

The Internet-Based MGS2 Control Sample: Self Report of Mental Illness

Alan R. Sanders, M.D.

Douglas F. Levinson, M.D.

Jubao Duan, Ph.D.

J. Michael Dennis, Ph.D.

Rick Li, Ph.D.

Kenneth S. Kendler, M.D.

John P. Rice, Ph.D.

Jianxin Shi, Ph.D.

Bryan J. Mowry, M.D.

Farooq Amin, M.D.

Jeremy M. Silverman, Ph.D.

Nancy G. Buccola, A.P.R.N., B.C.

William F. Byerley, M.D.

Donald W. Black, M.D.

Robert Freedman, M.D.

C. Robert Cloninger, M.D.

Pablo V. Gejman, M.D.

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

Method: Subjects gave consent to deposit phenotypic data and blood samples into a repository for general research use, with full anonymization of the sample. The authors compared the control sample with population census data for demographic data and with previous population surveys for anthropometrics and prevalences of psychiatric disorders as estimated by an Internet-administered questionnaire.

Results: The full MGS2 control sample includes 4,665 subjects (European-ancestry: N=3,364; African American: N=1,301), of whom 3,626 were included in the MGS2 genome-wide association study (GWAS). The sample is generally demographi-

cally 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 use 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 non-Hispanic European-ancestry and African American 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.

(*Am J Psychiatry* 2010; 167:854–865)

The Molecular Genetics of Schizophrenia (MGS2) project 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 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 (Menlo Park, Calif.), Internet recruited self-identified non-Hispanic European-ancestry and African American participants consenting to deposition of anonymized clinical data, a blood sample, and future genetic data in the National Institute of Mental Health (NIMH) repository. We note that the Wellcome Trust Case Control Consortium (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 reference 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 references 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, and 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 $\geq 20\%$ and with only a modest benefit for prevalence $\leq 5\%$.

This article is discussed in an editorial by Drs. Schwartz and Susser (p. 741).

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 idea is to draw from the same source population and independently of exposure status (marker genotype) (9, 10).

Method

Ascertainment

Knowledge Networks enrolled a total of 4,665 adults (18 years or older), 3,364 individuals of non-Hispanic European ancestry and 1,301 individuals of African American ethnicity. All European-ancestry and 529 African American control subjects were recruited from Knowledge Networks' nationwide panel of survey participants, which is recruited continuously by random digit dialing (11, 12) and resembles the population distribution in age, race, Hispanicity, region, employment, marital status, and other demographic features (Table 1 [also see Table S4 in reference 1]). An additional 772 African American controls were recruited by subcontract to Survey Sampling International (Shelton, Conn.), using Internet recruitment (banner ads on websites). Table 1 (ST1) in the data supplement accompanying the online version of this article lists the completion rate for the various stages of the study, from invitation to procedure completion. Of the targeted panel members, 15.5% (African American) to 21.7% (European-ancestry) 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 (two letters and three telephone contacts) achieved a 31% rate (13). Survey Sampling International completed 2.1% of targeted controls, but this was needed to achieve an adequate African American sample size.

Informed Consent

Institutional review board approval was obtained at North-Shore 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 data set with extensive genotypic data, such as this data set, if one also had access to a large number of genotypes and to subjects' identities in a second data set (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 the National Institutes of Health Genome-Wide Association Studies (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 reference 17]) and telephone interviews and is accurate compared with 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. (Irving, Tex.), to arrange blood specimen collection. The phlebotomist confirmed gender and race (European-ancestry or African American) and shipped the blood to the Rutgers University Cell and DNA Repository (Piscataway, N.J.) 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 age: European-ancestry=50.0 years [SD=16.4]; African American=45.4 years [SD=13.1]) 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 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 the discussion below of CIDI–SF dichotomous versus probabilistic methods and the disorder specific criteria discussions).

Ancestry

The questionnaire allowed controls to select ≥ 1 ancestry (race/ethnicity) for each grandparent, from 15 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 (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 15 categories for the present analysis.

Quality Control of Genotypic Data

Initially, we genotyped 14 schizophrenia candidate genes in 2,126 European-ancestry control and 1,952 European-ancestry case individuals (1). Extensive sample quality control (evaluating genotyping call rates, sex typing discrepancies, cryptic duplicates or relatives, and outliers in an analysis of ancestry-informative markers) 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 two; $\chi^2=14.45$, $df=1$, $p=0.00014$), 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 European-ancestry and African American distributions in analyses of ancestry-informative markers (63 versus 35; $\chi^2=10.18$, $df=1$, $p=0.0014$), 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 (European-ancestry: $N=2,817$; African American: $N=1,010$) of the 4,665 collected control samples and 4,196 (European-ancestry:

TABLE 1. Demographic Characteristics of European-Ancestry and African American Control Sample Relative to the U.S. Adult Population^a

Variable and Category	All Control Subjects		African American Control Subjects by Source		U.S. Population	
	European-Ancestry (N=3,364)	African American (N=1,301)	Knowledge Networks (N=529)	Survey Sampling International (N=772)	European-Ancestry	African American
Gender						
Male	46.9	39.5	33.6^b	43.5	47.9	42.1
Female	53.1	60.5	66.4^b	56.5	52.1	57.9
Age (years)						
18–24	5.0^b	5.7^b	5.7^b	5.7^b	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 ^b	30.1 ^b	31.1 ^b	20.7	20.3
55–64	18.1 ^b	16.8 ^b	17.0 ^b	16.7 ^b	14.8	13.0
≥65	21.8 ^b	6.9^b	9.5^b	5.2^b	18.7	14.2
Employment status						
In the labor force	57.1^b	59.3	55.6	61.8	65.0	58.3
Not in the labor force	42.9 ^b	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						
<High school diploma	7.9^b	4.0^b	6.8^b	2.1^b	10.4	20.0
High school diploma or equivalent	27.2^b	15.8^b	18.5^b	13.9^b	33.2	36.1
Some college	22.9 ^b	39.7 ^b	37.2 ^b	41.5 ^b	19.6	20.3
Associate degree	7.6^b	11.8 ^b	11.7 ^b	11.8 ^b	8.7	7.5
≥Bachelor's degree	34.3 ^b	28.7 ^b	25.7 ^b	30.7 ^b	28.1	16.1
Household income						
<\$10,000	5.3^b	10.4^b	13.6^b	8.2^b	6.6	17.8
\$10,000–\$24,999	17.1	23.0	25.9	21.0^b	18.2	25.8
\$25,000–\$49,999	33.1	32.4	28.7	34.8	34.0	31.9
\$50,000–\$74,999	23.8^b	17.1	17.2	17.1	25.6	16.3
≥\$75,000	20.7 ^b	16.1 ^b	14.6 ^b	17.2	15.6	8.1
Census region						
Northeast	18.3^b	16.1	15.7	16.3	23.9	17.5
Midwest	27.1	25.4 ^b	25.0 ^b	25.6 ^b	27.9	19.0
South	36.8 ^b	52.3	51.6	52.7	26.6	54.6
West	17.9^b	6.3^b	7.8	5.3^b	21.6	8.9

^a Values represent percentages of the control sample or subgroup or of the general population in each category for each variable, and data indicate self-identified European-ancestry and African American control subjects with blood data from complete collection (marital status not provided from Survey Sampling International, Inc.). The Knowledge Networks adult panel makeup data from January 2004 and adult U.S. Population data from the Current Population Survey in November 2003 (25) had previously been shown to match well (1). November 2003 Current Population Survey data presented is restricted to non-Hispanic adults listing white-only or black-only for race in the survey.

^b Values indicate differences in the control sample data ($p < 0.05$ based on chi-square comparison of counts in the respective categories) relative to the U.S. population data for European-ancestry or African American control subjects, as appropriate, with data in bold indicating a value lower than the U.S. population data.

N=2,838 ; African American: N=1,358) case samples with the Affymetrix 6.0 array; 3,626 (95%) control and 3,967 (95%) case samples passed stringent quality control (see Table 2 [ST2] in the data supplement). 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 Phe-

notypes [dbgap.ncbi.nlm.nih.gov], Study Accessions: phs000021.v2.p1 and phs000167.v1.p1), with 177 research projects already (December 7, 2009) 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

TABLE 2. Summary of Online Self-Report^a

Phenotype	Measure
Common psychiatric disorders	Composite International Diagnostic Interview–Short Form ^a (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 (38) score.
Personality	12-item Eysenck Personality Questionnaire Brief Version 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 Knowledge Networks and Survey Sampling International (confer with Table 1).

^a The modified questionnaire (Composite International Diagnostic Interview–Short Form 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).

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 (European-ancestry, African American–Knowledge Networks, and African American–Survey Sampling International) with the November 2003 U.S. Current Population Survey (25) for non-Hispanic adults reporting a single race (white or black). Differences between the control samples and the Current Population Survey include the following: more women in the African American–Knowledge Networks subsample; older mean age for subjects of European-ancestry (mean age for control samples=50.0 years [SD=16.4] versus 47.5 years [SD=17.1] for Current Population Survey, *t* test: $p<0.0001$); fewer employed European-ancestry subjects, perhaps reflecting older age (mean age for European-ancestry employed subjects was 43.8 years [SD=12.9] versus 58.1 years [SD=17.1] for nonworkers, *t* test: $p<0.0001$); European-ancestry and particularly African American control subjects were more educated; more European-ancestry subjects were in the highest income bracket, and more African American subjects (especially African American–Survey Sampling International) were in several upper brackets; more European-ancestry subjects were from the South and fewer from the Northeast and West, while more African American subjects were from the Midwest.

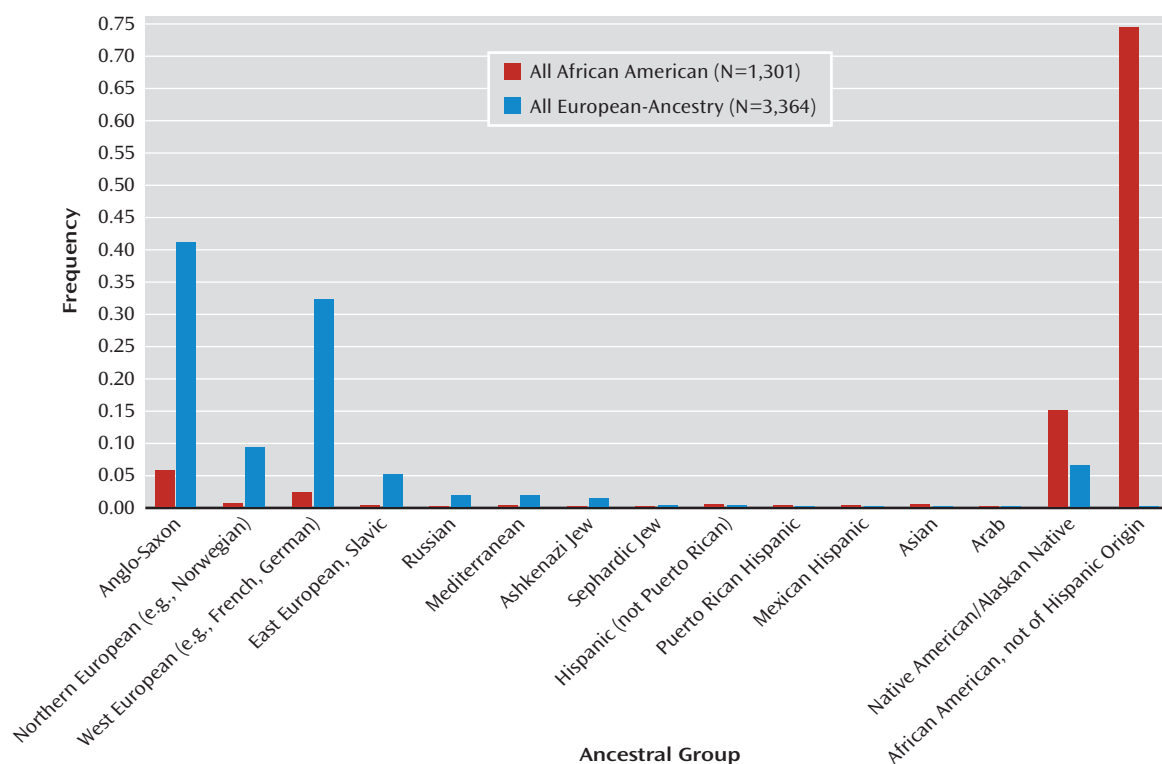
Ancestry

Numerical data for ≥ 2 grandparents were reported by 87% of subjects of European ancestry and 96% of African American 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% of European-ancestry and 80% of African American sub-

jects) versus those with any missing data (see Figures SF1 and SF2 in the data supplement). Data for all subjects (ignoring missing data) are shown in Figure 1. For European-ancestry 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 European-ancestry control sample and the U.S. Census show similar distributions of subcontinental European ancestries (see Figure SF3 in the data supplement).

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

Most European-ancestry 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 and 7% with all Western 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 reference 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).

FIGURE 1. Frequency of Self-Reported European-Ancestry and African American Control Subjects^a

^a Data show the results of the tabulation of the frequency with which self-identified European-ancestry and African American control subjects endorsed biological grandparents, with the ancestral background indicated on the x-axis (one particular grandparent may have more than one ancestry endorsed).

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 (European-ancestry, African American–Knowledge Networks); Table 4 provides additional data. The African American–Survey Sampling International subsample (Internet banner ad recruitment) has higher lifetime prevalence for common diagnoses, more comorbidity, higher neuroticism scores, and lower extraversion scores compared with the Knowledge Networks subsample that was recruited from a survey panel and is thus presented separately (see Tables ST3 and ST4 in the data supplement). Table 5 in the data supplement (ST5) includes comparison lifetime prevalence data from several epidemiological surveys (for all ancestries combined) for the Epidemiological Catchment Area Study (30), the National Comorbidity Survey (31), the National Comorbidity Survey Replication (32, 33), and the National Epidemiologic Survey on Alcohol and Related Conditions–Wave I (34, 35). Generally, male:female prevalence ratios were similar to those in previous studies for diagnoses compared with previous work (Table 3, Table 4 [also see ST5 in the data supplement]), and European-ancestry:African American–Knowledge Networks ratios were comparable within our study, but lifetime prevalences are higher in our sample (Table 3) than in the National Comorbidity Survey (31), which used CIDI. For

example, at least one anxiety disorder diagnosis (other than obsessive-compulsive disorder [OCD]) was assigned in 40.6% of European-ancestry and 42.9% of African American–Knowledge Networks subjects versus 24.9% in the National Comorbidity Survey (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 overreporting (20). By this method (see ST5 in the data supplement), 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 European-ancestry control subjects versus National Comorbidity Survey subjects for alcohol dependence (23.1% versus 14.1%) and drug dependence (12.2% versus 7.5%), with the expected excess in male subjects (Table 3). There was higher prevalence of alcohol dependence in European-ancestry versus African American control subjects (23.1% versus 17.8%; $\chi^2=7.17$, $df=1$, $p=0.0074$) but higher prevalence of drug dependence in African American–Knowledge Networks versus

TABLE 3. Common Psychiatric Diagnoses (DSM–IV From Composite International Diagnostic Interview–Short Form) in European-Ancestry and African American (Knowledge Networks) Control Sample^a

Condition	European-Ancestry Control Subjects			African American Control Subjects (From Knowledge Networks)		
	Men (N=1,578)	Women (N=1,786)	Total (N=3,364)	Men (N=178)	Women (N=351)	Total (N=529)
Substance use disorders						
Alcohol dependence	0.295	0.175	0.231	0.258	0.137	0.178
Drug dependence	0.139	0.108	0.122	0.213	0.137	0.163
Any substance dependence above	0.331	0.211	0.268	0.348	0.219	0.263
Substance-induced psychological symptoms (SIPS)	0.179	0.153	0.165	0.230	0.171	0.191
Major depressive episodes						
All	0.292	0.500	0.402	0.303	0.481	0.422
Mean age at onset	28.9	25.5	26.6	26.5	24.8	25.4
Single episode	0.077	0.101	0.090	0.051	0.083	0.072
Recurrent	0.215	0.399	0.313	0.253	0.399	0.350
Without SIPS	0.194	0.382	0.294	0.185	0.362	0.302
Without substance dependence	0.146	0.350	0.254	0.163	0.316	0.265
Generalized anxiety disorder						
All	0.143	0.255	0.202	0.135	0.219	0.191
Without major depressive episode	0.027	0.030	0.029	0.017	0.023	0.021
Without substance dependence	0.064	0.164	0.117	0.045	0.131	0.102
Anxiety disorders						
Specific phobia	0.132	0.233	0.185	0.213	0.251	0.238
Social phobia	0.114	0.169	0.143	0.135	0.154	0.147
Agoraphobia without panic attacks	0.020	0.038	0.029	0.039	0.060	0.053
Panic attacks	0.092	0.196	0.147	0.129	0.191	0.170
Panic attacks without agoraphobia	0.068	0.146	0.109	0.101	0.142	0.129
Panic attacks with agoraphobia	0.023	0.050	0.038	0.028	0.048	0.042
Generalized anxiety disorder	0.143	0.255	0.202	0.135	0.219	0.191
Any anxiety disorder above (generalized anxiety disorder, 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 without substance dependence	0.156	0.334	0.251	0.163	0.330	0.274
Obsessive-compulsive disorder	0.061	0.100	0.082	0.118	0.125	0.123

^a Data indicate European-ancestry and African American (Knowledge Networks) control subjects' condition frequencies; disorders are treated dichotomously (present/absent) in the present table for the study sample.

European-ancestry subjects (16.3% versus 12.2%; $\chi^2=6.23$, $df=1$, $p=0.013$). In the full (European-ancestry plus African American) Knowledge Networks sample, alcohol dependence showed the expected increase in men versus women (29.1% versus 16.8%, $\chi^2=83$, $df=1$, $p<0.0001$). Also, Knowledge Networks subjects diagnosed with drug dependence were younger (mean age: 43.7 years [$SD=11.8$] versus 50.3 years [$SD=16.5$], t test: $p<0.0001$) and had lower household income (<\$50,000 for 66.9% versus 55.8%, $\chi^2=21.4$, $df=1$, $p<0.0001$) but were more likely to be employed (62.0% versus 56.1%, $\chi^2=6$, $df=1$, $p=0.014$).

Major Depression

Our European-ancestry control subjects have a much higher lifetime prevalence of major depressive episodes compared with the National Comorbidity Survey sample (40.2% versus 17.1%; reduced to 34.6% for control subjects using probability of caseness as shown in ST5 in the data supplement); among European-ancestry subjects with major depressive episodes, mean age of onset was

26.6 years [$SD=13.8$], and 78% reported recurrent major depressive episodes (Table 3). It is known that CIDI-SF overdiagnoses major depression (37). Without interview data, we cannot determine whether there might be a true overrepresentation of individuals with depression in the Knowledge Networks panel, although this appears unlikely given that the controls scored lower than population means for the neuroticism scale (see below), which is often elevated in depression. CIDI-SF major depressive episode criteria ignore distress/impairment (20), although queried; if we restrict major depressive episode diagnosis to those who sought help, received treatment, or reported a lot or some impairment, lifetime prevalence was reduced by 3.1% for European-ancestry control subjects and by 7.3% for African American–Knowledge Networks control subjects. CIDI-SF also fails to differentiate major depressive episodes from the effects of substance use. Lifetime prevalence of major depressive episodes was 25.4% in controls without alcohol or drug dependence and 29.4% in controls who did not en-

TABLE 4. Additional Phenotypic Information in European-Ancestry and African American (Knowledge Networks) Control Sample^a

Condition	European-Ancestry Control Subjects			African American Control Subjects (From Knowledge Networks)		
	Men (N=1,578)	Women (N=1,786)	Total (N=3,364)	Men (N=178)	Women (N=351)	Total (N=529)
Other disorders, traits, and conditions						
Nicotine dependence (Fagerström Test for Nicotine Dependence score ≥ 4)	0.372	0.315	0.342	0.326	0.282	0.297
Mean Fagerström Test for Nicotine Dependence score, all	2.6	2.3	2.4	2.4	2.0	2.1
Mean Eysenck Personality Questionnaire Brief Version neuroticism score (0–12)	2.9	4.4	3.7	3.0	4.2	3.8
Mean Eysenck Personality Questionnaire Brief Version extra-version score (0–12)	6.8	6.9	6.9	7.7	7.6	7.7
Sexual identity						
Bisexual	0.025	0.032	0.029	0.039	0.023	0.028
Homosexual	0.029	0.015	0.021	0.034	0.006	0.015
Overweight or obese (current body mass index ≥ 25 kg/m ²)	0.688	0.611	0.647	0.685	0.738	0.720
Obese (current body mass index ≥ 30 kg/m ²)	0.314	0.368	0.342	0.404	0.501	0.469
Mean height (m)	1.79	1.64	1.71	1.79	1.65	1.70
Mean highest lifetime body mass index (kg/m ²)	31.0	32.4	31.8	31.8	35.9	34.5
Psychosis and mania screens—endorsed						
Diagnosis and/or treatment of schizophrenia and/or schizoaffective disorder	0.008	0.007	0.007	0.028	0.020	0.023
Diagnosis and/or treatment of auditory hallucinations and/or delusions	0.013	0.016	0.015	0.045	0.046	0.045
Diagnosis and/or treatment of bipolar disorder and/or manic depression	0.025	0.048	0.037	0.034	0.071	0.059
Diagnosis and/or treatment of any of the above three	0.035	0.055	0.045	0.079	0.091	0.087
Psychosis and mania screens—unsure or missing						
Diagnosis and/or treatment of schizophrenia and/or schizoaffective disorder	0.025	0.020	0.023	0.039	0.023	0.028
Diagnosis and/or treatment of auditory hallucinations and/or delusions	0.015	0.014	0.014	0.034	0.020	0.025
Diagnosis and/or treatment of bipolar disorder and/or manic depression	0.029	0.028	0.028	0.028	0.031	0.030
Diagnosis and/or treatment of any of the above three	0.042	0.038	0.040	0.056	0.054	0.055
Comorbidity of disorders^b						
No disorders and negative psychosis/mania screen	0.516	0.341	0.423	0.461	0.328	0.372
No disorders	0.532	0.348	0.435	0.472	0.339	0.384
Any disorder(s) (one or more)	0.468	0.652	0.565	0.528	0.661	0.616
One disorder	0.215	0.236	0.266	0.253	0.248	0.250
Two disorders	0.110	0.163	0.138	0.096	0.160	0.138
Three disorders	0.071	0.097	0.085	0.045	0.094	0.078
Four disorders	0.032	0.062	0.048	0.034	0.060	0.051
Five disorders	0.011	0.044	0.029	0.034	0.043	0.040
Six or more disorders	0.028	0.051	0.040	0.067	0.057	0.060
Proportion of disorders^b in control subjects with:						
One disorder	0.211	0.143	0.167	0.192	0.149	0.163
Two disorders	0.217	0.198	0.204	0.145	0.191	0.176
Three disorders	0.209	0.176	0.188	0.103	0.169	0.147
Four disorders	0.127	0.149	0.142	0.103	0.144	0.130
Five disorders	0.056	0.132	0.106	0.128	0.128	0.128
Six or more disorders	0.179	0.201	0.193	0.329	0.219	0.256

^a Data indicate European-ancestry and African American (Knowledge Networks) control subjects' condition frequencies; disorders are treated dichotomously (present/absent) in the present table for the study sample.

^b Disorders include alcohol or drug dependence, major depressive episodes, specific phobia, social phobia, agoraphobia, panic attacks, generalized anxiety disorder, and obsessive-compulsive disorder (i.e., excludes nicotine dependence).

dorse 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 African American–Knowledge Networks

subjects. In the full Knowledge Networks control sample, subjects with major depressive episodes were more likely to be female (men, 29.3%, versus women, 49.7%, $\chi^2=165$, $df=1$, $p<0.0001$), younger (mean age: 46.8 years [SD=14.5]

versus 51.3 years [SD=17.0], *t* test: $p < 0.0001$), unmarried (48.8% versus 34.8%, $\chi^2=75$, *df*=1, $p < 0.0001$), and to have lower household income (<\$50,000: 63.3% versus 53.0%, $\chi^2=41$, *df*=1, $p < 0.0001$).

Anxiety Disorders

Self-reported prevalence of generalized anxiety disorder was higher in European-ancestry control subjects than in the National Comorbidity Survey (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 major depressive episodes (though CIDI-SF does not ask about relative timing of generalized anxiety disorder symptoms versus major depressive episodes). We found a similar pattern and estimates for African American–Knowledge Networks subjects. All anxiety disorders were more prevalent in women. Prevalence of CIDI-SF anxiety disorders was similar in European-ancestry versus African American–Knowledge Networks subjects except for specific phobia (18.5% European-ancestry versus 23.8% African American; $\chi^2=7.82$, *df*=1, $p=0.0052$), agoraphobia without panic attacks (2.9% European-ancestry versus 5.3% African American; $\chi^2=7.27$, *df*=1, $p=0.0070$), and OCD (8.2% European-ancestry versus 12.3% African American; $\chi^2=9.03$, *df*=1, $p=0.0027$).

Nicotine Dependence

Nicotine dependence, defined as a Fagerström Test for Nicotine Dependence score ≥ 4 (38), was more common in men in our full Knowledge Networks control sample (36.7% versus 30.9%, $\chi^2=14.3$, *df*=1, $p=0.0002$), in European-ancestry versus African American–Knowledge Networks subjects (34.2% versus 29.7%, $\chi^2=3.91$, *df*=1, $p=0.048$), and in those with less education (high school completion or less for 45.7% of subjects with a Fagerström Test for Nicotine Dependence score ≥ 4 versus 27.8% for the rest, $\chi^2=124$, *df*=1, $p < 0.0001$).

Neuroticism and Extraversion

In European-ancestry control subjects, mean scores for neuroticism were 2.9 in male subjects and 4.4 in female subjects, compared with reported population means of 4.95 and 5.90, respectively (39); and for extraversion they were 6.8 for male subjects and 6.9 for female subjects (versus population means of 6.36 and 7.60, respectively). Mean scores were similar in African American–Knowledge Networks subjects except for higher extraversion scores (7.7 versus 6.9 in European-ancestry control subjects). Control subjects meeting CIDI-SF criteria for major depressive episode had higher mean neuroticism scores (5.8 versus 2.3 for European-ancestry control subjects, and 5.8 versus 2.4 for African American–Knowledge Networks control subjects) and lower mean extraversion scores (6.1 versus 7.4 for European-ancestry subjects, and 6.5 versus 8.5 for African American–Knowledge Networks subjects), compared with subjects with

no self-reported major depressive episode. 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 approximately 60% of individuals with at least one. Subjects with three or more diagnoses (20.2% of European-ancestry and 22.9% of African American–Knowledge Networks 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 European-ancestry and 37.2% of African American–Knowledge Networks subjects who received no CIDI-SF diagnoses and who denied all psychosis and bipolar screens.

Other Conditions and Traits

European-ancestry and African American subjects were similar in height, with male subjects approximately 15 cm taller than female subjects. Based on current body mass index (body mass index=kg/m²) and on highest nonpregnancy body mass index, the majority of subjects were either overweight (25 kg/m² \leq body mass index <30 kg/m²) or obese (body mass index ≥ 30 kg/m²), particularly African American women, similar to contemporaneous estimates from the National Health and Nutrition Examination Study (40). Similar to previous work (e.g., reference 41), in terms of sexual orientation (identity) for the European-ancestry 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 African American–Knowledge Networks sample, although the male:female ratios are slightly larger.

Psychosis and Mania

The proportion of European-ancestry control subjects endorsing previous treatment or diagnosis of schizophrenia or schizoaffective disorder was identical to that in the National Comorbidity Survey sample (0.7%) (31), but more of our control sample reported treatment or diagnosis of bipolar disorder (3.7% versus 1.6%). African American–Knowledge Networks subjects reported more schizophrenia/schizoaffective (2.3% versus 0.7% for European-ancestry subjects; $\chi^2=12.9$, *df*=1, $p=0.0003$) and bipolar disorder histories (5.9% versus 3.7% for European-ancestry subjects; $\chi^2=4.75$, *df*=1, $p=0.029$). 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 MGS2 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 Knowledge Networks 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 United States 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 European-ancestry subjects from the South and African American 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 European-ancestry and African American subjects, except that Native American ancestry appears to have been overreported (2). Molecular quality control 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 (Knowledge Networks, Survey Sampling International, Examination Management Services, Inc.), 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 National Comorbidity Survey 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 European-ancestry and African American subjects. Male:female ratios and patterns of comorbidity are similar to those reported previously (Table 3 [also see ST5 in the data supplement]).

4. The African American-Survey Sampling International subsample that was recruited by Internet ads had more CIDI-SF diagnoses and comorbidity, higher neuroticism scores, and lower extraversion scores compared with African American-Knowledge Networks panel subjects, consistent with previous work (42). We were unable to recruit the entire African American control sample through the more representative Knowledge Networks panel because our required African American:European-ancestry ratio (approximately 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 in reference 43) among African American versus European-ancestry panel members (15.5% versus 21.7% [see ST1 in the data supplement]).

Several factors could have contributed to the high rate of common psychiatric diagnoses (also see the Discussion in the data supplement for disorder-specific information). CIDI-SF generally overdiagnoses these disorders (e.g., references 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 nonpathological. 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 reference 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 coaggregate 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 an approximately 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., references 4, 8). Efficient collection of a large, anonymized control sample, generally representative of the U.S. adult European-ancestry and African American 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.

Received July 27, 2009; revisions received Dec. 14, 2009 and Jan. 14, 2010; accepted Jan. 15, 2010 (doi: 10.1176/appi.ajp.2010.09071050). From the Center for Psychiatric Genetics, Department of Psychiatry and Behavioral Sciences, NorthShore University HealthSystem, Evanston, Ill.; the Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, Calif.; Government and Academic Research, Knowledge Networks, Menlo Park, Calif.; the Departments of Psychiatry and Human Genetics, Virginia Commonwealth University, Richmond, Va.; the Department of Psychiatry, Washington University, St. Louis, Mo.; the Queensland Centre for Mental Health Research and Queensland Brain Institute, University of Queensland, Brisbane, Queensland, Australia; the Department of Psychiatry and Behavioral Sciences, Atlanta Veterans Affairs Medical Center and Emory University, Atlanta; the Department of Psychiatry, Mount Sinai School of Medicine, New York; the School of Nursing, Louisiana State University Health Sciences Center, New Orleans; the Department of Psychiatry, University of California at San Francisco, San Francisco; the Mental Health Clinical Research Center and Department of Psychiatry, University of Iowa Carver College of Medicine, Iowa City, IA; the Department of Psychiatry, University of Colorado Denver, Aurora, Colo. Address correspondence and reprint requests to Dr. Sanders, NorthShore University HealthSystem Research Institute, 1001 University Place, Evanston, IL 60201-3137; asanders@northshore.org (e-mail).

Dr. Black has received research support from AstraZeneca and Forest Laboratories; he is also a shareholder with Johnson and Johnson. Drs. Sanders, Levinson, Duan, Dennis, Li, Kendler, Rice, Shi, Mowry, Amin, Silverman, Byerley, Freedman, Cloninger, and Gejman and Ms. Buccola report no financial relationships with commercial interests.

The authors thank the individuals who volunteered for the control sample for their participation. Data and biomaterials from the National Institute of Mental Health Genetics Initiative MGS2 (Molecular Genetics of Schizophrenia) control sample were collected by NorthShore University HealthSystem, Evanston, Ill., R01 MH-59571, Pablo V. Gejman, M.D. (collaboration coordinator; principal investigator), as part of a collaborative R01 application comprised of 10 sites. From 2003–2006, the principal investigators and co-investigators were: NorthShore University HealthSystem, Evanston, Ill., R01 MH-59571, Pablo V. Gejman, M.D. (collaboration coordinator; principal investigator), Alan R. Sanders, M.D.; Emory University School of Medicine, Atlanta, R01 MH-59587, Farooq Amin, M.D. (principal investigator); University of California, San Francisco, R01 MH-60870, William F. Byerley, M.D. (principal investigator); University of Iowa, Iowa City, IA,

R01 MH-59566, Donald W. Black, M.D. (principal investigator), Raymond R. Crowe, M.D.; Washington University, St. Louis, Mo., R01 MH-60879, C. Robert Cloninger, M.D. (principal investigator); University of Colorado, Denver, R01 MH-59565, Robert Freedman, M.D. (principal investigator), Ann Olincy, M.D.; Stanford University, Palo Alto, Calif., R01 MH-61675, Douglas F. Levinson, M.D. (principal investigator); Louisiana State University, New Orleans, R01 MH-67257, Nancy G. Buccola, A.P.R.N., B.C., M.S.N. (principal investigator); University of Queensland, Brisbane, Queensland, Australia, R01 MH-59588, Bryan J. Mowry, M.D. (principal investigator); Mt. Sinai School of Medicine, New York, R01 MH-59586, Jeremy M. Silverman, Ph.D. (principal investigator). Our schizophrenia GWAS (genotyping and analyses) was supported by the Genetic Association Information Network (GAIN) and grants U01MH-79469 (Pablo V. Gejman, M.D.), U01MH-79470 (Douglas F. Levinson, M.D.), and R01MH-81800 (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

1. Sanders AR, Duan J, Levinson DF, Shi J, He D, Hou C, Burrell GJ, Rice JP, Nertney DA, Olincy A, Rozic P, Vinogradov S, Buccola NG, Mowry BJ, Freedman R, Amin F, Black DW, Silverman JM, Byerley WF, Crowe RR, Cloninger CR, Martinez M, Gejman PV: No significant association of 14 candidate genes with schizophrenia in a large European ancestry sample: implications for psychiatric genetics. *Am J Psychiatry* 2008; 165:497–506
2. Shi J, Levinson DF, Duan J, Sanders AR, Zheng Y, Pe'er I, Dudgeon F, Holmans PA, Whittemore AS, Mowry BJ, Olincy A, Amin F, Cloninger CR, Silverman JM, Buccola NG, Byerley WF, Black DW, Crowe RR, Oksenberg JR, Mirel DB, Kendler KS, Freedman R, Gejman PV: Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature* 2009; 460:753–757
3. Wellcome Trust Case Control Consortium: Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007; 447:661–678
4. McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, Hirschhorn JN: Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* 2008; 9:356–369
5. Schechter D, Lebovitch R: Normal controls are expensive to find: methods to improve cost-effectiveness of the screening evaluation. *Psychiatry Res* 2005; 136:69–78
6. Adami H, Elliott A, Zettlmeisl M, McMahon R, Thaker G: Use of telephone screens improves efficiency of healthy subject recruitment. *Psychiatry Res* 2002; 113:295–301
7. Heeren T, Edwards EM, Dennis JM, Rodkin S, Hingson RW, Rosenbloom DL: 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; 32:222–229
8. Morton NE, Collins A: Tests and estimates of allelic association in complex inheritance. *Proc Natl Acad Sci U S A* 1998; 95:11389–11393
9. Rothman KJ, Greenland S, Lash TL: Case-control studies, in *Modern Epidemiology*. Edited by Rothman KJ, Greenland S, Lash TL. Philadelphia, Lippincott Williams and Wilkins, 2008, pp 111–127
10. Schwartz S, Susser E, Gorman JM: Choosing controls in biologic psychiatry, in *Psychiatric Epidemiology: Searching for the Causes of Mental Disorders*. Edited by Susser E, Schwartz S, Morabia A, Bromet EJ. New York, Oxford University Press, 2006, pp 247–259
11. Couper M: Web surveys: a review of issues and approaches. *Public Opin Q* 2000; 64:464–494

12. Krotki K, Dennis JM: Probability-based survey research on the Internet, in *Proceedings of the Conference of the International Statistical Institute*. Seoul, South Korea, 2001
13. Firmann M, Mayor V, Vidal PM, Bochud M, Pécoud A, Hayoz D, Paccaud F, Preisig M, Song KS, Yuan X, Danoff TM, Stirnadel HA, Waterworth D, Mooser V, Waeber G, Vollenweider P: 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
14. Homer N, Szelinger S, Redman M, Duggan D, Tembe W, Muehling J, Pearson JV, Stephan DA, Nelson SF, Craig DW: Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. *PLoS Genet* 2008; 4:e1000167
15. Wittchen HU: Reliability and validity studies of the World Health Organization: Composite International Diagnostic Interview (CIDI): a critical review. *J Psychiatr Res* 1994; 28:57–84
16. Kessler RC, Wittchen H-U, Abelson JM, McGonagle K, Schwarz N, Kendler KS, Knauper B, Zhao S: Methodological studies of the Composite International Diagnostic Interview (CIDI) in the US National Comorbidity Survey. *Int J Methods Psychiatr Res* 1998; 7:33–55
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, Andrews G, Mroczek D, Üstün TB, Wittchen H-U: The World Health Organization Composite International Diagnostic Interview Short-Form (CIDI–SF). *Int J Methods Psychiatr Res* 1998; 7:171–185
19. Patten SB, Brandon-Christie J, Devji J, Sedmak B: Performance of the Composite International Diagnostic Interview Short Form for major depression in a community sample. *Chronic Dis Can* 2000; 21:68–72
20. Nelson CB, Kessler RC, Mroczek D: Scoring the World Health Organization's Composite International Diagnostic Interview Short Form (CIDI–SF, version 1.0). Geneva, 2001. <http://www.who.int/whosis/en/>
21. American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, 4th ed (DSM–IV). Washington, DC, American Psychiatric Publishing, 1994
22. Nurnberger JI Jr, Blehar MC, Kaufmann CA, York-Cooler C, Simpson SG, Harkavy-Friedman J, Severe JB, Malaspina D, Reich T: Diagnostic Interview for Genetic Studies: rationale, unique features, and training: NIMH Genetics Initiative. *Arch Gen Psychiatry* 1994; 51:849–859
23. Sklar P, Smoller JW, Fan J, Ferreira MA, Perlis RH, Chambert K, Nimgaonkar VL, McQueen MB, Faraone SV, Kirby A, de Bakker PI, Ogdie MN, Thase ME, Sachs GS, Todd-Brown K, Gabriel SB, Sougnez C, Gates C, Blumenstiel B, Defelice M, Ardlie KG, Franklin J, Muir WJ, McGhee KA, MacIntyre DJ, McLean A, VanBeck M, McQuillan A, Bass NJ, Robinson M, Lawrence J, Anjorin A, Curtis D, Scolnick EM, Daly MJ, Blackwood DH, Gurling HM, Purcell SM: Whole-genome association study of bipolar disorder. *Mol Psychiatry* 2008; 13:558–569
24. Scott LJ, Muglia P, Kong XQ, Guan W, Flickinger M, Upmanyu R, Tozzi F, Li JZ, Burmeister M, Absher D, Thompson RC, Francks C, Meng F, Antoniadis A, Southwick AM, Schatzberg AF, Bunney WE, Barchas JD, Jones EG, Day R, Matthews K, McGuffin P, Strauss JS, Kennedy JL, Middleton L, Roses AD, Watson SJ, Vincent JB, Myers RM, Farmer AE, Akil H, Burns DK, Boehnke M: Genome-wide association and meta-analysis of bipolar disorder in individuals of European ancestry. *Proc Natl Acad Sci U S A* 2009; 106:7501–7506
25. US Census Bureau: *Current Population Survey*. US Census Bureau, Washington, DC, 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; 138:375–383
27. Parra EJ, Marcini A, Akey J, Martinson J, Batzer MA, Cooper R, Forrester T, Allison DB, Deka R, Ferrell RE, Shriver MD: Estimating African American admixture proportions by use of population-specific alleles. *Am J Hum Genet* 1998; 63:1839–1851
28. Smith MW, Patterson N, Lautenberger JA, Truelove AL, McDonald GJ, Waliszewska A, Kessing BD, Malasky MJ, Scafe C, Le E, De Jager PL, Mignault AA, Yi Z, De The G, Essex M, Sankale JL, Moore JH, Poku K, Phair JP, Goedert JJ, Vlahov D, Williams SM, Tishkoff SA, Winkler CA, De La Vega FM, Woodage T, Sninsky JJ, Hafler DA, Altshuler D, Gilbert DA, O'Brien SJ, Reich D: A high-density admixture map for disease gene discovery in African Americans. *Am J Hum Genet* 2004; 74:1001–1013
29. Cheng CY, Kao WH, Patterson N, Tandon A, Haiman CA, Harris TB, Xing C, John EM, Ambrosone CB, Brancati FL, Coresh J, Press MF, Parekh RS, Klag MJ, Meoni LA, Hsueh WC, Fejerman L, Pawlikowska L, Freedman ML, Jandorf LH, Bandera EV, Ciupak GL, Nalls MA, Akyzbekova EL, Orwoll ES, Leak TS, Miljkovic I, Li R, Ursin G, Bernstein L, Ardlie K, Taylor HA, Boerwinkle E, Zmuda JM, Henderson BE, Wilson JG, Reich D: Admixture mapping of 15,280 African Americans identifies obesity susceptibility loci on chromosomes 5 and X. *PLoS Genet* 2009; 5:e1000490
30. Robins LN, Regier DA: *Psychiatric Disorders in America: The Epidemiological Catchment Area Study*. New York, Free Press, 1991
31. Kessler RC, McGonagle KA, Zhao S, Nelson CB, Hughes M, Eshleman S, Wittchen HU, Kendler KS: 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; 51:8–19
32. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE: Lifetime prevalence and age-of-onset distributions of DSM–IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry* 2005; 62:593–602
33. Kessler RC, Berglund P, Chiu WT, Demler O, Heeringa S, Hiripi E, Jin R, Pennell BE, Walters EE, Zaslavsky A, Zheng H: The US National Comorbidity Survey Replication (NCS–R): design and field procedures. *Int J Methods Psychiatr Res* 2004; 13:69–92
34. Grant BF, Stinson FS, Dawson DA, Chou SP, Dufour MC, Compton W, Pickering RP, Kaplan K: 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; 61:807–816
35. Conway KP, Compton W, Stinson FS, Grant BF: 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; 67:247–257
36. World Health Organization: *World Health Organization Composite International Diagnostic Interview, Version 1.0*. Geneva, World Health Organization, 1990
37. Aalto-Setälä T, Haarasilta L, Marttunen M, Tuulio-Henriksson A, Poikolainen K, Aro H, Lönnqvist J: Major depressive episode among young adults: CIDI–SF versus SCAN consensus diagnoses. *Psychol Med* 2002; 32:1309–1314
38. Heatherton TF, Kozlowski LT, Frecker RC, Fagerström KO: The Fagerström Test for Nicotine Dependence: a revision of the Fagerström Tolerance Questionnaire. *Br J Addict* 1991; 86:1119–1127
39. Eysenck SGB, Eysenck HJ, Barrett P: A revised version of the Psychoticism Scale. *Pers Individ Dif* 1985; 6:21–29
40. Wang Y, Beydoun MA, Liang L, Caballero B, Kumanyika SK: Will all Americans become overweight or obese? Estimating the

- progression and cost of the US obesity epidemic. *Obesity* (Silver Spring) 2008; 16:2323–2330
41. Laumann EO, Gagnon JH, Michael RT, Michaels S: Homosexuality, in *The Social Organization of Sexuality: Sexual Practices in the United States*. Chicago, University of Chicago Press, 1994, pp 283–321
 42. Halbreich U, Bakhai Y, Bacon KB, Goldstein S, Asnis GM, Endicott J, Lesser J: The normalcy of self-proclaimed “normal volunteers.” *Am J Psychiatry* 1989; 146:1052–1055
 43. Bussey-Jones J, Henderson G, Garrett J, Moloney M, Blumenthal C, Corbie-Smith G: Asking the right questions: views on genetic variation research among black and white research participants. *J Gen Intern Med* 2009; 24:299–304
 44. Orlando M, Burnam MA, Beckman R, Morton SC, London AS, Bing EG, Fleishman JA: 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; 11:75–82
 45. Johnson TP, Hougland JG, Moore RW: Sex differences in reporting sensitive behavior: a comparison of interview methods. *Sex Roles* 1991; 24:669–679
 46. 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; 6:315–323
 47. Paulsen AS, Crowe RR, Noyes R, Pfohl B: Reliability of the telephone interview in diagnosing anxiety disorders. *Arch Gen Psychiatry* 1988; 45:62–63
 48. Slutske WS, True WR, Scherrer JF, Goldberg J, Bucholz KK, Heath AC, Henderson WG, Eisen SA, Lyons MJ, Tsuang MT: Long-term reliability and validity of alcoholism diagnoses and symptoms in a large national telephone interview survey. *Alcohol Clin Exp Res* 1998; 22:553–558
 49. Fenig S, Levav I, Kohn R, Yelin N: Telephone vs face-to-face interviewing in a community psychiatric survey. *Am J Public Health* 1993; 83:896–898
 50. Newman JC, Des Jarlais DC, Turner CF, Gribble J, Cooley P, Paone D: The differential effects of face-to-face and computer interview modes. *Am J Public Health* 2002; 92:294–297
 51. Monsees GM, Tamimi RM, Kraft P: Genome-wide association scans for secondary traits using case-control samples. *Genet Epidemiol* 2009; 33:717–728
 52. Talati A, Fyer AJ, Weissman MM: A comparison between screened NIMH and clinically interviewed control samples on neuroticism and extraversion. *Mol Psychiatry* 2008; 13:122–130