Predicting Loneliness with Polygenic Scores of Social, Psychological, and Psychiatric Traits

Abdel Abdellaoui1,2, Michel G. Nivard1, Jouke-Jan Hottenga1, Iryna Fedko1, Karin J.H. Verweij2, Bart M.L. Baselmans1,3, Erik A. Ehli4, Gareth E. Davies4, Meike Bartels1,3,5, Dorret I. Boomsma1,3,5*, John T. Cacioppo4,6*

1 Department of Biological Psychology, Vrije Universiteit, Amsterdam, the Netherlands
2 Department of Psychiatry, Academic Medical Centre, University of Amsterdam, Amsterdam, the Netherlands
3 Amsterdam Public Health Research Institute, Vrije Universiteit Amsterdam, the Netherlands
4 Avera Institute for Human Genetics, Sioux Falls, South Dakota, USA
5 Amsterdam Neuroscience, Amsterdam, The Netherlands
6 Department of Psychology, University of Chicago, Chicago, Illinois, USA
* These authors contributed equally

Key words: Loneliness, Polygenic Scores, Major Depressive Disorder, Genetic Prediction, Genetic Correlation

Correspondence:
Abdel Abdellaoui
Academic Medical Centre, University of Amsterdam

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/gbb.12472

This article is protected by copyright. All rights reserved.
Abstract

Loneliness is a heritable trait that accompanies multiple disorders. The association between loneliness and mental health indices may partly be due to inherited biological factors. We constructed polygenic scores for 27 traits related to behavior, cognition, and mental health and tested their prediction for self-reported loneliness in a population-based sample of 8,798 Dutch individuals. Polygenic scores for major depressive disorder, schizophrenia, and bipolar disorder were significantly associated with loneliness. Of the Big Five personality dimensions, polygenic scores for neuroticism and conscientiousness also significantly predicted loneliness, as did the polygenic scores for subjective well-being, tiredness, and self-rated health. When including all polygenic scores simultaneously into one model, only two major depression polygenic scores remained as significant predictors of loneliness. When controlling only for these two MDD polygenic scores, only neuroticism and schizophrenia remain significant. The total variation explained by all polygenic scores collectively was 1.7%. The association between the propensity to feel lonely and the susceptibility to psychiatric disorders thus pointed to a shared genetic etiology. The predictive power of polygenic scores will increase as the power of the genome-wide association studies on which they are based increases and may lead to clinically useful polygenic scores that can inform on the genetic predisposition to loneliness and mental health.
Introduction

Loneliness is an aversive state that people experience when there is a discrepancy between desired and actual social relationships. The physiological and psychological reactions to loneliness are mechanisms which are likely to have evolved to put the body in a heightened state of alertness in order to prompt us to improve our social circumstances (Cacioppo et al., 2014, Goossens et al., 2015). Loneliness is an unwanted state whereas solitude indicates a preference for being alone. The capacity to tolerate solitude may have potential evolutionary benefits (e.g., less stress due to dominance hierarchy, less depletion of resources, increased freedom to choose one’s own mental and physical activities (Long & Averill, 2003)), which may have contributed to the evolution of differential preference for solitude, and thereby individual differences in the susceptibility to loneliness. Like many other quantitative dimensions that show individual differences, falling in the extreme of the distribution is likely to coincide with mental or physical health problems (Plomin et al., 2009). Chronic loneliness is characterized by high negative affectivity and social withdrawal and often encompasses psychiatric conditions such as major depression or schizophrenia (Cacioppo & Cacioppo, 2014). When these strong aversive signals of perceived isolation remain present for prolonged periods, they can have detrimental consequences to overall health (Cacioppo et al., 2015a, Cacioppo et al., 2015b, Cole et al., 2007, Hawkley & Cacioppo, 2010, Heinrich & Gullone, 2006, Holt-Lunstad et al., 2015, Miller, 2011). A meta-analysis of ~3.4 million subjects from 70 independent studies found loneliness to increase the likelihood of death with 26%-32% within the
7 years that subjects were monitored, which is comparable to the impact of obesity and cigarette smoking (Holt-Lunstad et al., 2015).

Like all human behavioral traits (Turkheimer, 2000), individual differences in the propensity to feel lonely are partly genetic. Heritability estimates from twin and family studies range from 26% to 58% in children, while in adults the largest study (N=8,683) estimated genetic influences at 37% (Bartels et al., 2008, Boomsma et al., 2007, Boomsma et al., 2006, Boomsma et al., 2005, Distel et al., 2010, McGuire & Clifford, 2000). The rapid developments in human molecular genetics will likely result in an improvement of the predictive value of measured genetic variants for complex psychological and psychiatric traits. The prediction of genetic mental health risks can be particularly useful during earlier development. The ability to estimate one’s genetic predisposition for loneliness and related health risks in an early stage could lead to more effective deployment of environmental interventions, which can eventually translate to improved public health and well-being. Genome-wide association studies (GWASs) have only recently reached sufficiently large sample sizes to detect robust and replicable associations between genetic variants and the highly polygenic psychiatric and psychological traits. Significantly associated single nucleotide polymorphisms (SNPs) however, generally explain very little variance individually (<1%), and all genome-wide significant SNPs together usually explain no more than a few percent, although with increases in sample size of GWASs this is rapidly changing. Several methods that look at SNP-based heritabilities, estimated by software packages such GCTA-GREML (Yang et al., 2011), or LD Score regression (Bulik-Sullivan et al., 2015b), and the association of polygenic scores with phenotypes (Dudbridge, 2013)) explain phenotypic variation based on larger sets of SNPs than only those SNP that reach genome-wide significance. These approaches indicate that the ensemble of
non-significant SNPs contain a substantial amount of signal due to true polygenic effects on a trait, which means that effect size estimates of many non-significant SNPs may still have a predictive value. GCTA-GREML and LD score regression have shown that the aggregate of all measured SNPs explain about 14% of the individual differences in loneliness (Abdellaoui et al., 2017, Gao et al., 2016). The predictive value of polygenic scores based on individual level DNA data have potential clinical utility beyond that of methods that only estimate SNP-based heritability (e.g., GCTA-GREML (Yang et al., 2013) and LD score regression (Bulik-Sullivan et al., 2015b)), which is why it is important to demonstrate their predictive value as we do in our study. Powerful polygenic scores can potentially be used in the clinic as estimates of genetic risk, and have additional applications in a research context, such as the study of interactions between genetic risk and environmental exposures (Peyrot et al., 2014, Peyrot et al., 2017).

In this study, we use genome-wide SNP genotypes to compute polygenic scores for a range of traits related to loneliness and assess their predictive value for loneliness. Polygenic scores are indicators of the genetic predisposition of a certain trait and are computed by summing all individual alleles weighted by the estimated effect sizes for a specific trait. The predictive power of the polygenic scores is strongly related to the statistical power of the GWASs that produce the effect size estimates.

As loneliness is associated with a wide range of psychological, social, and psychiatric traits on which large GWASs have been conducted, we will construct polygenic scores with effect size estimates from a large collection of those GWASs. Loneliness has been associated with a higher prevalence of several psychiatric and neurological disorders (Booth, 2000, Mushtaq et al., 2014), of which we included the following: major depressive disorder (MDD), bipolar disorder,
schizophrenia, autism, anorexia, anxiety disorder, ADHD, Alzheimer’s disease, and migraine. Polygenic scores were also be computed for the Big Five personality dimensions, which have been associated with loneliness in several studies, with neuroticism and extraversion generally showing the strongest associations with loneliness (positive and negative associations respectively) (Abdellaoui et al., 2017, Atak, 2009, Cacioppo et al., 2006a, Teppers et al., 2013). From epidemiological studies we also know that lonely individuals have lower levels of well-being (Ben-Zur, 2012), more depressive symptoms (Cacioppo et al., 2010, Cacioppo et al., 2006b), more substance use (Åkerlind & Hörnquist, 1992, Kim, 1999), more fatigue (Jaremka et al., 2014, Jaremka et al., 2013), lower self-rated health (Nummela et al., 2011), tend to live in poorer neighborhoods (Scharf & de Jong Gierveld, 2008), and reach lower educational attainment (Bishop & Martin, 2007, Distel et al., 2010, Page & Cole, 1991, Steptoe et al., 2013). Some of these associations may be explained by related socio-economic factors (Hawkley et al., 2008) but in this paper we test the alternative explanation, of a shared genetic etiology.

The primary aim of this study was to identify which social, psychological, and psychiatric traits share their genetic etiology with loneliness to such an extent that they can be used to produce polygenic scores that have predictive value for loneliness in a population-based sample. Before we tested the associations of polygenic scores with loneliness, we established which PGS would attain sufficient statistical power. The statistical power depends on the genetic correlation between the traits and loneliness, and the accuracy of the estimated allelic effect sizes, which depends on the power of the GWASs.
Methods and Materials

Subjects and Phenotype Data

Data on loneliness were collected by the Netherlands Twin Register (NTR) in >30,000 twins and their family members across the Netherlands between 2004 and 2016, of which 8,798 adult subjects (3,206 males and 5,592 females; ages 18-91, mean age = 45.3, median age = 43) were genotyped (Boomsma et al., 2005, Willemsen et al., 2010, Willemsen et al., 2013). Loneliness was measured by the short scale for assessing loneliness in large epidemiological studies, developed by Hughes et al. (Hughes et al., 2004), and contains three items from the R-UCLA loneliness scale (Russel et al., 1980): 1) How often do you feel left out, 2) How often do you feel isolated from others, and 3) How often do you feel that you lack companionship. Response categories are: 1) hardly ever, 2) some of the time, and 3) often. The three responses were summed to obtain the loneliness score, with higher scores indicating more loneliness. Because of a skewed distribution, the loneliness score was log-transformed for all analyses. This study was approved by the Central Ethics Committee on Research Involving Human Subjects of the VU University Medical Centre, Amsterdam, an Institutional Review Board certified by the US Office of Human Research Protections (IRB number IRB-2991 under Federal-wide Assurance-3703; IRB/institute codes, NTR 03-180). All subjects provided written informed consent.

Genotyping, Quality Control (QC), Imputation, and PCA
Genotyping was done on several genome-wide SNP micro-arrays (Lin et al., 2016). Genotyped data were cross-platform imputed using the Genome of the Netherlands (GoNL) (Boomsma et al., 2014, Francioli et al., 2014) as a reference set to infer the SNPs missing per platform in the combined data (Fedko et al., 2015). For pre-imputation QC we excluded alleles with reference set allele frequency differences of >10%, SNPs with MAF < .005, deviation from Hardy-Weinberg Equilibrium (HWE) with $p<10^{-12}$, and a genotyping call rate <.95. We excluded samples with a genotyping call rate <.90, inbreeding coefficient from PLINK (Purcell et al., 2007) (F) <-.075 or >.075, Affymetrix Contrast QC metric <.40, the Mendelian error rate >5 standard deviations (SDs) from the mean, or when the gender or Identity-by-State (IBS) status did not agree with known relationship status and genotypic assessment. Phasing and imputation was performed with MaCH-Admix (Liu et al., 2013) software. After imputation, SNPs that were significantly associated with genotyping platform ($p<10^{-5}$), that had an allele frequency difference of >10% with GoNL reference set, HWE $p<10^{-5}$, Mendelian error rate >5 SD from mean over all markers, or an imputation quality $R^2$<.90 were excluded. We then performed a Principal Components Analysis (PCA) to exclude individuals with a non-Dutch ancestry and control for Dutch population stratification following procedures described in Abdellaoui et al (2013). All SNPs that survived QC (N=1,224,793) were used to construct polygenic scores.

**Power Analysis**

GWAS summary statistics were available for 31 complex traits related to loneliness. We first investigated which of them would have sufficient power to predict loneliness assuming a genetic
correlation of .8, using the power calculation method devised by Dudbridge (Dudbridge, 2013). The power was computed as a function of seven parameters in the case of continuous traits, and nine in the case of binary traits: 1) the significance threshold (set at a Bonferroni corrected alpha of .05/24 = .0022; where 24 is the number of independent polygenic scores, derived by a principal component analysis [PCA] on the polygenic scores: the number of independent polygenic scores was set at the number of principal components [PCs] that explain >95% of the variance), 2) the genetic correlation between the trait and loneliness, 3) the sample size of the GWAS, 4) the sample size of the target sample (N=8,798), 5) the SNP-based heritability, which were based on LD score regression (Bulik-Sullivan et al., 2015b) estimates (with 14% for loneliness based on the Gao et al GWAS (Gao et al., 2016), since our own sample was underpowered for LD score regression and gave heritability estimates of 0%; the heritability of the other traits are depicted in Table 1), 6) the number of independent SNPs in the target sample (148,681 SNPs, computed by pruning for LD in PLINK (Purcell et al., 2007)), 7) the assumed fraction of causal markers (which we set to .3 based on estimates of previous studies on cognitive and psychiatric traits (Hugh-Jones et al., 2016, Vilhjálmsson et al., 2015)), and in the case of binary traits, also 8) the trait prevalence in the general population, and 9) case/control sampling fraction in the GWAS study. We only included traits in subsequent analyses that reached statistical power of >50% assuming a genetic correlation of .8. For the traits that reached sufficient power, we re-computed the power four times with an adjusted Bonferroni significance threshold (.05 / number of independent polygenic scores computed with a PCA as the number of PCs that explain >95% of the variance) assuming genetic correlations of .2, .4, .6, and .8 (Table 1). Using the power calculation method and R-code developed by Dudbridge
(2013), we built a web-version of the power-calculator for general use that can be found at https://eagenetics.shinyapps.io/power_website/.

**Polygenic scores**

Polygenic scores were created with the estimated effect sizes from recent large GWASs (see references in the first column of Table 1). If NTR studies were part of the meta-analysis, the summary statistics were re-computed excluding NTR subjects in order to avoid an over-estimation of the association between the PRSs and loneliness (Wray et al., 2013). The polygenic scores were computed using LDpred (Vilhjálmsdóttir et al., 2015), which models linkage disequilibrium (LD) using the LD structure of a reference sample (all five European populations from the 1000 Genomes dataset in our case: Utah Residents (CEPH) with Northern and Western European Ancestry, Finnish, British, Iberian, and Toscani individuals, N=381). Vilhjálmsdóttir et al. (Vilhjálmssson et al., 2015) showed with simulations and empirically that this method outperforms traditional approaches (Vilhjálmsdóttir et al., 2015). This method needs the assumed fraction of causal markers as an input parameter, which we set at .3, based on estimates of previous studies on cognitive and psychiatric traits (Hugh-Jones et al., 2016, Vilhjálmsdóttir et al., 2015). Association analyses were carried out using generalized estimation equations (GEE) in SPSS 22.0. An exchangeable conditional covariance matrix was used to account for the relatedness among subjects (i.e., we allowed for correlated residuals between members of the same family) and tests were based on the robust (sandwich-corrected) standard errors (Minică et al., 2015). The first ten genomic PCs, age, and sex were
included in the model as fixed effects. The association analyses were first conducted for each polygenic score separately, then with all polygenic scores simultaneously in one model.
Results

We first investigated the statistical power of the polygenic score prediction given a genetic correlation of .8 with loneliness. There were four polygenic scores with less than 50% power to detect an association, namely: loneliness based on the GWAS in the Health and Retirement Study (Gao et al., 2016) (continuous and categorical; power: 14% and 21% respectively), ADHD (Middeldorp et al., 2016) (power = 26%), and melancholic MDD (Cai et al., 2015) (power = 47%). These traits were excluded from subsequent analyses, and the power was recomputed for the 27 remaining traits assuming genetic correlations of .2, .4, .6, and .8 (Table 1).

The GEE association analyses were corrected for sex, age, and the first ten genetic PCs. When including only the covariates, only sex and age were significantly associated with loneliness, with women and younger individuals reporting higher levels of loneliness (sex: standardized B = -.124, p = 8.3 × 10^-9, age: standardized B = -.071, p = 2.5 × 10^-10). Out of the 27 polygenic scores that survived the power analyses, twelve were significantly associated with loneliness when tested separately, with standardized B’s varying between .04 and .08: MDD (two distinct studies, one based on self-report in a web-based survey (Hyde et al., 2016), and one on structured clinical interviews, clinician-administered checklists, hospital/medical records, or self-report (Wray et al., 2017)), neuroticism (two distinct studies, one based on a web-based implementation of the Big Five Inventory (BFI) (Lo et al., 2016), and one on a combination of the NEO Personality Inventory, Eysenck Personality Questionnaire, and the International Personality and Item Pool Inventory (Okbay et al., 2016a)), schizophrenia, subjective well-being, the genes shared between schizophrenia and bipolar disorder, depressive symptoms, tiredness, bipolar disorder,
conscientiousness, and self-rated health ($p < .002$, see Figure 1). Polygenic scores for extraversion, anorexia, openness, and autism reached nominal significance ($p < .05$). Twelve polygenic scores did not reach significance, despite sufficient power to detect associations with traits with medium to high genetic correlations. All psychiatric disorders were positively associated with loneliness, i.e., a higher genetic risk for psychiatric disease was associated with increased loneliness. A higher genetic predisposition for neuroticism, openness, tiredness, or self-rated health was associated with higher levels of loneliness. Negative associations were observed for well-being, conscientiousness, and extraversion.

As expected from the genetic correlations among psychological and psychiatric traits, the polygenic scores show considerable correlations with each other (see correlation matrix in Figure 2). We analyzed all 27 polygenic scores simultaneously in one model in order to assess their independent contributions: only the two MDD scores reached significance after multiple testing correction (Figure 1). In order to test whether these two MDD scores were responsible for associations with the other traits, we repeated the analyses for all other scores while correcting for the two MDD scores by including them as covariates in the GEE model. Only neuroticism and schizophrenia remained significant after controlling for the two MDD scores (Figure 1).

The polygenic scores collectively explain 1.7% of the variance, while the two MDD scores together explain 1.1%.
Discussion

As genomic studies are advancing, it is becoming evident that much, if not most, of human phenotypic variation is influenced by genetic variants of pleiotropic nature (Gratten & Visscher, 2016). Psychiatric disorders show a substantial genetic overlap with each other and with non-psychiatric cognitive and behavioral traits (Bulik-Sullivan et al., 2015a). In this study, we investigated the predictive power of polygenic scores of a large collection of personality, cognition, and physical and mental health related traits on feeling lonely. We constructed the polygenic scores from genome-wide SNP data in a Dutch population-based cohort using summary statistics from a wide range of GWASs. Significant predictive power was observed for polygenic scores of about half of the traits tested. When including all polygenic scores simultaneously in one model, only two polygenic scores remained significantly associated with loneliness after multiple testing correction, both for indices of MDD. The independent contributions of the two MDD scores, which only capture a relatively small part of MDD heritability, imply that they contain unique information about genetic risk for MDD and its overlap with loneliness (which is also reflected in the relatively low correlation between the two MDD scores of .28 in Figure 2). One MDD index was based on a recent large GWAS conducted by 23andMe (Hyde et al., 2016), where clinical diagnoses of depression were identified through self-report in web-based surveys. The second score was based on a GWAS from the Psychiatric Genomics Consortium (Wray et al., 2017), where cases were identified through structured diagnostic instruments from direct interviews by trained interviewers, clinician-administered DSM-IV checklists, hospital/medical records, or self-reported clinical diagnoses. It is not clear whether the difference between the unique genetic risks captured by the two scores are due to a difference in the phenotype or sample ascertainment or due to chance fluctuations in effect.
size estimates of individual SNPs between studies. MDD has a substantial genetic overlap with all traits that show an association with loneliness in our study (Bulik-Sullivan et al., 2015a, Deary et al., 2017, Harris et al., 2016, Hyde et al., 2016, Lee et al., 2013b, Lo et al., 2016, Okbay et al., 2016a, Smoller et al., 2013), and the genetic factors responsible for that overlap partly explain their association with loneliness. In other words, genes that cause a relationship between loneliness and personality, well-being, schizophrenia, bipolar disorder, tiredness, or self-rated health include genes that are involved in depression as well. The MDD scores did not fully explain the associations with all other traits: neuroticism, schizophrenia, and schizophrenia & bipolar disorder remained significant after correcting for only the two MDD scores.

Loneliness and depression are both aversive and unpleasant states, but there is much evidence showing that they are statistically and conceptually different constructs (Cacioppo et al., 2010, Cacioppo et al., 2006b, Weeks et al., 1980). Loneliness has been characterized as negative feelings about one’s perceived inadequacy of social connections, while depression entails negative feelings in general (Weiss, 1973). Highly prevalent behavioral and mental states such as loneliness and depression may have been beneficial adaptations of our ancestors, because they may have increased their chances of survival and reproduction under certain conditions. It remains challenging to identify such adaptive roots with much certainty (Smith, 2016). The prevailing hypothesis on the evolutionary benefits of loneliness is based on humans having adapted to live in groups; for social creatures, loneliness may have increased the chances for survival and reproduction through the motivation to improve one’s social circumstances (Cacioppo et al., 2014). A related evolutionary mechanism that has been widely proposed for depression is the social risk hypothesis, in which depression causes one to minimize the risk of social exclusion due to an
imbalance of one’s social value and social burden (Allen & Badcock, 2003, Badcock et al., 2017). The analytical rumination hypothesis (Andrews & Thomson Jr, 2009) for depression is also in line with existing evolutionary explanations of loneliness, in which depression-related problems (in the case of loneliness: perceived social isolation) are given priority access to the limited processing resources of the brain by decreasing positive affect and desire for other activities. The analytical rumination hypothesis would explain why loneliness increases depressive symptoms, while depression influences loneliness in much lesser extent (Cacioppo et al., 2010, Cacioppo et al., 2006b), and why sensitivity for negative social cues is increased in lonely people (Cacioppo & Hawkley, 2009, Cacioppo et al., 2009, Cacioppo et al., 2015b, Cacioppo et al., 2015c, Duck et al., 1994). This increase of sensitivity of negative social cues is in line with the evidence for the existence of different subtypes of low mood to cope with different kinds of fitness-relevant situations (Keller & Nesse, 2005). In this context, loneliness could be interpreted as the employment of specific biological “depression” faculties to cope with or warn against a specific fitness-threatening situation (social isolation), just as there are specific types of pain to warn against different types of physical injury. Such a relationship would likely result in the significant genetic overlap between loneliness and depression that we observe in this study.

After depression, neuroticism showed the strongest association with loneliness. Cacioppo et al. (Cacioppo et al., 2010) showed that the relationship between loneliness and depression is independent from the relationship with sensitivity for negative emotions, i.e., neuroticism; our data showed that this independent relationship is measurable at the DNA level. An essential difference between neuroticism and depression as well as loneliness is that neuroticism is a relatively permanent individual characteristic, while depression and loneliness usually reflect a temporary
change in one’s state. Neuroticism nevertheless shows a strong genetic correlation with depression (.75) (Okbay et al., 2016a). We showed in an earlier molecular genetic study that loneliness and neuroticism also have a considerable shared genetic etiology with an estimated genetic correlation between .7 and .8 (Abdellaoui et al., 2017). When analyzing the polygenic scores separately, we see a strong negative association with the polygenic score for subjective well-being which, despite relatively low power, suggests a genetic correlation with subjective well-being in the same range (or higher) as the genetic correlation with neuroticism and depression. This genetic correlation is likely due to genes that subjective well-being shares with neuroticism and/or depression, which have genetic correlations with well-being of approximately -.8 (Okbay et al., 2016a). The genetic influences these traits share are most likely those that underlie processes related to mood and/or the sensitivity to negative emotions.

After depression and neuroticism, schizophrenia showed the strongest association with loneliness. The association with schizophrenia was due to the genetic component that schizophrenia shares with bipolar disorder, which also remains significant after accounting for the two MDD scores. The genetic component that differentiates between schizophrenia and bipolar disorder was not associated with loneliness, despite a relatively powerful polygenic score (standardized B = .005, p = .681). A meta-analysis on the effectiveness of interventions for loneliness proved addressing maladaptive social cognition to be the most effective intervention, as compared to interventions directed at improving social skills, enhancing social support, or increasing opportunities for social contact (Masi et al., 2010). Impaired cognition is a core feature of both schizophrenia and bipolar disorder, with schizophrenia showing more impaired social than non-social cognition compared to bipolar disorder (Lee et al., 2013a), which may explain the
stronger association between loneliness and schizophrenia. As the predictive power of molecular genetic data increases, polygenic scores based on GWASs that capture these features may aid in further narrowing down which individuals would benefit most from interventions that target social cognition.

Gao et al (Gao et al., 2016) investigated the association between loneliness in ~7,000 unrelated individuals from the Health and Retirement Study (HRS) and polygenic scores for six traits: neuroticism, extraversion, schizophrenia, bipolar disorder, major depressive disorder, and depressive symptoms. They found a significant association only with neuroticism and depressive symptoms. The reason that our strongest predictors, MDD, did not reach significance in their study was likely due to the differences in the sample sizes of the MDD GWASs: 18,759 in Gao et al (2016) and over 160,000 in our study. The power to detect associations with additional traits that did not reach significance in their analyses (schizophrenia and bipolar disorder) may be due to our larger sample size and the use of LDpred to construct polygenic scores, which increases power by optimizing the effect size estimates of individual SNPs by incorporating the genome-wide LD structure (Vilhjálmsson et al., 2015).

Polygenic scores tend to explain a few percent of the variance for most human behavioral traits, making the 1.7% explained variance for loneliness relatively promising, especially considering the plateau has not yet been reached, and considering that age and gender combined explained only 0.9% of the individual differences. When effect sizes of genome-wide SNPs are mapped accurately enough, polygenic scores can potentially reach a stronger predictive power than that of family history. While perfect prediction will never be reached, more powerful genetic studies may lead to clinically useful polygenic scores and could be employed to help combat pre-
clinical mental health outcomes such as loneliness. Our results show the presence of a shared genetic etiology between the propensity to feel lonely and traits related to personality, mood, negative affect, well-being, somatic health, and susceptibility to psychiatric disorders.
**Acknowledgements**

We would like to thank all the twins and family members. This project was made possible by the National Institutes of Health (NIH, R37 AG033590-08 to JC). MGN is supported by Royal Netherlands Academy of Science Professor Award (PAH/6635 to DIB). Data collection was supported by the Netherlands Organization for Scientific Research (NWO: MagW/ZonMW grants 904-61-090, 985-10-002, 904-61-193, 480-04-004, 400-05-717, NWO-bilateral agreement 463-06-001, NWO-VENI 451-04-034, Addiction-31160008, Middelgroot-911-09-032, Spinozapremie 56-464-14192), Biobanking and Biomolecular Resources Research Infrastructure (BBMRI –NL, 184.021.007). Genotyping was also supported by the Avera Institute for Human Genetics, Sioux Falls, South Dakota (USA) and the National Institutes of Health (NIH, R01D0042157-01A), the Genetic Association Information Network (GAIN) of the Foundation for the US National Institutes of Health (NIMH, MH081802) and by the Grand Opportunity grants 1RC2MH089951-01 and 1RC2 MH089995-01 from the NIMH. Part of the analyses were carried out on the Genetic Cluster Computer (http://www.geneticcluster.org), which is financially supported by the Netherlands Scientific Organization (NWO 480-05-003), the Dutch Brain Foundation, and the department of Psychology and Education of the VU University Amsterdam.

**Financial Disclosures:**

The authors declare no conflict of interest.
References


Figure 1: Results of the generalized estimation equations (GEE) association analyses between loneliness and polygenic scores (ordered on effect size of the analyses of individual scores); N=8,798. Bonferroni corrected α: .05 / 24 independent tests = .002, where the independence was determined by PCA.
Figure 2: Partial correlations between 27 polygenic scores, adjusted for sex, age, and the first ten genetic principal components. The size of circles corresponds to the strength of the correlation.
Table 1: Sample sizes (N-GWAS), heritability estimates ($h^2$ LDSC) of the GWAS summary statistics from LD score regression, and the power to detect an association between loneliness and the polygenic scores given a genetic correlation ($r_g$) of .2, .4, .6, and .8.

<table>
<thead>
<tr>
<th>Polygenic Score</th>
<th>N-GWAS</th>
<th>$h^2$ LDSC</th>
<th>Power if $r_g = .2$</th>
<th>Power if $r_g = .4$</th>
<th>Power if $r_g = .6$</th>
<th>Power if $r_g = .8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MDD (23andMe)</td>
<td>297 102</td>
<td>.07</td>
<td>.08</td>
<td>.59</td>
<td>.97</td>
<td>1</td>
</tr>
<tr>
<td>MDD (MDD2-PGC)</td>
<td>160 125</td>
<td>.11</td>
<td>.11</td>
<td>.75</td>
<td>.99</td>
<td>1</td>
</tr>
<tr>
<td>Neuroticism</td>
<td>59 206</td>
<td>.11</td>
<td>.05</td>
<td>.41</td>
<td>.89</td>
<td>1</td>
</tr>
<tr>
<td>Schizophrenia</td>
<td>79 845</td>
<td>.45</td>
<td>.84</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Neuroticism</td>
<td>164 489</td>
<td>.09</td>
<td>.16</td>
<td>.86</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Subjective Well-being</td>
<td>298 420</td>
<td>.03</td>
<td>.07</td>
<td>.56</td>
<td>.96</td>
<td>1</td>
</tr>
<tr>
<td>Bipolar &amp; Schizophrenia</td>
<td>39 202</td>
<td>.37</td>
<td>.27</td>
<td>.97</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Depressive symptoms</td>
<td>146 251</td>
<td>.05</td>
<td>.06</td>
<td>.46</td>
<td>.92</td>
<td>1</td>
</tr>
<tr>
<td>Tiredness</td>
<td>108 976</td>
<td>.07</td>
<td>.06</td>
<td>.49</td>
<td>.94</td>
<td>1</td>
</tr>
<tr>
<td>Bipolar Disorder</td>
<td>63 766</td>
<td>.43</td>
<td>.15</td>
<td>.85</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Conscientiousness (23andMe)</td>
<td>59 176</td>
<td>.09</td>
<td>.04</td>
<td>.31</td>
<td>.79</td>
<td>.98</td>
</tr>
<tr>
<td>Self-Rated Health</td>
<td>111 749</td>
<td>.10</td>
<td>.11</td>
<td>.73</td>
<td>.99</td>
<td>1</td>
</tr>
<tr>
<td>Extraversion</td>
<td>59 225</td>
<td>.18</td>
<td>.10</td>
<td>.71</td>
<td>.99</td>
<td>1</td>
</tr>
<tr>
<td>Anorexia</td>
<td>17 767</td>
<td>.56</td>
<td>.22</td>
<td>.94</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Openness (23andMe)</td>
<td>59 176</td>
<td>.10</td>
<td>.04</td>
<td>.36</td>
<td>.85</td>
<td>.99</td>
</tr>
<tr>
<td>Autism</td>
<td>10 263</td>
<td>.46</td>
<td>.07</td>
<td>.53</td>
<td>.95</td>
<td>1</td>
</tr>
<tr>
<td>Social Deprivation Index</td>
<td>112 151</td>
<td>.04</td>
<td>.03</td>
<td>.24</td>
<td>.68</td>
<td>.95</td>
</tr>
<tr>
<td>Migraine</td>
<td>196 685</td>
<td>.04</td>
<td>.02</td>
<td>.16</td>
<td>.51</td>
<td>.86</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>59 173</td>
<td>.08</td>
<td>.03</td>
<td>.26</td>
<td>.72</td>
<td>.97</td>
</tr>
<tr>
<td>Alzheimer's disease</td>
<td>54 162</td>
<td>.07</td>
<td>.01</td>
<td>.10</td>
<td>.34</td>
<td>.69</td>
</tr>
<tr>
<td>Educational Attainment</td>
<td>321 852</td>
<td>.12</td>
<td>.53</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Extraversion (GPC)</td>
<td>56 614</td>
<td>.06</td>
<td>.02</td>
<td>.16</td>
<td>.52</td>
<td>.86</td>
</tr>
<tr>
<td>Bipolar vs Schizophrenia</td>
<td>16 381</td>
<td>.33</td>
<td>.06</td>
<td>.49</td>
<td>.93</td>
<td>1</td>
</tr>
<tr>
<td>Alcohol use</td>
<td>66 700</td>
<td>.05</td>
<td>.02</td>
<td>.15</td>
<td>.50</td>
<td>.85</td>
</tr>
<tr>
<td>Smoking: ever vs never</td>
<td>69 207</td>
<td>.08</td>
<td>.03</td>
<td>.21</td>
<td>.63</td>
<td>.93</td>
</tr>
<tr>
<td>Smoking: cigs per day</td>
<td>35 173</td>
<td>.06</td>
<td>.01</td>
<td>.08</td>
<td>.27</td>
<td>.59</td>
</tr>
<tr>
<td>Income</td>
<td>112 151</td>
<td>.06</td>
<td>.05</td>
<td>.42</td>
<td>.89</td>
<td>1</td>
</tr>
</tbody>
</table>