects with data for four grandparents (69% EA and 80% AA subjects) versus those with any missing data (Supplementary Figures, SF1 and SF2). Data for all subjects (ignoring missing data) are shown in Figure 1. For EA subjects, 83% of ancestry was reported as Anglo-Saxon (i.e., British Isles), Northern European (e.g., Norwegian), or West European (e.g., French,

German), with the remaining entries distributed across other European as well as Native American ancestries (Figure 1). Our EA control sample and the U.S. Census show similar distributions of subcontinental European ancestries (SF3).

For AA subjects, 75% ancestry was reported as AA and 15% as Native American, and most of the rest as EA (Figure 1). Genotypic data suggest that Native American ancestry was over-reported for both AA and EA subjects (refer to SF3 in [2] which shows that few subjects had the elevated “Asian” ancestry scores that are observed for Native Americans). Reasons for over-reporting of Native American ancestry have been discussed in [26]. Among AA subjects, 42% reported that all four grandparents were only AA. The proportion of EA admixture in the AA sample is captured more accurately by genotypic data, and our sample (see SF3 in [2]) looks similar to other AA samples [27–29].

Most EA subjects (73%) reported multiple ancestries and/or had some missing data, but some reported four grandparents of the same ancestry (e.g., 15% with all Anglo-Saxon, 7% with all West European ancestry). Ancestry-informative principle component scores computed from our GWAS data demonstrated the accuracy of these self-reported data, given that the scores produced distinct clusters of individuals with single ancestries (see SF6 in [2]), with two components predicting North-South and East-West geographical origins respectively, and another component predicting Ashkenazi Jewish ancestry. These scores were used in our GWAS analyses to exclude outliers and to correct association tests for case-control ancestry differences [2].

Diagnosis

Table 3 shows the lifetime prevalence of common DSM-IV diagnoses [21] by CIDI-SF scoring criteria for the population-based portion of the control sample (EA; AA-KN); Table 4 provides additional data. The AA-SSI subsample (internet banner ad recruitment) has higher lifetime prevalence for common diagnoses, more comorbidity, higher neuroticism scores, and lower extraversion scores, compared with the KN subsample that was recruited from a survey panel, and is thus presented separately (ST3 and ST4). ST5 includes comparison lifetime prevalence data from several epidemiological surveys (for all ancestries combined), for the Epidemiological Catchment Area (ECA) study [30], the National Comorbidity Survey (NCS) [31], the NCS Replication (NCS-R) [32, 33], and the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC) – Wave I [34, 35]. Generally, male:female prevalence ratios were similar to those in previous studies for diagnoses compared with previous work (Tables 3 and 4, ST5) and EA:AA-KN ratios were comparable within our study, but lifetime prevalences are higher in our sample (Table 3) than in the NCS [31] which used the CIDI interview. For example, at least one anxiety disorder diagnosis (other than OCD) was assigned in 40.6% of EA and 42.9% of AA-KN subjects versus 24.9% in NCS [31].

An alternative CIDI-SF scoring system assigns a probability of caseness to each subject [36], which can be summed across subjects to estimate prevalence. This estimate attempts to account for unmeasured exclusion criteria (such as organic causes) and over-reporting [20]. By this method (ST5), lifetime prevalence estimates are reduced for major depression and specific phobia, but not for alcohol or drug dependence (perhaps because organic exclusions

are not pertinent for these diagnoses). We restricted further diagnostic analyses to the dichotomous method.

Alcohol and Drug Dependence

Prevalences were higher for EA control subjects versus NCS for alcohol dependence (23.1% versus 14.1%) and drug dependence (12.2% versus 7.5%), with the expected excess in males (Table 3). There was higher prevalence of alcohol dependence in EA versus AA control subjects (23.1% versus 17.8%; $\chi^2=7.17$, $p=0.0074$), but higher prevalence of drug dependence in AA-KN versus EA subjects (16.3% versus 12.2%; $\chi^2=6.23$, $p=0.013$). In the full (EA+AA) KN sample, alcohol dependence showed the expected increase in males versus females (29.1% versus 16.8%, $\chi^2=83$, $p<0.0001$). Also, KN subjects diagnosed with drug dependence were younger (43.7 versus 50.3 mean age, t-test $p<0.0001$) and had lower household income (<\$50,000 for 66.9% versus 55.8%, $\chi^2=21.4$, $p<0.0001$), but were more likely to be employed (62.0% versus 56.1%, $\chi^2=6$, $p=0.014$).

Major Depression

Our EA control subjects have a much higher lifetime prevalence of major depressive episodes (MDEs) compared with NCS (40.2% versus 17.1%; reduced to 34.6% for control subjects using probability of caseness as in ST5); among EA subjects with MDEs, mean age of onset was 26.6, and 78% reported recurrent MDEs (Table 3). It is known that CIDI-SF over-diagnoses major depression [37]. Without interview data we cannot determine whether there might be a true over-representation of individuals with depression in the KN panel, although this appears unlikely given that the controls scored lower than population means for neuroticism scale (see below) which is often elevated in depression. CIDI-SF MDE criteria ignore distress/impairment [20], although queried; if we restrict MDE diagnosis to those who sought help, received treatment, or reported a lot or some impairment, lifetime prevalence was reduced by 3.1% for EA and by 7.3% for AA-KN control subjects. CIDI-SF also fails to differentiate MDE from the effects of substance use. Lifetime prevalence of MDE was 25.4% in controls without alcohol or drug dependence and 29.4% in controls who did not endorse a CIDI-SF item about emotional or psychological problems (anhedonia, depression, and paranoia) due to substance use (alcohol and/or drugs). A similar pattern was observed in AA-KN subjects. In the full KN control sample, subjects with MDE were more likely to be female (male 29.3% versus female 49.7%, $\chi^2=165$, $p<0.0001$), younger (46.8 versus 51.3 mean age, t-test $p<0.0001$), unmarried (51.2% versus 65.2%, $\chi^2=75$, $p<0.0001$), and to have lower household income (<\$50,000, 63.3% versus 53.0%, $\chi^2=41$, $p<0.0001$).

Anxiety Disorders

Self-reported prevalence of generalized anxiety disorder (GAD) was higher in EA control subjects than in the NCS (20.2% versus 5.1%), with prevalence of 11.7% when excluding those with alcohol and/or drug dependence, or 2.9% when excluding subjects with any MDE (though CIDI-SF does not ask about relative timing of GAD symptoms versus MDEs). We found a similar pattern and estimates for AA-KN subjects. All anxiety disorders were more prevalent in women. Prevalence of CIDI-SF anxiety disorders was similar in EA versus AA-KN subjects except for specific phobia (18.5% EA versus 23.8% AA, $\chi^2=7.82$, $p=0.0052$),

agoraphobia without panic attacks (2.9% EA versus 5.3% AA, $\chi^2=7.27$, $p=0.0070$), and OCD (8.2% EA versus 12.3% AA, $\chi^2=9.03$, $p=0.0027$).

Nicotine Dependence

Nicotine dependence, defined as FTND 4 (Fagerström Test for Nicotine Dependence, [38]), was more common in men in our full KN control sample (36.7% versus 30.9%, $\chi^2=14.3$, $p=0.0002$), in EA versus AA-KN (34.2% versus 29.7%, $\chi^2=3.91$, $p=0.048$), and in those with less education (high school completion or less for 45.7% for FTND 4 subjects versus 27.8% for the rest, $\chi^2=124$, $p<0.0001$).

Neuroticism and Extraversion

In EA control subjects, mean scores for neuroticism were 2.9 in males and 4.4 in females, compared with reported population means of 4.95 and 5.90 respectively [39]; and for extraversion they were 6.8 for males and 6.9 for females (versus population means of 6.36 and 7.60). Mean scores were similar in AA-KN subjects except for higher extraversion scores (7.7 versus 6.9 in EA). Control subjects meeting CIDI-SF criteria for MDE had higher mean neuroticism scores (5.8 versus 2.3 in EA, and 5.8 versus 2.4 for AA-KN) and lower mean extraversion scores (6.1 versus 7.4 in EA, and 6.5 versus 8.5 in AA-KN), compared with subjects with no self-reported MDE. Similar patterns of elevated neuroticism and lowered extraversion scores were seen with various anxiety diagnoses (data not shown).

Comorbidity

As shown in Table 4, there was extensive comorbidity, with more than one CIDI-SF diagnosis for around 60% of individuals with at least one. Subjects with three or more diagnoses (20.2% of EA and 22.9% of AA-KN subjects) accounted for most of the total CIDI-SF diagnoses in their groups (62.9% and 66.1%, respectively). The most highly screened control group would consist of the 42.3% of EA and 37.2% AA-KN subjects who received no CIDI-SF diagnoses, and who denied all psychosis and bipolar screens.

Other Conditions and Traits

EA and AA subjects were similar in height, with males about 15 cm taller than females. Based on current body mass index (BMI, kg/m²) and on highest non-pregnancy BMI, the majority of subjects were either overweight (25 BMI<30) or obese (BMI 30), particularly AA women, similar to contemporaneous estimates from the National Health and Nutrition Examination Study (NHANES) [40]. Similar to previous work (e.g., [41]), in terms of sexual orientation (identity) for the EA sample, we find more homosexual men (2.9%) than homosexual women (1.5%), a higher ratio of bisexuals to homosexuals in women (2.13) than in men (0.86), and similar percentages for the AA-KN sample though the M:F ratios are slightly larger.

Psychosis and Mania

The proportion of EA control subjects endorsing previous treatment or diagnosis of schizophrenia or schizoaffective disorder was identical to that in the NCS sample (0.7%) [31], but more of our control sample reported treatment or diagnosis of bipolar disorder

(3.7% versus 1.6%). AA-KN subjects reported more schizophrenia/schizoaffective (2.3% versus 0.7% EA; $\chi^2=12.9$, $p=0.0003$) and bipolar disorder histories (5.9% versus 3.7% EA; $\chi^2=4.75$, $p=0.029$). Without more direct data (i.e., interview data of probands, family informants, and/or review of medical records), we cannot determine the true rates of these disorders in these samples, but for MGS analyses we chose to exclude all individuals who endorsed or failed to answer any of the relevant items (i.e., the three psychosis/bipolar screening questions) [1, 2].

DISCUSSION

The challenge was to find a cost-effective method to collect a control sample as large as our case sample. The KN panel provided an opportunity to recruit a representative national sample and a modest cost, and the online lifetime CIDI-SF provided an efficient method for screening of common psychiatric disorders. Our methods allowed us to recruit individuals throughout the U. S. and from all demographic strata, while obtaining written informed consent for use of biomaterials, clinical information, and genetic information for any medical research, and for anonymization of the sample to ensure access by future investigators.

The results demonstrate that the MGS2 control sample has the following characteristics:

1. It is representative of the U.S. population demographically except for a bias toward higher age, unemployment, education and income (but all strata are represented), and with an excess of EA subjects from the South and AA subjects from the Midwest. Older control subjects are past the typical ages of onset for the assayed psychiatric disorders, which might be an advantage for studies of those common disorders. Selection of control subjects based on educational level might be considered for studies of disorders manifesting with cognitive difficulties.
2. It is similar to the population in anthropometrics, ancestral background, and sexual orientation (identity). Self-reported ancestral data were consistent with genotypic analyses for both EA and AA subjects, except that Native American ancestry appears to have been over-reported [2]. Molecular QC analyses demonstrated that DNA quality and consistency of self-reported and genotypic ancestry were similar for MGS2 control and case samples, despite the fact that control subjects were recruited and evaluated online with no direct contact besides the company staff (KN, SSI, EMSI), and the case subjects had direct contact (recruitment, assessment, venipuncture) with the research group.
3. The diagnostic prevalences of common psychiatric disorders were substantially higher in these CIDI-SF self-reported control subjects compared with the CIDI interviewed NCS sample. The excess prevalence was less pronounced when measured by probabilities of caseness or in those without alcohol or substance dependence. On the other hand, mean neuroticism scores (which are typically associated with depressive and anxiety disorders) are lower than previously

reported population means, for both EA and AA subjects. Male:female ratios and patterns of comorbidity are similar to those reported previously (Table 3, ST5).

4. The AA-SSI subsample that was recruited by internet ads had more CIDI-SF diagnoses and comorbidity, higher neuroticism scores, and lower extraversion scores, compared to AA-KN panel subjects, consistent with previous work [42]. We were unable to recruit the entire AA control sample through the more representative KN panel because our required AA:EA ratio (~1:2) to match the case collection was much higher than that in the general population, and because the volunteering rate was lower (see review for possible reasons [43]) among AA versus EA panel members (15.5% versus 21.7%, ST1).

Several factors could have contributed to the high rate of common psychiatric diagnoses (also see Supplementary Discussion for disorder-specific information). CIDI-SF generally over-diagnoses these disorders (e.g., [37, 44]), perhaps due to a combination of failure to elicit key exclusion criteria (medical disorders, effects of medications and of substance use, bereavement, temporal overlap of syndromes, etc.) or inclusion criteria (impairment, distress, etc.), and subjects' tendency to endorse items based on experiences that a clinician interviewer would probe and then recognize as non-pathological. Volunteers for this type of study might also be more likely to have common psychiatric disorders, although the low neuroticism scores observed here suggest otherwise. Finally, the effects of using a "private" evaluation method (in this case an online interview) as opposed to a face-to-face interview might influence the probability of reporting psychopathology. The directionality of this effect can vary according to subject matter and context (see discussion in [45]), but more anonymous methods may lead to more accurate responses to sensitive topics [46–50].

Users of this control sample have many options. For studies of common disorders, one might exclude individuals with the disorder under study (or a spectrum of disorders that co-aggregate in families) based on dichotomous CIDI-SF criteria (at the standard thresholds, or at lower or higher thresholds) or caseness probabilities. We have constructed factor scores (depression, anxiety, "internalizing disorders," nicotine dependence, substance dependence) that might more accurately measure the probability of a clinically significant disorder in each individual (manuscript in preparation), and others have shown that restricting GWAS analyses of such secondary traits to the controls in a case-control sample provides appropriate Type I error rates as long as either the marker or the secondary trait is not associated with disease risk (for MGS2, schizophrenia) in the study base [51]. It is also possible to select a "hypernormal" control group (although with a ~60% reduction in sample size), by selecting only individuals with no CIDI-SF diagnosis who also denied any psychotic or bipolar disorder screens. For example, for a sample of 534 MGS2 control subjects selected in this way (from a pool of 1,844 "hypernormal" MGS2 control subjects), neuroticism and extraversion scores did not differ from those of 90 age-matched individuals with no lifetime psychiatric disorder or family history of anxiety disorder by structured interview [52]. To minimize selection bias, the control group should be chosen with the same restrictions as the cases in association studies [9, 10], which is one reason why we did not restrict the studied control sample to "hypernormal" control subjects in our MGS2 GWAS [2].

Clinical screening information on control samples is useful (e.g., [4, 8]). Efficient collection of a large, anonymized control sample, generally representative of the U.S. adult EA and AA population for many aspects of demography, ancestry, and morbidity, was enabled by the internet-based recruitment and assessment design utilized for the MGS2 control sample. Information presented in this manuscript will further enable users of this control sample to follow standard guidelines for selection of controls in their own case-control studies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the individuals who volunteered for the control sample for their participation. Data and biomaterials from the NIMH GI (Genetics Initiative) MGS2 (Molecular Genetics of Schizophrenia) control sample were collected by NorthShore University HealthSystem, Evanston, IL, R01 MH59571, Pablo V. Gejman, M.D. (Collaboration Coordinator; PI) as part of a collaborative R01 application comprised of ten sites. From 2003–2006, the Principal Investigators and Co-Investigators were: NorthShore University HealthSystem, Evanston, IL, R01 MH59571, Pablo V. Gejman, M.D. (Collaboration Coordinator; PI), Alan R. Sanders, M.D.; Emory University School of Medicine, Atlanta, GA, R01 MH59587, Farooq Amin, M.D. (PI); University of California, San Francisco, CA, R01 MH60870, William F. Byerley, M.D. (PI); University of Iowa, Iowa, IA, R01 MH59566, Donald W. Black, M.D. (PI), Raymond R. Crowe, M.D.; Washington University, St. Louis, MO, R01 MH60879, C. Robert Cloninger, M.D. (PI); University of Colorado, Denver, CO, R01 MH59565, Robert Freedman, M.D. (PI), Ann Olincy, M.D.; Stanford University, Palo Alto, CA, R01 MH61675, Douglas F. Levinson, M.D. (PI); Louisiana State University, New Orleans, LA, R01 MH67257, Nancy G. Buccola, APRN, BC, MSN (PI); University of Queensland, Brisbane, Queensland, Australia, R01 MH59588, Bryan J. Mowry, M.D. (PI); Mt. Sinai School of Medicine, New York, NY, R01 MH59586, Jeremy M. Silverman, Ph.D. (PI). Our schizophrenia GWAS (genotyping and analyses) was supported by the Genetic Association Information Network (GAIN) and grants U01MH79469 (Pablo V. Gejman, M.D.), U01MH79470 (Douglas F. Levinson, M.D.), and R01MH81800 U01MH79469 (Pablo V. Gejman, M.D.). We also thank the individuals at each participating institution for their various contributions, especially Roberta Fishman, John Corcoran, and Douglas A. Fugman.

REFERENCES (MAIN TEXT)

1. Sanders AR, et al. No significant association of 14 candidate genes with schizophrenia in a large European ancestry sample: implications for psychiatric genetics. *Am J Psychiatry*. 2008; 165(4): 497–506. [PubMed: 18198266]
2. Shi J, et al. Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature*. 2009; 460(7256):753–757. [PubMed: 19571809]
3. WTCCC . Wellcome Trust Case Control Consortium: Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007; 447(7145):661–678. [PubMed: 17554300]
4. McCarthy MI, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet*. 2008; 9(5):356–369. [PubMed: 18398418]
5. Schechter D, Lebovitch R. Normal controls are expensive to find: methods to improve cost-effectiveness of the screening evaluation. *Psychiatry Res*. 2005; 136(1):69–78. [PubMed: 16023733]
6. Adami H, et al. Use of telephone screens improves efficiency of healthy subject recruitment. *Psychiatry Res*. 2002; 113(3):295–301. [PubMed: 12559485]
7. Heeren T, et al. A comparison of results from an alcohol survey of a prerecruited Internet panel and the National Epidemiologic Survey on Alcohol and Related Conditions. *Alcohol Clin Exp Res*. 2008; 32(2):222–229. [PubMed: 18162076]
8. Morton NE, Collins A. Tests and estimates of allelic association in complex inheritance. *Proc Natl Acad Sci U S A*. 1998; 95(19):11389–11393. [PubMed: 9736746]
9. Rothman, KJ, Greenlan, S, Lash, TL. Case-Control Studies. In: Rothman, KJ, Greenlan, S, Lash, TL, editors *Modern Epidemiology*. Lippincott Williams & Wilkins; Philadelphia: 2008. 111–127.

10. Schwartz, S, Susser, E, Gorman, JM. Choosing Controls in Biologic Psychiatry. In: Susser, E, , et al., editors *Psychiatric Epidemiology: Searching for the Causes of Mental Disorders*. Oxford University Press; New York: 2006. 247–259.
11. Couper M. Web surveys: a review of issues and approaches. *Public Opin Q*. 2000; 64(4):464–494. [PubMed: 11171027]
12. Krotki, K; Dennis, JM. Probability-based survey research on the internet. Conference of the International Statistical Institute; 2001; Seoul, South Korea.
13. Firmann M, et al. The CoLaus study: a population-based study to investigate the epidemiology and genetic determinants of cardiovascular risk factors and metabolic syndrome. *BMC Cardiovasc Disord*. 2008; 8:6. [PubMed: 18366642]
14. Homer N, et al. Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. *PLoS Genet*. 2008; 4(8):e1000167. [PubMed: 18769715]
15. Wittchen HU. Reliability and validity studies of the WHO--Composite International Diagnostic Interview (CIDI): a critical review. *J Psychiatr Res*. 1994; 28(1):57–84. [PubMed: 8064641]
16. Kessler RC, et al. Methodological studies of the Composite International Diagnostic Interview (CIDI) in the US National Comorbidity Survey. *Int J Methods Psychiatr Res*. 1998; 7(1):33–55.
17. Gigantesco A, Morosini P. Development, reliability and factor analysis of a self-administered questionnaire which originates from the World Health Organization's Composite International Diagnostic Interview - Short Form (CIDI-SF) for assessing mental disorders. *Clin Pract Epidemiol Ment Health*. 2008; 4:8.
18. Kessler RC, et al. The World Health Organization Composite International Diagnostic Interview short-form (CIDI-SF). *Int J Methods Psychiatr Res*. 1998; 7(4):171–185.
19. Patten SB, et al. Performance of the composite international diagnostic interview short form for major depression in a community sample. *Chronic Dis Can*. 2000; 21(2):68–72. [PubMed: 11007657]
20. Nelson CB, Kessler RC, Mroczek D. 2001 Scoring the World Health Organization's Composite International Diagnostic Interview Short Form (CIDI-SF; v1.0 NOV98 for all disorders except OCD which is from v1.1 MAR99).
21. APA. *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV)*. Washington, DC: American Psychiatric Association; 1994.
22. Nurnberger JI Jr, et al. Diagnostic interview for genetic studies. Rationale, unique features, and training. NIMH Genetics Initiative. *Arch Gen Psychiatry*. 1994; 51(11):849–859. [PubMed: 7944874]
23. Sklar P, et al. Whole-genome association study of bipolar disorder. *Mol Psychiatry*. 2008; 13(6): 558–569. [PubMed: 18317468]
24. Scott LJ, et al. Genome-wide association and meta-analysis of bipolar disorder in individuals of European ancestry. *Proc Natl Acad Sci U S A*. 2009; 106(18):7501–7506. [PubMed: 19416921]
25. Current Population Survey. U. S. Census Bureau; Nov, 2003
26. Klimentidis YC, Miller GF, Shriver MD. Genetic admixture, self-reported ethnicity, self-estimated admixture, and skin pigmentation among Hispanics and Native Americans. *Am J Phys Anthropol*. 2009; 138(4):375–383. [PubMed: 18951390]
27. Parra EJ, et al. Estimating African American admixture proportions by use of population-specific alleles. *Am J Hum Genet*. 1998; 63(6):1839–1851. [PubMed: 9837836]
28. Smith MW, et al. A high-density admixture map for disease gene discovery in African Americans. *Am J Hum Genet*. 2004; 74(5):1001–1013. [PubMed: 15088270]
29. Cheng CY, et al. Admixture mapping of 15,280 African Americans identifies obesity susceptibility Loci on chromosomes 5 and X. *PLoS Genet*. 2009; 5(5):e1000490. [PubMed: 19461885]
30. Robins, LN, Regier, DA. *Psychiatric Disorders in America: The Epidemiological Catchment Area Study*. New York: The Free Press; 1991.
31. Kessler RC, et al. Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. Results from the National Comorbidity Survey. *Arch Gen Psychiatry*. 1994; 51(1): 8–19. [PubMed: 8279933]

32. Kessler RC, et al. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry*. 2005; 62(6):593–602. [PubMed: 15939837]
33. Kessler RC, et al. The US National Comorbidity Survey Replication (NCS-R): design and field procedures. *International Journal of Methods in Psychiatric Research*. 2004; 13(2):69–92. [PubMed: 15297905]
34. Grant BF, et al. Prevalence and co-occurrence of substance use disorders and independent mood and anxiety disorders: results from the National Epidemiologic Survey on Alcohol and Related Conditions. *Arch Gen Psychiatry*. 2004; 61(8):807–816. [PubMed: 15289279]
35. Conway KP, et al. Lifetime comorbidity of DSM-IV mood and anxiety disorders and specific drug use disorders: results from the National Epidemiologic Survey on Alcohol and Related Conditions. *J Clin Psychiatry*. 2006; 67(2):247–257. [PubMed: 16566620]
36. WHO. World Health Organization. Composite International Diagnostic Interview, Version 1.0. Geneva: The World Health Organization; 1990.
37. Aalto-Setälä T, et al. Major depressive episode among young adults: CIDI-SF versus SCAN consensus diagnoses. *Psychol Med*. 2002; 32(7):1309–1314. [PubMed: 12420900]
38. Heatherton TF, et al. The Fagerstrom Test for Nicotine Dependence: a revision of the Fagerstrom Tolerance Questionnaire. *Br J Addict*. 1991; 86(9):1119–1127. [PubMed: 1932883]
39. Eysenck SBG, Eysenck HJ, Barrett P. A revised version of the psychoticism scale. *Pers Individ Dif*. 1985; 6:21–29.
40. Wang Y, et al. Will all Americans become overweight or obese? Estimating the progression and cost of the US obesity epidemic. *Obesity (Silver Spring)*. 2008; 16(10):2323–2330. [PubMed: 18719634]
41. Laumann EO, et al. The social organization of sexuality: Sexual practices in the United States. The University of Chicago Press; Chicago: 1994. Homosexuality; 283–321.
42. Halbreich U, et al. The normalcy of self-proclaimed “normal volunteers”. *Am J Psychiatry*. 1989; 146(8):1052–1055. [PubMed: 2750979]
43. Bussey-Jones J, et al. Asking the right questions: views on genetic variation research among black and white research participants. *J Gen Intern Med*. 2009; 24(3):299–304. [PubMed: 19101773]
44. Orlando M, et al. Re-estimating the prevalence of psychiatric disorders in a nationally representative sample of persons receiving care for HIV: results from the HIV Cost and Services Utilization Study. *Int J Methods Psychiatr Res*. 2002; 11(2):75–82. [PubMed: 12459797]
45. Johnson TP, Hougland JG, Moore RW. Sex differences in reporting sensitive behavior: A comparison of interview methods. *Sex Roles*. 1991; 24(11/12):669–679.
46. ACSF. Analysis of sexual behaviour in France (ACSF). A comparison between two modes of investigation: telephone survey and face-to-face survey. ASCF principal investigators and their associates. *AIDS*. 1992; 6(3):315–323. [PubMed: 1567577]
47. Paulsen AS, et al. Reliability of the telephone interview in diagnosing anxiety disorders. *Arch Gen Psychiatry*. 1988; 45(1):62–63. [PubMed: 3337610]
48. Slutske WS, et al. Long-term reliability and validity of alcoholism diagnoses and symptoms in a large national telephone interview survey. *Alcohol Clin Exp Res*. 1998; 22(3):553–558. [PubMed: 9622431]
49. Fenig S, et al. Telephone vs face-to-face interviewing in a community psychiatric survey. *Am J Public Health*. 1993; 83(6):896–898. [PubMed: 8498632]
50. Newman JC, et al. The differential effects of face-to-face and computer interview modes. *Am J Public Health*. 2002; 92(2):294–297. [PubMed: 11818309]
51. Monsees GM, Tamimi RM, Kraft P. Genome-wide association scans for secondary traits using case-control samples. *Genet Epidemiol*. 2009; 33(8):717–728. [PubMed: 19365863]
52. Talati A, Fyer AJ, Weissman MM. A comparison between screened NIMH and clinically interviewed control samples on neuroticism and extraversion. *Mol Psychiatry*. 2008; 13(2):122–130. [PubMed: 17938631]

Figure 1. Results of the tabulation of the frequency with which self-identified EA controls ($n = 3,364$, grey bars) or AA controls ($n = 1,301$, black bars) endorsed biological grandparents with the ancestral background indicated on the x-axis (one particular grandparent may have more than one ancestry endorsed).

Table 1

Demographics of EA & AA Control Sample versus U.S. Adults

Variable	Category	AA Controls by					
		All Controls			U.S. Population		
		EA	AA	KN	SSI	EA	AA
Sample Size:		3,364	1,301	529	772		
Gender	Men	46.9	39.5	33.6[*]	43.5	47.9	42.1
	Women	53.1	60.5	66.4 [*]	56.5	52.1	57.9
Age	18–24	5.0[*]	5.7[*]	5.7[*]	5.7[*]	10.3	13.2
	25–34	14.9	17.5	17.8	17.4	15.6	18.2
	35–44	20.4	22.4	20.0	24.0	20.0	21.1
	45–54	19.8	30.7 [*]	30.1 [*]	31.1 [*]	20.7	20.3
	55–64	18.1 [*]	16.8 [*]	17.0 [*]	16.7 [*]	14.8	13.0
	65 or over	21.8 [*]	6.9[*]	9.5[*]	5.2[*]	18.7	14.2
Employment Status	In the labor force	57.1[*]	59.3	55.6	61.8	65.0	58.3
	Not in the labor force	42.9 [*]	40.7	44.4	38.2	35.0	41.7
Marital Status	Married	62.7	n.a.	39.3	n.a.	61.6	39.0
	Not married	37.3	n.a.	60.7	n.a.	38.4	61.0
Level of Education	< HS diploma	7.9[*]	4.0[*]	6.8[*]	2.1[*]	10.4	20.0
	HS diploma or equivalent	27.2[*]	15.8[*]	18.5[*]	13.9[*]	33.2	36.1
	Some college	22.9 [*]	39.7 [*]	37.2 [*]	41.5 [*]	19.6	20.3
	Associate degree	7.6[*]	11.8 [*]	11.7 [*]	11.8 [*]	8.7	7.5
	Bachelor's degree	34.3 [*]	28.7 [*]	25.7 [*]	30.7 [*]	28.1	16.1
Household Income	Under \$10,000	5.3[*]	10.4[*]	13.6[*]	8.2[*]	6.6	17.8
	\$10,000–\$24,999	17.1	23.0	25.9	21.0[*]	18.2	25.8

Variable	AA Controls by					
	All Controls			Source		U.S. Population
	EA	AA	KN	SSI	EA	AA
Sample Size:						
\$25,000–\$49,999	33.1	32.4	28.7	34.8	34.0	31.9
\$50,000–\$74,999	23.8 *	17.1	17.2	17.1	25.6	16.3
\$75,000 or more	20.7*	16.1*	14.6*	17.2	15.6	8.1
Census Region						
Northeast	18.3 *	16.1	15.7	16.3	23.9	17.5
Midwest	27.1	25.4*	25.0*	25.6*	27.9	19.0
South	36.8*	52.3	51.6	52.7	26.6	54.6
West	17.9 *	6.3 *	7.8	5.3 *	21.6	8.9

Values represent percentages of the control sample or subgroup, or of the general population in each category for each variable. Self-identified EA and AA control subjects with blood data from complete collection (marital status not provided from SSI). HS = high school. Knowledge Networks (KN) Adult Panel makeup data from January 2004 and adult U.S. Population data from CPS (Current Population Survey) in November 2003 [25] had previously been shown to match well [1]. November 2003 CPS data here is restricted to non-Hispanic adults listing white-only or black-only for race in the CPS.

* Asterisks indicate differences in the control sample data ($p < 0.05$ based on χ^2 comparison of counts in the respective categories) compared to the U.S. Population data for EA or AA, as appropriate, with **bolding** indicating a value lower than the U.S. Population data.

Table 2**Summary of Online Self-Report Assessment**

Phenotype	Measure
Common psychiatric disorders	CIDI-SF [18], assessing major depression, generalized anxiety, specific phobia, panic attack, agoraphobia, social phobia, obsessive-compulsive disorder, alcohol dependence and drug dependence
Psychotic and bipolar disorders	Three items inquiring about lifetime diagnosis of, or treatment for, schizophrenia or schizoaffective disorder, psychotic symptoms, or bipolar disorder
Nicotine dependence	Screening items for nicotine dependence, including all items for generating a Fagerström Test for Nicotine Dependence (FTND, [38]) score
Personality	Brief (12-item) Eysenck Neuroticism and Extraversion scales [39]
Anthropometrics	Current height and weight; highest lifetime weight outside of pregnancy
Sexual identity	Single item
Ancestry	Selection of all applicable ancestries (race/ethnicity) for each grandparent.
Demographics	Previously collected by KN and SSI, confer Table 1

The modified questionnaire (CIDI-SF was originally developed to assess disorders for the past 12 months [18]) allowed estimation of the lifetime percentages (uncorrected for age of risk of onset) of the above psychiatric diagnoses via DSM-IV [21], with some limitations (chiefly due to questionnaire time/cost and complexity constraints).

Table 3
Common Psychiatric Diagnoses (DSM-IV from CIDI-SF) in EA & AA (KN) Control Sample

Condition	EA Controls			AA Controls (from KN)		
	Males	Females	Total	Males	Females	Total
Sample Size:	1,578	1,786	3,364	178	351	529
Substance Use Disorders						
<i>Alcohol dependence</i>	0.295	0.175	0.231	0.258	0.137	0.178
<i>Drug dependence</i>	0.139	0.108	0.122	0.213	0.137	0.163
<i>Any substance dependence above</i>	0.331	0.211	0.268	0.348	0.219	0.263
<i>Substance-induced psychological sx (SIPS)</i>	0.179	0.153	0.165	0.230	0.171	0.191
Major Depressive Episodes (MDE)						
<i>All</i>	0.292	0.500	0.402	0.303	0.481	0.422
<i>Mean age at onset</i>	28.9	25.5	26.6	26.5	24.8	25.4
<i>Single episode</i>	0.077	0.101	0.090	0.051	0.083	0.072
<i>Recurrent</i>	0.215	0.399	0.313	0.253	0.399	0.350
<i>Without SIPS</i>	0.194	0.382	0.294	0.185	0.362	0.302
<i>Without substance dependence</i>	0.146	0.350	0.254	0.163	0.316	0.265
Generalized Anxiety Disorder (GAD)						
<i>All</i>	0.143	0.255	0.202	0.135	0.219	0.191
<i>Without MDE</i>	0.027	0.030	0.029	0.017	0.023	0.021
<i>Without substance dependence</i>	0.064	0.164	0.117	0.045	0.131	0.102
Anxiety Disorders						
<i>Specific Phobia</i>	0.132	0.233	0.185	0.213	0.251	0.238
<i>Social Phobia</i>	0.114	0.169	0.143	0.135	0.154	0.147
<i>Agoraphobia without Panic Attacks</i>	0.020	0.038	0.029	0.039	0.060	0.053
<i>Panic Attacks</i>	0.092	0.196	0.147	0.129	0.191	0.170
<i>Panic Attacks without Agoraphobia</i>	0.068	0.146	0.109	0.101	0.142	0.129
<i>Panic Attacks with Agoraphobia</i>	0.023	0.050	0.038	0.028	0.048	0.042
<i>Generalized Anxiety Disorder</i>	0.143	0.255	0.202	0.135	0.219	0.191

Condition	EA Controls			AA Controls (from KN)		
	Males	Females	Total	Males	Females	Total
Sample Size:	1,578	1,786	3,364	178	351	529
Any anxiety disorder above (GAD, panic, phobia)	0.316	0.485	0.406	0.326	0.481	0.429
Any of the above, without SIPS	0.205	0.368	0.292	0.191	0.353	0.299
Any of the above w/o subst dep	0.156	0.334	0.251	0.163	0.330	0.274
Obsessive Compulsive Disorder (OCD)	0.061	0.100	0.082	0.118	0.125	0.123

EA and AA (KN) control subjects' condition frequencies. SIPS = Substance-induced psychological symptoms; sx = symptoms; Disorders are treated dichotomously (present/absent) in this table for the studied sample.

Table 4

Additional Phenotypic Information in EA and AA (KN) Control Sample

Condition	EA Controls			AA Controls (from KN)		
	Males	Females	Total	Males	Females	Total
Sample Size:	1,578	1,786	3,364	178	351	529
Other disorders, traits, and conditions						
<i>Nicotine dependence (FTND 4)</i>	0.372	0.315	0.342	0.326	0.282	0.297
<i>Mean FTND score, all</i>	2.6	2.3	2.4	2.4	2.0	2.1
<i>Mean Eysenck brief Neuroticism score (0–12)</i>	2.9	4.4	3.7	3.0	4.2	3.8
<i>Mean Eysenck brief Extraversion score (0–12)</i>	6.8	6.9	6.9	7.7	7.6	7.7
<i>Sexual identity: bisexual</i>	0.025	0.032	0.029	0.039	0.023	0.028
<i>Sexual identity: homosexual</i>	0.029	0.015	0.021	0.034	0.006	0.015
<i>Overweight or obese (current BMI 25)</i>	0.688	0.611	0.647	0.685	0.738	0.720
<i>Obese (current BMI 30)</i>	0.314	0.368	0.342	0.404	0.501	0.469
<i>Mean height (m)</i>	1.79	1.64	1.71	1.79	1.65	1.70
<i>Mean highest lifetime BMI (kg/m²)</i>	31.0	32.4	31.8	31.8	35.9	34.5
Psychosis & mania screens - endorsed						
<i>Dx and/or Tx of SZ and/or SA</i>	0.008	0.007	0.007	0.028	0.020	0.023
<i>Dx and/or Tx of AH and/or delusions</i>	0.013	0.016	0.015	0.045	0.046	0.045
<i>Dx and/or Tx of BP and/or MDI</i>	0.025	0.048	0.037	0.034	0.071	0.059
<i>Dx and/or Tx of any of the above 3</i>	0.035	0.055	0.045	0.079	0.091	0.087
Psychosis & mania screens – unsure or missing						
<i>Dx and/or Tx of SZ and/or SA</i>	0.025	0.020	0.023	0.039	0.023	0.028
<i>Dx and/or Tx of AH and/or delusions</i>	0.015	0.014	0.014	0.034	0.020	0.025
<i>Dx and/or Tx of BP and/or MDI</i>	0.029	0.028	0.028	0.028	0.031	0.030
<i>Dx and/or Tx of any of the above 3</i>	0.042	0.038	0.040	0.056	0.054	0.055
Comorbidity of disorders*						
<i>No disorders and negative psychosis/mania screen</i>	0.516	0.341	0.423	0.461	0.328	0.372
<i>No disorders</i>	0.532	0.348	0.435	0.472	0.339	0.384

Condition	EA Controls				AA Controls (from KN)			
	Males	Females	Total		Males	Females	Total	
Sample Size:	1,578	1,786	3,364		178	351	529	
<i>Any disorder(s) (one or more)</i>	0.468	0.652	0.565		0.528	0.661	0.616	
<i>One disorder</i>	0.215	0.236	0.226		0.253	0.248	0.250	
<i>Two disorders</i>	0.110	0.163	0.138		0.096	0.160	0.138	
<i>Three disorders</i>	0.071	0.097	0.085		0.045	0.094	0.078	
<i>Four disorders</i>	0.032	0.062	0.048		0.034	0.060	0.051	
<i>Five disorders</i>	0.011	0.044	0.029		0.034	0.043	0.040	
<i>Six or more disorders</i>	0.028	0.051	0.040		0.067	0.057	0.060	
<i>Proportion of disorders* in controls with</i>								
<i>One disorder</i>	0.211	0.143	0.167		0.192	0.149	0.163	
<i>Two disorders</i>	0.217	0.198	0.204		0.145	0.191	0.176	
<i>Three disorders</i>	0.209	0.176	0.188		0.103	0.169	0.147	
<i>Four disorders</i>	0.127	0.149	0.142		0.103	0.144	0.130	
<i>Five disorders</i>	0.056	0.132	0.106		0.128	0.128	0.128	
<i>Six or more disorders</i>	0.179	0.201	0.193		0.329	0.219	0.256	

EA and AA (KN) control subjects' condition frequencies. FTND = Fagerström Test for Nicotine Dependence (0–12); BMI = body mass index; Dx = diagnosis; Tx = treatment; SZ = schizophrenia; SA = schizoaffective disorder; AH = auditory hallucinations; BP = bipolar disorder; MDI = manic-depression; disorders* = alcohol or drug dependence, MDE, specific phobia, social phobia, agoraphobia, panic attacks, GAD, or OCD (i.e., excludes nicotine dependence). Disorders are treated dichotomously (present/absent) in this table for the studied sample.