Published in final edited form as:
Drug Alcohol Depend. 2017 February 01; 171: 117–121. doi:10.1016/j.drugalcdep.2016.09.022.

Short communication: Genetic association between schizophrenia and cannabis use

Karin J.H. Verweij[1,2,3], **Abdel Abdellaoui**[1], **Michel G. Nivard**[1], **Alberto Sainz Cort**[1], **Lannie Ligthart**[1], **Harmen H.M. Draisma**[1,2], **Camelia C. Minică**[1], **International Cannabis Consortium**, **Nathan A. Gillespie**[4], **Gonneke Willemsen**[1], **Jouke-Jan Hottenga**[1], **Dorret I. Boomsma**[1], and **Jacqueline M. Vink**[1,3]

[1]Department of Biological Psychology/Netherlands Twin Register, VU University, van der Boechorststraat 1, 1081 BT Amsterdam, The Netherlands [2]Neuroscience Campus Amsterdam, De Boelelaan 1085, 1081 HV Amsterdam, The Netherlands [3]Behavioural Science Institute, Radboud University, Montessorilaan 3, 6525 HR Nijmegen, The Netherlands [4]Virginia Institute for Psychiatric and Behavior Genetics, Virginia Commonwealth University, 800 E Leigh St, Richmond, Virginia 23219, USA

Abstract

Background and aim—Previous studies have shown a relationship between schizophrenia and cannabis use. As both traits are substantially heritable, a shared genetic liability could explain the association. We use two recently developed genomics methods to investigate the genetic overlap between schizophrenia and cannabis use.

Methods—Firstly, polygenic risk scores for schizophrenia were created based on summary statistics from the largest schizophrenia genome-wide association (GWA) meta-analysis to date. We analysed the association between these schizophrenia polygenic scores and multiple cannabis use phenotypes (lifetime use, regular use, age at initiation, and quantity and frequency of use) in a sample of 6,931 individuals. Secondly, we applied LD-score regression to the GWA summary statistics of schizophrenia and lifetime cannabis use to calculate the genome-wide genetic correlation.

Results—Polygenic risk scores for schizophrenia were significantly ($\alpha<0.05$) associated with five of the eight cannabis use phenotypes, including lifetime use, regular use, and quantity of use, with risk scores explaining up to 0.5% of the variance. Associations were not significant for age at initiation of use and two measures of frequency of use analyzed in lifetime users only, potentially because of reduced power due to a smaller sample size. The LD-score regression revealed a significant genetic correlation of $r_g=0.22$ (SE=0.07, p=0.003) between schizophrenia and lifetime cannabis use.

Corresponding Author: Karin J.H. Verweij, VU University, Department of Biological Psychology, Van der Boechorststraat 1, 1081 BT Amsterdam, The Netherlands, Phone: +31205985747, karin.verweij@vu.nl.

Declarations of competing interest: The authors have nothing to declare

Europe PMC Funders Author Manuscripts

Conclusions—Common genetic variants underlying schizophrenia and lifetime cannabis use are partly overlapping. Individuals with a stronger genetic predisposition to schizophrenia are more likely to initiate cannabis use, use cannabis more regularly, and consume more cannabis over their lifetime.

Keywords

cannabis; schizophrenia; polygenic risk; genetic correlation; quantitative genetics

1 Introduction

Numerous studies have observed an association between schizophrenia and cannabis use (McGrath et al., 2010; Moore et al., 2007; Morrison et al., 2009; van Os et al., 2002), although the direction of causation is still under debate. As both traits are substantially heritable, with heritability estimates of around 45% for lifetime cannabis use (Verweij et al., 2010) and 80% for schizophrenia (Sullivan et al., 2003), a shared genetic liability could explain the relationship. With methodological advances in molecular genetics and increased sample sizes in genome-wide association (GWA) studies it has become viable to use measured genetic variation among individuals to examine this relationship. Power et al. (2014) used results from the 2013 schizophrenia GWA study that included 13,833 schizophrenia cases and 18,310 controls (Ripke et al., 2013) to predict cannabis use in a target sample of 2082 Australian individuals. They found that the genetic variants underlying schizophrenia significantly predicted lifetime cannabis use, and quantity of use within users (maximum variance explained of R^2=0.47% and 0.85%, respectively), but not age at initiation of cannabis use.

Here, we used two recently developed methods to further investigate the genetic covariation between both traits. Firstly we performed a polygenic risk analysis to determine the extent to which common genetic variants that affect the risk of schizophrenia predict various measures of cannabis use. Based on the results from the latest and largest schizophrenia GWA study (with up to 36,989 schizophrenia cases and 113,075 controls; Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014) we generated genetic risk scores for schizophrenia in a large independent target sample from the Netherlands. We used LDpred to create the risk score, a method shown to have greater prediction accuracy than the conventional risk prediction approach involving linkage disequilibrium (LD) pruning followed by P-value thresholding (Vilhjalmsson et al., 2015). We then determined the association of the genetic risk score for schizophrenia with various cannabis use phenotypes, including lifetime (ever versus never) use, regular use, age at initiation of use, and quantity and frequency of use.

Secondly, we applied LD-score regression to the summary statistics of the largest GWA meta-analyses of schizophrenia (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014) and lifetime cannabis use (Stringer et al., in press) to estimate the genome-wide genetic correlation. To our knowledge, this is the first study to determine the strength of the genetic correlation between schizophrenia and cannabis use; the information can provide useful etiological insights into the relationships between the two traits.

Europe PMC Funders Author Manuscripts

2 Methods

2.1 Polygenic risk prediction analysis

The target sample comprised 6,931 participants (64.1% females) from 3,244 twin families registered at the Netherlands Twin Registry (Boomsma et al., 2006). Participants were between 18 and 94 years old (M=43.0, SD=15.7). Data on cannabis use were obtained from five waves of self-report questionnaires sent out between 1993 and 2013 (see Willemsen et al., 2013). For subjects who participated in more than one wave we used data from their last questionnaire.

We analysed various cannabis use variables:

1) Lifetime cannabis use was analysed as a dichotomous variable with never users marked as 0 and ever users as 1.

2) Lifetime regular cannabis use (2a) was analysed as a dichotomous variable. Individuals were asked whether they had ever used cannabis regularly during their lifetime. Those who had used regularly were marked as 1 and those that had never used regularly as 0. We also repeated this analysis excluding experimental users, i.e. those that had used regularly were marked as 1 and those that had never used as 0 (2b).

3) Age at onset of cannabis use in years.

4) Lifetime frequency of use was analysed as a continuous variable (4a). Participants were asked about their lifetime frequency of cannabis use. Response categories were: 0 = never; 1 = 1-2 times; 2 = 3-5 times; 3 = 6-10 times; 4 = 11-19 times; 6 = 40 times or more. We also repeated this analysis in lifetime users only (4b).

5) Frequency of cannabis use was analysed as a continuous variable (5a). Participants were asked about their frequency of cannabis use during the period of heaviest use. Response categories were: 0 = never used; 1 = once a month or less often; 2 = 2-4 times a month; 3 = 2-3 times per week; 4 = 4-5 times a week; 6 = six days a week or daily. We also repeated this analysis in lifetime users only (5b).

Details of the variables, descriptives statistics and sample sizes are provided in Table 1.

Genotyping was performed across five platforms and standard quality control checks were performed prior to imputation. Genotype data were imputed using the 1000 Genomes phase1 release v3 reference set. Stringent post-imputation quality thresholds were used (Nivard et al., 2014). Only single nucleotide polymorphisms (SNPs) with an imputation quality score above 0.95 were retained; SNPs were removed if they had a Minor Allele Frequency (MAF) <0.05 or deviated from Hardy-Weinberg Equilibrium (HWE) with $p<0.001$. Individuals were excluded if their genotype missing rate was >10%, if they had excess genome-wide homozygosity, or if they were of non-Dutch ancestry (see Abdellaoui et al., 2013). After quality control 3,622,018 SNPs remained.

Europe PMC Funders Author Manuscripts

Polygenic risk scores were created using LDpred (Vilhjalmsson et al., 2015), which takes into account LD among the SNPs in creating the polygenic risk scores. Risk scores in the target sample were generated by calculating the mean causal effect size of each marker using the SNP effect sizes from the latest schizophrenia GWA meta-analysis results (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014) and the LD structure from the European populations in the 1000 Genomes reference set. We created the schizophrenia polygenic scores with the expected fraction of causal genetic variants (the fraction of markers with non-zero effects) set at 30%, because in a previous paper the risk score based on this fraction was shown to optimally predict schizophrenia (see Vilhjalmsson et al., 2015).

Generalized estimating equation (GEE) modelling was applied to test whether the polygenic risk scores for schizophrenia predicted cannabis use phenotypes in the target cohort. A covariance matrix was used to account for family relatedness and tests were based on robust (sandwich-corrected) standard error (see Minica et al., 2014). Age, sex, birth cohort (three dummy variables representing four 20-year time intervals), and ten genetic PCs (three ancestry-informative PCs and seven PCs accounting for genotyping batch effects) were included as covariates in the model. Analyses also accounted for two environmental factors (level of urbanisation and socio-economic status) that have previously been found to be associated with schizophrenia and cannabis use (Daniel et al., 2009; Kuepper et al., 2011; Martino et al., 2008; Vassos et al., 2012; Werner et al., 2007). Based on postal code, we determined the level of urbanisation of participants' residence (ranging from 1 (very high, >2500 addresses per km^2) to 5 (very low, <500 addresses per km^2) and socio-economic status as measured by the average income of their residential area. Variance explained by the polygenic risk scores was calculated in regression analyses as the R^2 (or Nagelkerke's pseudo-R^2 for dichotomous variables) of the model including polygenic risk scores and covariates minus the R^2 of the model including only covariates.

2.2 *LD*-score regression analysis

With LD-Score regression (Bulik-Sullivan et al., 2015b) the variance in a trait that can be explained by all SNPs based on GWA summary statistics is estimated. The method is based on the fact that an estimated SNP effect-size incorporates effects of all SNPs in LD with that SNP. For polygenic traits, SNPs with high LD have on average higher χ^2 statistics than SNPs with low LD (the more genetic variation a SNP tags, the higher the probability that the SNP will tag a causal variant). When regressing the χ^2 statistics from a GWA study against the LD score for each SNP, the slope of the resulting regression line provides an estimate of the proportion of trait variance accounted for by all genotyped SNPs (Bulik-Sullivan et al., 2015b). Cross-trait LD-Score regression is an extension to estimate the genetic covariation between traits using GWA summary statistics of multiple traits (Bulik-Sullivan et al., 2015c). The genetic covariance is estimated using the slope from the regression of the product of z-scores from two GWA studies on the LD score. The estimate obtained from this method represents the genetic correlation between the two traits based on all polygenic effects captured by SNPs.

Here, we estimated the genetic correlation between schizophrenia and lifetime cannabis use based on the summary statistics from the largest GWA meta-analyses for schizophrenia (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014) and cannabis use (International Cannabis Consortium, Stringer et al., in press) to date. From the cannabis use GWA study we included only SNPs present in all contributing cohorts were included. Furthermore, we only included HapMap-3 SNPs (as suggested by Bulik-Sullivan et al., 2015b), resulting in 1,125,108 and 433,918 SNPs for schizophrenia and cannabis use, respectively. Standard LD scores were used as provided by Bulik-Sullivan and colleagues (Bulik-Sullivan et al., 2015b) based on the 1000 genomes reference set, restricted to European populations.

3 Results

Individuals from more urbanised residential areas were more likely to have used cannabis (Nagelkerke R^2=3.3%, p<0.001), whereas socio-economic status was not associated with lifetime cannabis use (p=0.35). Polygenic risk scores for schizophrenia were significantly (α<0.05) associated with five of the eight tested cannabis use phenotypes in the target sample, including lifetime use, regular use, and quantity of use, with risk scores explaining up to 0.5% of the variance (see Table 1). This indicates that individuals with a genetic predisposition to schizophrenia were significantly more likely to use cannabis. Associations were not significant for *Age at initiation of use, Lifetime frequency of cannabis use in users,* and *Frequency of use during period of most use in users*, potentially because of reduced power due to a smaller sample size for these variables.

The LD-score regression analysis revealed a significant genetic correlation of r_g=0.22 (SE=0.07, p=0.003) between schizophrenia and lifetime cannabis use, implying that part of the common genetic variants overlap between schizophrenia and lifetime cannabis use.

4 Discussion

Overall, results from our two analyses support the hypothesis that the association between schizophrenia and cannabis use is partly due to a shared genetic etiology. Polygenic risk scores based on the SNP effect sizes from a schizophrenia GWA meta-analysis significantly predicted most of the cannabis use variables in our target sample. Individuals with an increased genetic predisposition to schizophrenia were more likely to have initiated cannabis use, to have used cannabis regularly, and to have used more cannabis over their lifetime. These results are in line with the study by Power et al. (Power et al., 2014) who found an association between individuals' burden of schizophrenia risk alleles and lifetime cannabis use and quantity of use, but not age at initiation. The overall variance explained by the combined SNPs was low (up to 0.5%), but consistent with many previous cross-disorder predictions. This is not surprising because summing estimates of SNP effects also sums the error component of those estimates, and thus does not yield an accurate and unbiased estimate of the variance explained by the aggregate of all SNP effects. This method relies heavily on the accuracy of the SNP effect estimates in the discovery sample, which will increase when larger samples sizes become available.

More accurate predictions may also be obtained by better phenotyping; our measures of cannabis use were limited by the potential for inaccuracy and bias in retrospective self-reports. Additionally, the genetic predictions may be stronger for more extreme forms of cannabis use (e.g. problematic use).

With the LD-score regression analysis we quantified for the first time the genetic correlation between schizophrenia and cannabis use. The genetic correlation was estimated to be rather high, r_g=0.22, suggesting that a substantial part of the well-known association between schizophrenia and cannabis use can be explained by shared genetic influences (although we are unable to differentiate between pleiotropy and causality). The estimated genetic correlation is expectably weaker than the genetic correlation between schizophrenia and bipolar disorder or schizophrenia and depression, but is similar in magnitude to the genetic correlation of schizophrenia with ADHD and anorexia, and stronger than the genetic correlation of schizophrenia with all other diseases and traits investigated by Bulik-Sullivan et al. (2015a). Bioinformatics work and future studies with advanced technologies, novel statistical approaches (such as Mendelian randomisation), and larger sample sizes, should aim to determine the nature of the genetic association between schizophrenia and cannabis use and identify common genes and biological mechanisms that can explain the genetic association.

Acknowledgements

KJHV is supported by a 2014 NARSAD Young Investigator Grant from the Brain & Behavior Research Foundation. JMV is supported by ERC Starting Grant ‘Beyond the Genetics of Addiction’, grant number 284167 (PI: JMV). Funding for the Netherlands Twin Registry was obtained from the Netherlands Organization for Scientific Research (NWO) and MagW/ZonMW, BBMRI-NL (184.021.007), the VU University Institute for Health and Care Research (EMGO+) and Neuroscience Campus Amsterdam, the European Science Council (ERC) Genetics of Mental Illness (230374), Avera Institute for Human Genetics and US National Institute of Mental Health (1RC2MH089951-01 and 1RC2 MH089995).

References

Abdellaoui A, Hottenga JJ, de Knijff P, Nivard MG, Xiao X, Scheet P, Brooks A, Ehli EA, Hu Y, Davies GE, Hudziak JJ, et al. Population structure, migration, and diversifying selection in the Netherlands. European journal of human genetics : EJHG. 2013; 21:1277–1285. [PubMed: 23531865]

Boomsma DI, De Geus EJC, Vink JM, Stubbe JH, Distel MA, Hottenga J-J, Posthuma D, Van Beijsterveldt TCEM, Hudziak JJ, Bartels M. Netherlands Twin Register: from twins to twin families. Twin Research and Human Genetics. 2006; 9:849–857. [PubMed: 17254420]

Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR. An atlas of genetic correlations across human diseases and traits. Nature genetics. 2015a; 47:1236–1241. [PubMed: 26414676]

Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. 2015b; 47:291–295.

Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nature genetics. 2015c; 47:291–295. [PubMed: 25642630]

Daniel JZ, Hickman M, Macleod J, Wiles N, Lingford-Hughes A, Farrell M, Araya R, Skapinakis P, Haynes J, Lewis G. Is socioeconomic status in early life associated with drug use? A systematic review of the evidence. Drug and alcohol review. 2009; 28:142–153. [PubMed: 19320699]

Kuepper R, van Os J, Lieb R, Wittchen HU, Henquet C. Do cannabis and urbanicity co-participate in causing psychosis? Evidence from a 10-year follow-up cohort study. Psychol Med. 2011; 41:2121–2129. [PubMed: 21466750]

Martino SC, Ellickson PL, McCaffrey DF. Developmental trajectories of substance use from early to late adolescence: a comparison of rural and urban youth. Journal of studies on alcohol and drugs. 2008; 69:430–440. [PubMed: 18432386]

McGrath J, Welham J, Scott J, Varghese D, Degenhardt L, Hayatbakhsh MR, Alati R, Williams GM, Bor W, Najman JM. Association between cannabis use and psychosis-related outcomes using sibling pair analysis in a cohort of young adults. Archives of general psychiatry. 2010; 67:440–447. [PubMed: 20194820]

Minica CC, Boomsma DI, Vink JM, Dolan CV. MZ twin pairs or MZ singletons in population family-based GWAS? More power in pairs. Molecular psychiatry. 2014; 19:1154–1155. [PubMed: 25266125]

Moore TH, Zammit S, Lingford-Hughes A, Barnes TR, Jones PB, Burke M, Lewis G. Cannabis use and risk of psychotic or affective mental health outcomes: a systematic review. Lancet (London, England). 2007; 370:319–328.

Morrison PD, Zois V, McKeown DA, Lee TD, Holt DW, Powell JF, Kapur S, Murray RM. The acute effects of synthetic intravenous Delta9-tetrahydrocannabinol on psychosis, mood and cognitive functioning. Psychol Med. 2009; 39:1607–1616. [PubMed: 19335936]

Nivard MG, Mbarek H, Hottenga JJ, Smit JH, Jansen R, Penninx BW, Middeldorp CM, Boomsma DI. Further confirmation of the association between anxiety and CTNND2: replication in humans. Genes, brain, and behavior. 2014; 13:195–201.

Power RA, Verweij KJ, Zuhair M, Montgomery GW, Henders AK, Heath AC, Madden PA, Medland SE, Wray NR, Martin NG. Genetic predisposition to schizophrenia associated with increased use of cannabis. Molecular psychiatry. 2014

Ripke S, O'Dushlaine C, Chambert K, Moran JL, Kahler AK, Akterin S, Bergen SE, Collins AL, Crowley JJ, Fromer M, Kim Y, et al. Genome-wide association analysis identifies 13 new risk loci for schizophrenia. Nature genetics. 2013; 45:1150–1159. [PubMed: 23974872]

Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. Nature. 2014; 511:421–427. [PubMed: 25056061]

Stringer S, Minica CC, Verweij KJH, Mbarek H, Consortium I, Gillespie NA, Derks EM, Vink JM. Genome-wide association study of lifetime cannabis use based on a large meta-analytic sample of 32,330 subjects from the International Cannabis Consortium. Translational psychiatry. in press.

Sullivan PF, Kendler KS, Neale MC. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. Archives of general psychiatry. 2003; 60:1187–1192. [PubMed: 14662550]

van Os J, Bak M, Hanssen M, Bijl RV, de Graaf R, Verdoux H. Cannabis use and psychosis: a longitudinal population-based study. American journal of epidemiology. 2002; 156:319–327. [PubMed: 12181101]

Vassos E, Pedersen CB, Murray RM, Collier DA, Lewis CM. Meta-analysis of the association of urbanicity with schizophrenia. Schizophrenia bulletin. 2012; 38:1118–1123. [PubMed: 23015685]

Verweij KJH, Zietsch BP, Lynskey MT, Medland SE, Neale MC, Martin NG, Boomsma DI, Vink JM. Genetic and environmental influences on cannabis use initiation and problematic use: a meta-analysis of twin studies. Addiction. 2010; 105:417–430. [PubMed: 20402985]

Vilhjalmsson BJ, Yang J, Finucane HK, Gusev A, Lindstrom S, Ripke S, Genovese G, Loh PR, Bhatia G, Do R, Hayeck T, et al. Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. American journal of human genetics. 2015; 97:576–592. [PubMed: 26430803]

Werner S, Malaspina D, Rabinowitz J. Socioeconomic status at birth is associated with risk of schizophrenia: population-based multilevel study. Schizophrenia bulletin. 2007; 33:1373–1378. [PubMed: 17443013]

Willemsen G, Vink JM, Abdellaoui A, den Braber A, van Beek JHDA, Draisma HHM, van Dongen J, van't Ent D, Geels LM, van Lien R. The Adult Netherlands Twin Register: twenty-five years of survey and biological data collection. Twin Research and Human Genetics. 2013; 16:271–281. [PubMed: 23298648]

Europe PMC Funders Author Manuscripts

Drug Alcohol Depend. Author manuscript; available in PMC 2018 February 01.

Table 1

Genetic risk prediction analysis. Cannabis variables in the target sample, sample size, descriptive statistics, and results of the polygenic risk score analyses for schizophrenia predicting the various cannabis use phenotypes.

	Dichotomous variables	**N**	**Prevalence**	**Mean (SD)**	**B**	**R^2**	**P-value**	**P_{FDR}**
1	Lifetime cannabis use (Dichotomous, 0=never, 1=ever use)	6931	24.8%	--	0.101	0.24%	0.002	0.020
2a	Lifetime regular cannabis use (Dichotomous, 0=never used regularly, 1=used regularly)	6603	3.8%	--	0.189	0.43%	0.009	0.025
2b	Lifetime regular cannabis use versus never use (Dichotomous, 0=never used, 1=used regularly)	5457	4.5%	--	0.204	0.49%	0.007	0.025
	Continuous variables							
3	Age at initiation of cannabis use (in users only)	1400	--	18.8 (5.3)	-0.265	0.19%	0.069	0.079
4a	Lifetime frequency of cannabis use (in whole sample, never-users set at zero) [a]	5985	--	0.4 (1.1)	0.033	0.07%	0.035	0.056
4b	Lifetime frequency of cannabis use (in users only) [a]	776	--	2.8 (1.8)	0. 134	0.30%	0.052	0.070
5a	Frequency of cannabis use during period of heaviest use (in whole sample, never-users set at zero) [b]	5924	--	0.2 (0.6)	0.020	0.06%	0.026	0.052
5b	Frequency of cannabis use during period of heaviest use (in users only) [b]	715	--	1.5 (1.1)	0.057	0.11%	0.210	0.210

N=number of individuals in the target sample, B=effect size, R^2 = variance explained, for dichotomous traits R^2= Nagelkerke's pseudo R^2, P_{FDR} =P-value corrected for multiple testing by false discovery rate correction (FDR); note that this correction is relatively conservative given that most outcome variables are strongly interrelated