

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/96136/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Gage, Suzanne H, Jones, Hannah J, Burgess,, Stephen, Bowden, Jack, Davey Smith, George, Zammit, Stanley and Munafo, Marcus R 2017. Assessing causality in associations between cannabis use and schizophrenia risk: a two-sample Mendelian randomization study. *Psychological Medicine* 47 (5) , pp. 971-980. 10.1017/S0033291716003172

Publishers page: <https://doi.org/10.1017/S0033291716003172>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



Assessing causality in associations between cannabis use and schizophrenia risk: A two-sample Mendelian randomization study

Suzanne H. Gage (PhD) ^{1,2}, Hannah J. Jones (PhD) ¹, Stephen Burgess (PhD) ³, Jack Bowden (PhD) ^{1,4,5}, George Davey Smith (MD) ^{1,4}, Stanley Zammit (PhD) ^{4,6}, Marcus R. Munafò (PhD) ^{1,2}

1. MRC Integrative Epidemiology Unit (IEU) at the University of Bristol, United Kingdom.
2. UK Centre for Tobacco and Alcohol Studies, School of Experimental Psychology, University of Bristol, United Kingdom.
3. Department of Public Health and Primary Care, University of Cambridge, United Kingdom.
4. School of Social and Community Medicine, University of Bristol, United Kingdom.
5. MRC Biostatistics Unit, University of Cambridge, United Kingdom.
6. MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, United Kingdom.

Corresponding author: Suzanne H. Gage, School of Experimental Psychology, University of Bristol, 12a Priory Road, Bristol, BS8 1TU, United Kingdom. E: suzi.gage@bristol.ac.uk T: 0117 9289943

Key words: cannabis, schizophrenia, genetics, Mendelian randomization.

Word count: 3685

Support from the Medical Research Council (MC_UU_12013/1, MC_UU_12013/6, MR/M006727/1) is gratefully acknowledged. Stephen Burgess is supported by a fellowship from the Wellcome Trust (100114). Jack Bowden is supported by an MRC Methodology Research Fellowship (grant code MR/N501906/1).

Abstract

Background: Observational associations between cannabis and schizophrenia are well documented, but ascertaining causation is more challenging. We used Mendelian randomization (MR), utilizing publicly available data as a method for ascertaining causation from observational data.

Methods: We performed bi-directional two-sample MR using summary level genomewide data from the International Cannabis Consortium (ICC) and the Psychiatric Genomics Consortium (PGC2). Single nucleotide polymorphisms (SNPs) associated with cannabis initiation ($P < 10^{-5}$) and schizophrenia ($P < 5 \times 10^{-8}$) were combined using an inverse-variance weight fixed-effects approach. We also used height and education genomewide-association study data, representing negative and positive control analyses.

Results: There was some evidence consistent with a causal effect of cannabis initiation on risk of schizophrenia (OR 1.04 per doubling odds of cannabis initiation, 95% CI 1.01, 1.07, $P = 0.019$). There was strong evidence consistent with a causal effect of schizophrenia risk on likelihood of cannabis initiation (OR 1.10 per doubling of the odds of schizophrenia, 95% CI 1.05, 1.14, $P = 2.64 \times 10^{-5}$). Findings were as predicted for the negative control (height OR 1.00, 95% CI 0.99 to 1.01, $P = 0.90$) but weaker than predicted for the positive control (years in education OR 0.99, 95% CI 0.97 to 1.00, $P = 0.066$) analyses.

Conclusions: Our results provide some that cannabis initiation increases the risk of schizophrenia, though the size of the causal estimate is small. We find stronger evidence that schizophrenia risk predicts cannabis initiation, possibly as genetic instruments for schizophrenia are stronger than for cannabis initiation.

Assessing causality in associations between cannabis use and schizophrenia: A 2-sample Mendelian randomization study

Introduction

It is clear from human experimental studies that tetrahydrocannabinol (THC) intoxication can induce transient psychotic experiences (D'Souza et al., 2004). However, whether chronic cannabis use is causally associated with psychotic disorders such as schizophrenia is harder to ascertain (Gage et al., 2013, Gage et al., 2016a). A systematic review and meta-analysis (Moore et al., 2007) of longitudinal studies found evidence for a modest association between cannabis and later psychotic outcomes. However, the authors noted that the studies that adjusted for more potential confounding factors found that the point estimates attenuated to a greater degree than those adjusting for fewer, indicating that residual confounding could still be present. Some studies (Callaghan et al., 2012, Rossler et al., 2012), although not all (Gage et al., 2014), conducted since this meta-analysis have also found evidence of an association between cannabis use and later psychotic experiences or schizophrenia. As well as these observational studies, studies have been undertaken to investigate the genetic relationships between cannabis and schizophrenia. Power and colleagues (Power et al., 2014) found that a genetic risk score for schizophrenia predicted cannabis use in a sample of 2,082 healthy individuals. Verweij and colleagues (Verweij et al., in press) used LD-score regression to investigate the genetic correlation to investigate this association, and found evidence of a genetic correlation between lifetime cannabis use and risk of schizophrenia. This could be interpreted as shared genetic aetiology, but could also reflect a causal association between schizophrenia and risk of cannabis use (Gage et al., 2015).

Observational epidemiology alone cannot allow researchers to determine causality. Residual confounding and reverse causation are difficult or impossible to

rule out, and associations that appear robust observationally are sometimes shown to be spurious when experimentally tested (Davey Smith and Ebrahim, 2001). A further difficulty in conducting observational studies to investigate the association between cannabis use and psychosis is that it is not possible to rule out the impact of psychotic phenomena occurring during intoxication in regular users due to the long half-life of THC (Huestis et al., 1992). One technique that can allow causal inference from observational data is Mendelian randomization (MR). It has been described in detail elsewhere (Davey Smith and Hemani, 2014, Gage et al., 2016b); briefly, if genetic variants can be identified that robustly predict an exposure of interest, these can be used as unconfounded proxies for that exposure. Associations between the variants and the outcome of interest can thus provide evidence of causation whilst (subject to certain assumptions) eliminating problems of confounding or reverse causation. A recent genomewide association study (GWAS) of cannabis initiation (Stringer et al., 2016) did not find any single nucleotide polymorphisms (SNPs) that reached genomewide significance, but identified a number of SNPs with weaker evidence of association with cannabis initiation.

Two-sample MR allows for the use of publicly available summary data, where variant-exposure associations are identified from one GWAS, and variant-outcome associations from another (Burgess et al., 2015b). This can provide the large sample sizes required to provide adequate power to identify the small effect sizes of causal effects of common genetic variants in such studies. The current study uses two-sample MR to assess the likelihood of a causal association between cannabis initiation and schizophrenia, and Egger regression (Bowden et al., 2015) to investigate whether any association observed is due to pleiotropic effects of SNPs rather than causal effects of cannabis on schizophrenia. We also use two-sample MR with two other outcomes, years in education and height, as positive and negative controls respectively. Evidence of association between cannabis and a positive control (where causal effects are more strongly established), and lack of evidence of

association with a negative control (where causal effects are implausible) add support to causal interpretations. We chose education as a positive control as previous epidemiological evidence has suggested a probable causal association between cannabis and educational attainment (Meier et al., 2012). We chose height as a negative control outcome because cannabis is unlikely to have a causal effect on height since cannabis use is not common before mid- to late-adolescence (Eastwood, 2013).

Methods

The International Cannabis Consortium (ICC) GWAS, across the genome, explained 13-20% of the observed phenotypic variation of ever use of cannabis (Stringer et al., 2016). Although no genomewide-significant SNPs were identified, 153 SNPs were associated with cannabis initiation (ever/never use of cannabis) at $P < 10^{-5}$. A number of these SNPs were in high linkage disequilibrium. Where $R^2 > 0.9$, one SNP was randomly selected and the other highly correlated SNPs were removed from the analyses. 22 SNPs remained after this pruning, and a further SNP was removed due to poor quality (Vink, 2016, personal communication). The final 21 SNPs are listed in Table 1. Correlations $r^2 < 0.9$ were included in the analysis as a correlation matrix (Burgess et al., 2015b) (listed in Supplementary Table 1).

These 21 SNPs were then extracted from the schizophrenia Psychiatric Genetics Consortium (PGC2) GWAS (Schizophrenia-Working-Group-of-the-Psychiatric-Genomics-Consortium, 2014), a GWAS of educational attainment by the Social Science Genetics Association Consortium (SSGAC) (Rietveld et al., 2013) and the Genetic Investigation of Anthropometric Traits (GIANT) (Wood et al., 2014) GWAS of height. All 21 were in the schizophrenia GWAS, 8 were in the education GWAS, and 9 were in the height GWAS. We used the PhenoScanner tool (Staley et al., 2016) (<http://phenoscanner.medschl.cam.ac.uk>) to identify proxy SNPs using $R^2 > 0.9$ correlations calculated in European-descent individuals from 1000 Genomes

Phase 3, and a further 4 were identified, meaning that 12 SNPs were available for the cannabis-education and cannabis-height associations. These SNPs were used for the main analyses for the positive and negative controls. The full 21 SNPs were used for the main analysis to investigate the association between cannabis initiation and schizophrenia. We also conducted a sensitivity analysis restricting the cannabis-schizophrenia analysis to the same 12 SNPs used for the positive and negative control analyses. We used PhenoScanner to obtain estimates of the mutual correlations between variants, and a correlation matrix was included in the regression analysis model (Burgess et al., 2015a) (correlation matrix shown in Supplementary Table 1).

To assess the likelihood of a causal association between schizophrenia and cannabis initiation (schizophrenia risk predicting likelihood of cannabis initiation), we extracted 128 SNPs independently associated with schizophrenia at genomewide significance. These SNPs explain approximately 18% of the observed variance in schizophrenia risk (Schizophrenia-Working-Group-of-the-Psychiatric-Genomics-Consortium, 2014). Of the 128 SNPs, 107 were also present in the GWAS of cannabis initiation (listed in Supplementary Table 2).

Effect of cannabis initiation on schizophrenia risk

Log odds ratios and standard errors for the associations between the significant ($P < 10^{-5}$) SNPs and cannabis initiation were recorded. The same SNPs were then identified in the PGC2 GWAS of schizophrenia, and the corresponding log odds ratios and standard errors were recorded. The SNP-exposure and SNP-outcome coefficients were combined using an inverse-variance weighted approach to give an overall estimate of causal effect. This is equivalent to a weighted regression of the SNP-outcome coefficients on the SNP-exposure coefficients with the intercept constrained to zero.

The results of the analyses were converted to odds ratios. Given that schizophrenia is a binary outcome, the resulting causal effect estimate represents the odds for schizophrenia risk per unit increase in the log odds ratio for cannabis initiation, and is therefore somewhat hard to interpret. ORs have therefore been converted (by multiplying log odds ratios by 0.693 and then exponentiating) in order to represent the odds ratio per doubling in odds of the binary exposure.

Effect of schizophrenia risk on cannabis initiation

As above, the 107 SNPs associated with schizophrenia were extracted from the GWAS schizophrenia, and then identified in the GWAS of cannabis initiation, and log odds ratios and standard errors recorded. The SNP-exposure and SNP-outcome coefficients were combined as described above.

Positive and negative controls

The 12 cannabis initiation SNPs (or proxies where appropriate) were identified in GWAS of education and height, and beta coefficients and standard errors recorded, as described above. Although these GWASs were conducted on continuous outcomes, the resulting coefficients have been transformed as described above (multiplied by 0.693) for consistency.

Sensitivity analyses

We used MR-Egger regression to formally test for potential violations of Mendelian randomization assumptions. This method relaxes the assumption that the effects of genetic variants on the outcome operate entirely via the exposure, by allowing an intercept term in the weighted regression described above. The intercept parameter represents the overall pleiotropic effect of the SNPs on the outcome (i.e., a direct effect on the outcome rather than via the exposure, which would violate MR assumptions). If the intercept deviates from the null, this provides evidence that there

may be bias in the standard causal estimate due to pleiotropy. The beta coefficient for this analysis provides a causal estimate, assuming that the pleiotropic effect of SNPs on the outcome is not correlated with the associations between the SNPs and the exposure (MR Egger is described in more detail in Corbin et al., 2016).

With correlated genetic variants, the inverse-variance weighted regression model is extended to a generalized weighted regression model, as described previously (Burgess et al., 2015a). For MR-Egger regression, the same generalized weighted regression analysis was undertaken except including an intercept term in the regression model. The genetic associations with the outcome ($\hat{\beta}_Y$) were regressed on the genetic associations with the exposure (cannabis initiation, $\hat{\beta}_X$) using inverse-variance weights. The regression model is:

$$\hat{\beta}_Y = \theta_0 + \theta_E \hat{\beta}_X + \varepsilon$$

where θ_0 is the intercept term and θ_E is the Egger causal estimate and ε is an error term with zero mean and covariance matrix Ω . The weighting matrix Ω has terms $\Omega_{j_1 j_2} = \sigma_{Y j_1} \sigma_{Y j_2} \rho_{j_1 j_2}$, where $\sigma_{Y j}$ is the standard error of the genetic association with the outcome for the j th SNP, and $\rho_{j_1 j_2}$ is the correlation between the j_1 th and j_2 th SNPs. This regression gives the same estimates as standard Egger regression with uncorrelated variants when the correlations between different variants are all zero.

As another sensitivity analysis, the cannabis-schizophrenia association was also assessed using the 12 SNPs available for the education and height analyses, in order to allow a direct comparison. Given the potentially weak instruments used to predict cannabis initiation, we ran a further sensitivity analysis where we systematically removed each SNP from the analysis, in order to check the robustness of the findings.

Finally, we performed a calculation in order to estimate the relative risk of schizophrenia predicted by our upper confidence bound (UCB) comparing an individual who is likely to use cannabis (50% probability of initiation) with an

individual with a low probability of smoking cannabis (10% probability of initiation) using the calculation: $\exp((\log(0.5/(1-0.5))-\log(0.1/(1-0.1)))*\log(\text{UCB}/0.693))$.

Results

Effect of cannabis initiation on schizophrenia risk

In 36,989 schizophrenia cases and 113,075 controls, an MR analysis using 21 SNPs associated ($P < 10^{-5}$) with cannabis initiation indicated a small association with schizophrenia case status (OR 1.04, 95% CI 1.01 to 1.07 per doubling of the odds of initiation, $P = 0.019$).

Effect of schizophrenia risk on cannabis initiation

In 14,388 individuals who have used cannabis in their lifetime and 17,942 non-cannabis users, an MR analysis using 107 SNPs associated with schizophrenia at genome-wide significance ($P < 5 \times 10^{-8}$) provided strong evidence of an association between schizophrenia case status and lifetime cannabis initiation (OR 1.10, 95% CI 1.05 to 1.14 per doubling of the odds of schizophrenia, $P = 2.64 \times 10^{-5}$).

Positive and negative controls

In 126,599 individuals, an MR analysis using 12 SNPs associated ($P < 10^{-5}$) with cannabis initiation indicated a weak association with fewer years in education (OR 0.99, 95% CI 0.97 to 1.00 per doubling of the odds of initiation, $P = 0.066$) (the positive control). In 253,288 individuals, an MR analysis using these SNPs found no evidence of an association with height (OR 1.00, 95% CI 0.99 to 1.01, $P = 0.90$) (the negative control). These results are shown in Table 2.

Sensitivity analyses

A sensitivity MR-Egger regression indicated very little evidence of pleiotropy in the association between cannabis and schizophrenia using the full list of 21 SNPs

(intercept OR 0.99, 95% CI 0.98 to 1.01, $P = 0.50$). The strength of evidence for a causal association was considerably weaker in this analysis with wider CIs (OR 1.01, 95% CI 0.93 to 1.10, $P = 0.815$), although not inconsistent with the findings from the primary analysis.

There was no strong evidence of pleiotropy in the association between schizophrenia (exposure) and cannabis initiation (outcome) (intercept OR 1.00, 95% CI 0.98 to 1.01, $P = 0.52$). The point estimate was larger in this analysis as compared to the standard MR approach, however the strength of evidence of an association was weaker (OR 1.17, 95% CI 0.96 to 1.43, $P = 0.13$). The Egger plots for these analyses are shown in the supplementary materials (supplementary figures 1 and 2).

MR-Egger regression suggested weak evidence of pleiotropy in the positive and negative control outcome analyses using 12 SNPs (intercept OR education 0.99, 95% CI 0.97 to 1.00, $P = 0.18$; height 0.99, 95% CI 0.98 to 1.00, $P = 0.093$). The causal estimate from the MR-Egger regression for the association of cannabis initiation with years of education was slightly larger than for the main analysis, although the strength of evidence was very similar (OR 0.94, 95% CI 0.88 to 1.00, $P = 0.062$). The causal estimate from the MR-Egger regression analysis of cannabis initiation and height provided weak evidence for cannabis initiation being associated with shorter height (OR 0.95, 95% CI 0.88 to 1.01, $P = 0.095$), but was still consistent overall with the lack of association indicated in the primary analysis. These results are shown in Table 3.

When the association between cannabis initiation and schizophrenia risk was restricted to the 12 SNPs available for the education and height analyses, the resulting association was similar to the primary analysis, although confidence intervals were wider (OR 1.03, 95% CI 0.99 to 1.08, $P = 0.15$). This is shown in Tables 2 and 3.

When we systematically removed each SNP that predicted cannabis initiation in turn there was no evidence that the size of the effect was being driven by one SNP (Supplementary Figure 3).

Finally, we calculated the relative risk of schizophrenia for an individual who is likely to smoke cannabis (50% probability of initiation) to be at most 24% higher compared with an individual with a low probability of smoking cannabis (10% probability of initiation) based on the upper confidence interval for the causal effect estimate.

Discussion

Our results provide some evidence in support of the hypothesis that cannabis initiation increases the risk of schizophrenia, although the size of the causal estimate is small. Our findings therefore do not contradict the current observational literature suggesting an association between cannabis initiation and schizophrenia of a moderate size, and may suggest any true causal effect size is smaller still. Given the small effect sizes generally seen in the observational data (Moore et al., 2007) particularly when considering cannabis initiation phenotypes, and in these results using a different methodology, this converging evidence supports the theory that although cannabis use might be a risk factor for schizophrenia, any effect cannabis initiation has is likely to be small. For example, based on the upper bound of the 95% CI from our primary analysis, an individual who is more likely to smoke cannabis (50% probability of initiation) has at most a 24% higher relative risk of suffering from schizophrenia compared with an individual with a low probability of smoking cannabis (10% probability of initiation). However, these results are unlikely to be directly comparable to the observational studies in terms of comparing effect sizes, particularly with regard to heavier use of cannabis, or use of different strains of cannabis where ratios of cannabinoids may differ. The ratio of THC and cannabidiol (CBD) is likely to be particularly important in determining the psychotomimetic

properties of cannabis (D'Souza et al., 2004, Iseger and Bossong, 2015), although genetic variants that index the strength of cannabis used have not been identified as yet.

We found strong evidence in support of the reverse causation hypothesis, that schizophrenia risk predicts likelihood of cannabis initiation. These findings are consistent with a recent study that showed genetic predisposition for schizophrenia was associated with increased use of cannabis (Power et al., 2014). The authors of that study interpreted their findings as evidence for shared genetic aetiology between cannabis use and schizophrenia; however, an alternate interpretation, supported by the more directional analyses reported here, is that schizophrenia risk increases the risk of cannabis use. Our study cannot provide information about potential mechanisms through which this could happen. Causation could be via either biological or social mechanisms. For example, people with a higher genetic risk for schizophrenia could be more predisposed to try cannabis because they experience more pleasurable or positive effects of intoxication, or because they are more likely to associate with people who use cannabis (Nordsletten et al., 2016). However, these explanations remain speculative, although could be tested in further research, for example using a recall-by-genotype design. The findings from the positive and negative control analyses were broadly consistent with our prediction; our results suggest that cannabis initiation is associated with fewer years in education (though with weaker evidence than predicted), while cannabis initiation does not predict height.

Strengths and limitations

MR studies allow stronger evidence of causality than is possible from observational epidemiology, which is a strength of this study. Utilising both MR and positive and negative control designs allows for a triangulation of approaches with different limitations (Gage et al., 2016b). A particular strength of two-sample MR is

that it can provide the large sample sizes required to identify the small effect sizes that are likely when using SNPs as proxies to study complex phenotypes. However, there are a number of limitations to our study. Firstly, none of the SNPs used to predict cannabis initiation reached genomewide significance in the original GWAS (Stringer et al., 2016). As these SNPs are not strong predictors of cannabis initiation (and some may represent false positives), it is likely that our effect size estimates will be overly-conservative. The weak evidence of association with our positive control could be explained by the weakness of our genetic instrument. While these SNPs did not reach genomewide significance in the cannabis initiation GWAS, they were nevertheless associated with initiation at $P < 10^{-5}$. Inferences and estimates from a MR analysis may still be valid even if genetic variants are included that are not truly associated with the exposure; their inclusion is equivalent to adding noise to a genetic risk score. Including false positive variants is not ideal as it reduces power and could introduce pleiotropic variants to the analysis. However, as we do not know which variants are true associations and which are false positives, our approach makes the best possible use of the available data. We also conduct MR Egger sensitivity analyses, which are more robust to pleiotropy.

Secondly, there is a small overlap in the cohorts included in the schizophrenia and cannabis GWAS, as both include data from the Estonia Genome Centre cohort (EGCUT). EGCUT data makes up 2.3% of the PGC2 dataset, and approximately 10% of the Cannabis Consortium dataset. The true overlap is 1,500 individuals (Esko, 2016, personal communication). A small overlap such as this is unlikely to result in substantial bias (Burgess et al., 2016). Large overlap would bias the results in the direction of the observational association (and therefore could lead to an overestimation of the effect). We reran our main analyses in both directions using a version of the PGC2 data with this sample removed. Results did not change from those presented.

Thirdly, and perhaps more problematic, is that it is not clear what the cannabis initiation phenotype used in the Cannabis Consortium GWAS actually represents; participants were asked 'Have you ever used cannabis?', so there is no distinction between having tried cannabis once and having used it every day for many years. If there is a causal association between cannabis and schizophrenia, it is more likely that exposure intensity would be the relevant risk factor. Observationally, the association between ever use of cannabis and later psychosis is much weaker than for heavy cannabis use (Moore et al., 2007). Furthermore, those who choose to try cannabis (rather than become regular or heavy users) might be more impulsive or risk taking than those who never try it, and this GWAS could therefore be a measure of that or other related phenotypes, rather than cannabis initiation per se. Our lack of understanding of the function of the genetic variants identified in the cannabis initiation GWAS and their role in cannabis metabolism means that we do not understand the mechanism by which they impact on cannabis use.

A recent GWAS of cannabis use identified 3 SNPs at genomewide significance in a sample of 14754 participants (Sherva et al., 2016). However, this GWAS conflates cannabis initiation with dependence phenotypes, which would make interpretation of any association we might find challenging. Also, given our inability to stratify participants in the schizophrenia PGC sample by cannabis use status, we deemed it inappropriate to use these SNPs as genetic influences on cannabis dependence cannot occur unless a person has used cannabis. This analysis needs to be conducted in cannabis users separately from non-users; if the association between cannabis and schizophrenia is causal, there should be an association observed between the cannabis dependence SNPs and schizophrenia in cannabis dependent subjects, but no association observed in non-users.

A further limitation relates to the nature of schizophrenia. Since the vast majority of the population do not have schizophrenia, we are making an assumption

that genetic predictors of schizophrenia also have an impact in the general population. There is evidence that psychotic symptoms exist on a continuum (Linscott and van Os, 2013), which would mean this is an appropriate assumption to make. It has also been found that genetic risk for schizophrenia predicts childhood psychopathology (Jones et al., 2016).

Finally, a further limitation is our choice of positive and negative controls. Whilst weak evidence of association may be due to the use of a weak genetic instrument, as discussed above, it is also possible that our assumption of causality between cannabis and education is misplaced. Since we began our analyses, two studies have been published suggesting that the association between cannabis and education may be more strongly confounded than previously believed (Jackson et al., 2016, Mokrysz et al., 2016). With regards to height as a negative control, there is some evidence of a negative correlation between height and schizophrenia risk, that could be due to pleiotropy (Bacanu et al., 2013). Identifying appropriate positive and negative controls is extremely challenging when investigating complex phenotypes such as substance use and mental health. However, in the interests of completeness we present both our positive and negative control analyses as originally conceived.

Given that schizophrenia and cannabis initiation are both complex phenotypes, likely to be influenced by multiple genetic and environmental risk factors, a consideration of pleiotropy is important. We conducted MR Egger analyses to formally test for biological pleiotropy (where genetic variants have a direct impact on more than one phenotype). Biological pleiotropy violates the assumptions of MR, but we find only weak evidence of biological pleiotropy. It is distinct from mediated pleiotropy (whereby a genetic variant may have a mediated impact on a later phenotype via a direct effect on an upstream phenotype), which does not violate the assumptions of MR. A recent paper (Verweij et al., in press) used LD-score regression to assess the association between cannabis use and schizophrenia, and

found evidence of a genetic correlation. This could mean shared genetic architecture, but it is also consistent with a causal relationship in one or both directions.

Conclusion

Our results provide some evidence that cannabis initiation might be causally associated with odds of developing schizophrenia, using two-sample MR to support stronger causal inference, although the size of the association is small. The results provide stronger evidence that schizophrenia risk predicts likelihood of cannabis initiation, but the weakness of the currently-available genetic instrument for cannabis use limits our interpretation of these results. Whilst stronger instruments are likely to be identified with increasing GWAS sample sizes, future GWAS studies of schizophrenia should prioritise recording cumulative cannabis exposure in cases and controls to allow the causal association between heaviness of cannabis use and schizophrenia to be investigated in stratified samples.

Acknowledgments

SHG and MRM are members of the UK Centre for Tobacco and Alcohol Studies, a UK Clinical Research Council Public Health Research: Centre of Excellence. Funding from British Heart Foundation, Cancer Research UK, Economic and Social Research Council, Medical Research Council, and the National Institute for Health Research, under the auspices of the UK Clinical Research Collaboration, is gratefully acknowledged. Support from the Medical Research Council (MC_UU_12013/1, MC_UU_12013/6, MR/M006727/1) is also gratefully acknowledged. Stephen Burgess is supported by a fellowship from the Wellcome Trust (100114). Jack Bowden is supported by an MRC Methodology Research Fellowship (grant code MR/N501906/1). The authors would like to thank Dr Tõnu Esko from the EGCUT for his assistance in obtaining the PGC2 data without the EGCUT sample.

References

- BACANU, S. A., CHEN, X. & KENDLER, K. S. 2013. The genetic overlap between schizophrenia and height. *Schizophrenia Research*, 151, 226-8.
- BOWDEN, J., DAVEY SMITH, G. & BURGESS, S. 2015. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *International Journal of Epidemiology*, 44, 512-525.
- BURGESS, S., DAVIES, N. M. & THOMPSON, S. G. 2016. Bias due to participant overlap in two-sample Mendelian randomization. *Genetic Epidemiology*, 40, 597-608.
- BURGESS, S., DUDBRIDGE, F. & THOMPSON, S. G. 2015a. Combining information on multiple instrumental variables in Mendelian randomization: comparison of allele score and summarized data methods. *Statistics in Medicine*, 35, 1880-1906.
- BURGESS, S., SCOTT, R. A., TIMPSON, N. J., DAVEY SMITH, G., THOMPSON, S. G. & CONSORTIUM, E.-I. 2015b. Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors. *European Journal of Epidemiology*, 30, 543-52.
- CALLAGHAN, R. C., CUNNINGHAM, J. K., ALLEBECK, P., ARENOVICH, T., SAJEEV, G., REMINGTON, G., BOILEAU, I. & KISH, S. J. 2012. Methamphetamine use and schizophrenia: a population-based cohort study in California. *American Journal of Psychiatry*, 169, 389-96.
- CORBIN, L. J., RICHMOND, R. C., WADE, K. H., BURGESS, S., BOWDEN, J., SMITH, G. D. & TIMPSON, N. J. 2016. BMI as a Modifiable Risk Factor for Type 2 Diabetes: Refining and Understanding Causal Estimates Using Mendelian Randomization. *Diabetes*, 65, 3002-7.
- D'SOUZA, D. C., PERRY, E., MACDOUGALL, L., AMMERMAN, Y., COOPER, T., WU, Y. T., BRALEY, G., GUEORGUIEVA, R. & KRYSTAL, J. H. 2004. The psychotomimetic effects of intravenous delta-9-tetrahydrocannabinol in healthy individuals: implications for psychosis. *Neuropsychopharmacology*, 29, 1558-72.
- DAVEY SMITH, G. & EBRAHIM, S. 2001. Epidemiology--is it time to call it a day? *International Journal of Epidemiology*, 30, 1-11.
- DAVEY SMITH, G. & HEMANI, G. 2014. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Human Molecular Genetics*, 23, R89-98.
- EASTWOOD, P. 2013. *Statistics on drug misuse: England 2013*, Leeds, Health and Social Care Information Centre.
- GAGE, S. H., DAVEY SMITH, G. & MUNAFO, M. 2015. *Comment on PMID 24957864: Genetic predisposition to schizophrenia associated with increased use of cannabis*. [Online]. National Library of Medicine. Available: <https://www.ncbi.nlm.nih.gov/pubmed/24957864> [Accessed].
- GAGE, S. H., HICKMAN, M., HERON, J., MUNAFO, M., LEWIS, G., MACLEOD, J. & ZAMMIT, S. 2014. Associations of cannabis and cigarette use with psychotic experiences at age 18: findings from the Avon Longitudinal Study of Parents and Children. *Psychological Medicine*, 44, 3435-3444.
- GAGE, S. H., HICKMAN, M. & ZAMMIT, S. 2016a. Association Between Cannabis and Psychosis: Epidemiologic Evidence. *Biological Psychiatry*, 79, 549-556.

- GAGE, S. H., MUNAFO, M. R. & DAVEY SMITH, G. 2016b. Causal Inference in Developmental Origins of Health and Disease (DOHaD) Research. *Annual Review of Psychology*, 67, 567-85.
- GAGE, S. H., ZAMMIT, S. & HICKMAN, M. 2013. Stronger evidence is needed before accepting that cannabis plays an important role in the aetiology of schizophrenia in the population. *F1000 Medicine Reports*, 5, 2.
- HUESTIS, M. A., HENNINGFIELD, J. E. & CONE, E. J. 1992. Blood cannabinoids. I. Absorption of THC and formation of 11-OH-THC and THCCOOH during and after smoking marijuana. *Journal of Analytical Toxicology*, 16, 276-82.
- ISEGER, T. A. & BOSSONG, M. G. 2015. A systematic review of the antipsychotic properties of cannabidiol in humans. *Schizophrenia Research*, 162, 153-161.
- JACKSON, N. J., ISEN, J. D., KHODDAM, R., IRONS, D., TUVBLAD, C., IACONO, W. G., MCGUE, M., RAINE, A. & BAKER, L. A. 2016. Impact of adolescent marijuana use on intelligence: Results from two longitudinal twin studies. *Proceedings of the National Academy of Sciences*, 113, E500-8.
- JONES, H. J., STERGIAKOULI, E., TANSEY, K. E., HUBBARD, L., HERON, J., CANNON, M., HOLMANS, P., LEWIS, G., LINDEN, D. E. J., JONES, P. B., DAVEY SMITH, G., O'DONOVAN, M. C., OWEN, M. J., WALTER, J. T. & ZAMMIT, S. 2016. Phenotypic manifestation of genetic risk for schizophrenia during adolescence in the general population. *JAMA Psychiatry*, 73, 221-228.
- LINSCOTT, R. J. & VAN OS, J. 2013. An updated and conservative systematic review and meta-analysis of epidemiological evidence on psychotic experiences in children and adults: on the pathway from proneness to persistence to dimensional expression across mental disorders. *Psychological Medicine*, 43, 1133-49.
- MEIER, M. H., CASPI, A., AMBLER, A., HARRINGTON, H., HOUTS, R., KEEFE, R. S., MCDONALD, K., WARD, A., POULTON, R. & MOFFITT, T. E. 2012. Persistent cannabis users show neuropsychological decline from childhood to midlife. *Proceedings of the National Academy of Sciences*, 109, E2657-64.
- MOKRYSZ, C., LANDY, R., GAGE, S. H., MUNAFO, M. R., ROISER, J. P. & CURRAN, H. V. 2016. Are IQ and educational outcomes in teenagers related to their cannabis use? A prospective cohort study. *Journal of Psychopharmacology*, 30, 159-68.
- MOORE, T. H., ZAMMIT, S., LINGFORD-HUGHES, A., BARNES, T. R., JONES, P. B., BURKE, M. & LEWIS, G. 2007. Cannabis use and risk of psychotic or affective mental health outcomes: a systematic review. *Lancet*, 370, 319-28.
- NORDSLETTEN, A. E., LARSSON, H., CROWLEY, J. J., ALMQVIST, C., LICHTENSTEIN, P. & MATAIX-COLS, D. 2016. Patterns of Nonrandom Mating Within and Across 11 Major Psychiatric Disorders. *JAMA Psychiatry*, 73, 354-61.
- POWER, R. A., VERWEIJ, K. J., ZUHAIR, M., MONTGOMERY, G. W., HENDERS, A. K., HEATH, A. C., MADDEN, P. A., MEDLAND, S. E., WRAY, N. R. & MARTIN, N. G. 2014. Genetic predisposition to schizophrenia associated with increased use of cannabis. *Molecular Psychiatry*, 19, 1201-1204.
- RIETVELD, C. A., MEDLAND, S. E., DERRINGER, J., YANG, J., ESKO, T., MARTIN, N. W., WESTRA, H. J., SHAKHBAZOV, K., ABDELLAOUI, A., AGRAWAL, A., ALBRECHT, E., ALIZADEH, B. Z., AMIN, N.,

- BARNARD, J., BAUMEISTER, S. E., BENKE, K. S., BIELAK, L. F., BOATMAN, J. A., BOYLE, P. A., DAVIES, G., DE LEEUW, C., EKLUND, N., EVANS, D. S., FERHMANN, R., FISCHER, K., GIEGER, C., GJESSING, H. K., HAGG, S., HARRIS, J. R., HAYWARD, C., HOLZAPFEL, C., IBRAHIM-VERBAAS, C. A., INGELSSON, E., JACOBSSON, B., JOSHI, P. K., JUGESSUR, A., KAAKINEN, M., KANONI, S., KARJALAINEN, J., KOLCIC, I., KRISTIANSOON, K., KUTALIK, Z., LAHTI, J., LEE, S. H., LIN, P., LIND, P. A., LIU, Y., LOHMAN, K., LOITFELDER, M., MCMAHON, G., VIDAL, P. M., MEIRELLES, O., MILANI, L., MYHRE, R., NUOTIO, M. L., OLDMEADOW, C. J., PETROVIC, K. E., PEYROT, W. J., POLASEK, O., QUAYE, L., REINMAA, E., RICE, J. P., RIZZI, T. S., SCHMIDT, H., SCHMIDT, R., SMITH, A. V., SMITH, J. A., TANAKA, T., TERRACCIANO, A., VAN DER LOOS, M. J., VITART, V., VOLZKE, H., WELLMANN, J., YU, L., ZHAO, W., ALLIK, J., ATTIA, J. R., BANDINELLI, S., BASTARDOT, F., BEAUCHAMP, J., BENNETT, D. A., BERGER, K., BIERUT, L. J., BOOMSMA, D. I., BULTMANN, U., CAMPBELL, H., CHABRIS, C. F., CHERKAS, L., CHUNG, M. K., CUCCA, F., DE ANDRADE, M., DE JAGER, P. L., DE NEVE, J. E., DEARY, I. J., DEDOUSSIS, G. V., DELOUKAS, P., DIMITRIOU, M., EIRIKSDOTTIR, G., ELDERSON, M. F., ERIKSSON, J. G., et al. 2013. GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science*, 340, 1467-71.
- ROSSLER, W., HENGARTNER, M. P., ANGST, J. & AJDACIC-GROSS, V. 2012. Linking substance use with symptoms of subclinical psychosis in a community cohort over 30 years. *Addiction*, 107, 1174-84.
- SCHIZOPHRENIA-WORKING-GROUP-OF-THE-PSYCHIATRIC-GENOMICS-CONSORTIUM 2014. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, 511, 421-427.
- SHERVA, R., WANG, Q., KRANZLER, H., ZHAO, H., KOESTERER, R., HERMAN, A., FARRER, L. A. & GELERNTER, J. 2016. Genome-wide Association Study of Cannabis Dependence Severity, Novel Risk Variants, and Shared Genetic Risks. *JAMA Psychiatry*, 73, 472-80.
- STALEY, J. R., BLACKSHAW, J., KAMAT, M. A., ELLIS, S., SURENDRAN, P., SUN, B. B., PAUL, D. S., FREITAG, D., BURGESS, S., DANESH, J., YOUNG, R. & BUTTERWORTH, A. S. 2016. PhenoScanner: a database of human genotype-phenotype associations. *Bioinformatics*, 32, 3207-3209.
- STRINGER, S., MINICA, C. C., VERWEIJ, K. J., MBAREK, H., WALTERS, R., ABDDELLAOUI, A., BRUINENBERG, M., DERRINGER, J. & VAN EIJK, K. R. 2016. Genome wide association study of lifetime cannabis use based on a large meta-analytic sample of 32,330 subjects from the International Cannabis Consortium: Identification of four risk genes including NCAM1 *Translational Psychiatry*, 29, e769.
- VERWEIJ, K. J., ABDELLAOUI, A., NIVARD, M. G., SAINZ CORT, A., LIGTHART, L., DRAISMA, H. H. M., MINICA, C. C., CONSORTIUM, I. C., GILLESPIE, N. A., HOTTENGA, J. J., BOOMSMA, D. I. & VINK, J. M. in press. Brief Report. Genetic association between schizophrenia and cannabis use. *Drug and Alcohol Dependence*.
- WOOD, A. R., ESKO, T., YANG, J., VEDANTAM, S., PERS, T. H., GUSTAFSSON, S., CHU, A. Y., ESTRADA, K., LUAN, J., KUTALIK, Z.,

AMIN, N., BUCHKOVICH, M. L., CROTEAU-CHONKA, D. C., DAY, F. R., DUAN, Y., FALL, T., FEHRMANN, R., FERREIRA, T., JACKSON, A. U., KARJALAINEN, J., LO, K. S., LOCKE, A. E., MAGI, R., MIHAILOV, E., PORCU, E., RANDALL, J. C., SCHERAG, A., VINKHUYZEN, A. A., WESTRA, H. J., WINKLER, T. W., WORKALEMAHU, T., ZHAO, J. H., ABSHER, D., ALBRECHT, E., ANDERSON, D., BARON, J., BEEKMAN, M., DEMIRKAN, A., EHRET, G. B., FEENSTRA, B., FEITOSA, M. F., FISCHER, K., FRASER, R. M., GOEL, A., GONG, J., JUSTICE, A. E., KANONI, S., KLEBER, M. E., KRISTIANSSEN, K., LIM, U., LOTAY, V., LUI, J. C., MANGINO, M., MATEO LEACH, I., MEDINA-GOMEZ, C., NALLS, M. A., NYHOLT, D. R., PALMER, C. D., PASKO, D., PECHLIVANIS, S., PROKOPENKO, I., RIED, J. S., RIPKE, S., SHUNGIN, D., STANCAKOVA, A., STRAWBRIDGE, R. J., SUNG, Y. J., TANAKA, T., TEUMER, A., TROMPET, S., VAN DER LAAN, S. W., VAN SETTEN, J., VAN VLIET-OSTAPTCHOUK, J. V., WANG, Z., YENGO, L., ZHANG, W., AFZAL, U., ARNLOV, J., ARSCOTT, G. M., BANDINELLI, S., BARRETT, A., BELLIS, C., BENNETT, A. J., BERNE, C., BLUHER, M., BOLTON, J. L., BOTTCHER, Y., BOYD, H. A., BRUINENBERG, M., BUCKLEY, B. M., BUYSKE, S., CASPERSEN, I. H., CHINES, P. S., CLARKE, R., CLAUDI-BOEHM, S., COOPER, M., DAW, E. W., DE JONG, P. A., DEELEN, J., DELGADO, G., et al. 2014. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genetics*, 46, 1173-86.

Table 1: List of SNPs (correlated < 0.9) associated with cannabis initiation ($P < 10^{-5}$), and proxies where used.

rs number	Chromosome	Nearest gene (within 100k base pairs)	beta (se) cannabis initiation	beta (se) schizophrenia	Proxy	r ² for proxy
rs7675351	4	SCOC	0.15 (0.03)	0.00 (0.02)		
rs7700636	5	NR3C1	0.15 (0.03)	0.01 (0.01)		
rs17237367	15	RORA	0.12 (0.03)	0.01 (0.01)		
rs4984458*	15	-	0.10 (0.02)	-0.01 (0.01)	rs12906344	1.000
rs4984460	15	-	0.12 (0.03)	-0.01 (0.01)		
rs3738226	1	RCSD1	0.10 (0.02)	0.00 (0.01)		
rs74944517*	2	LOC129656	0.29 (0.06)	0.02 (0.02)	rs1526674	1.000
rs2326313*	3	CADM2	0.10 (0.02)	-0.01 (0.01)	rs9819830	1.000
rs6840574	4	-	0.15 (0.03)	0.01 (0.01)		
rs1554927	8	CALB1	0.10 (0.02)	0.02 (0.01)		
rs12789616	11	NCAM1	0.09 (0.02)	0.02 (0.01)		
rs8041045*	15	-	0.10 (0.02)	0.00 (0.01)	rs12439562	0.971
rs73067624 *	1	KCNT2	0.18 (0.04)	0.02 (0.02)		
rs12518098 *	5	ZSWIM6	0.10 (0.02)	0.02 (0.01)		
rs353253 *	5	PCYOX1L	0.11 (0.03)	-0.01 (0.01)		
rs8102250 *	19	CRX	0.13 (0.03)	0.01 (0.01)		
rs113019398 *	20	LINC00687	0.17 (0.04)	-0.03 (0.02)		
rs2033867 *	2	SP9	0.24 (0.05)	0.05 (0.02)		
rs13063578 *	3	SETD2	0.11 (0.02)	0.02 (0.01)		
rs7107987 *	11	CHID1	0.27 (0.06)	0.01 (0.01)		
rs12313672 *	12	RNA5SP356	0.14 (0.03)	0.02 (0.01)		

* SNPs not in education / height GWAS

Table 2: Associations between cannabis initiation and various outcomes and schizophrenia and cannabis initiation using 2-sample MR.

Exposure	Outcome	OR*	95% CI	p-value	N SNPs
Cannabis initiation	Schizophrenia (binary outcome)	1.04 [†]	1.01 to 1.07	0.019	21
Cannabis initiation	Education (years)	0.99 [†]	0.97 to 1.00	0.066	12
Cannabis initiation	Height (z-score)	1.00 [†]	0.99 to 1.01	0.901	12
Cannabis initiation	Schizophrenia (binary outcome)	1.03 [†]	0.99 to 1.08	0.145	12 [#]
Schizophrenia	Cannabis initiation (binary outcome)	1.10	1.05 to 1.14	2.64E-05	107

* per doubling in odds of exposure; [†] correlated likelihood method used; [#] restricted number of SNPs to be comparable to positive/negative control analyses

Table 3: Egger regression analyses showing intercept and causal estimate values for associations between cannabis initiation and various outcomes and schizophrenia and cannabis initiation.

Exposure	Outcome		OR*	95% CI	p	N SNPs
Cannabis initiation	Schizophrenia (binary outcome)	Intercept	0.99 [†]	0.98 to 1.01	0.50	21
		Causal estimate	1.01 [†]	0.93 to 1.10	0.815	
Cannabis initiation	Education (years)	Intercept	0.99 [†]	0.99 to 1.00	0.182	12
		Causal estimate	0.94 [†]	0.88 to 1.00	0.062	
Cannabis initiation	Height (z-score)	Intercept	0.99 [†]	0.98 to 1.00	0.093	12
		Causal estimate	0.95 [†]	0.88 to 1.01	0.095	
Cannabis initiation	Schizophrenia (binary outcome)	Intercept	1.00 [†]	0.97 to 1.03	0.801	12 [#]
		Causal estimate	1.01 [†]	0.88 to 1.17	0.854	
Schizophrenia	Cannabis initiation (binary outcome)	Intercept	1.00	0.98 to 1.01	0.524	107
		Causal estimate	1.17	0.96 to 1.43	0.128	

* per doubling in odds of exposure; [†] correlated likelihood method used; [#] restricted number of SNPs to be comparable to positive/negative control analyses

Supplementary material

Supplementary Table 1: Correlation matrix of SNPs associated with cannabis initiation.

	rs4984460	rs4984458	rs8041045
rs4984460	1	0.18	0.12
rs4984458	0.18	1	0.65
Rs8041045	0.12	0.65	1

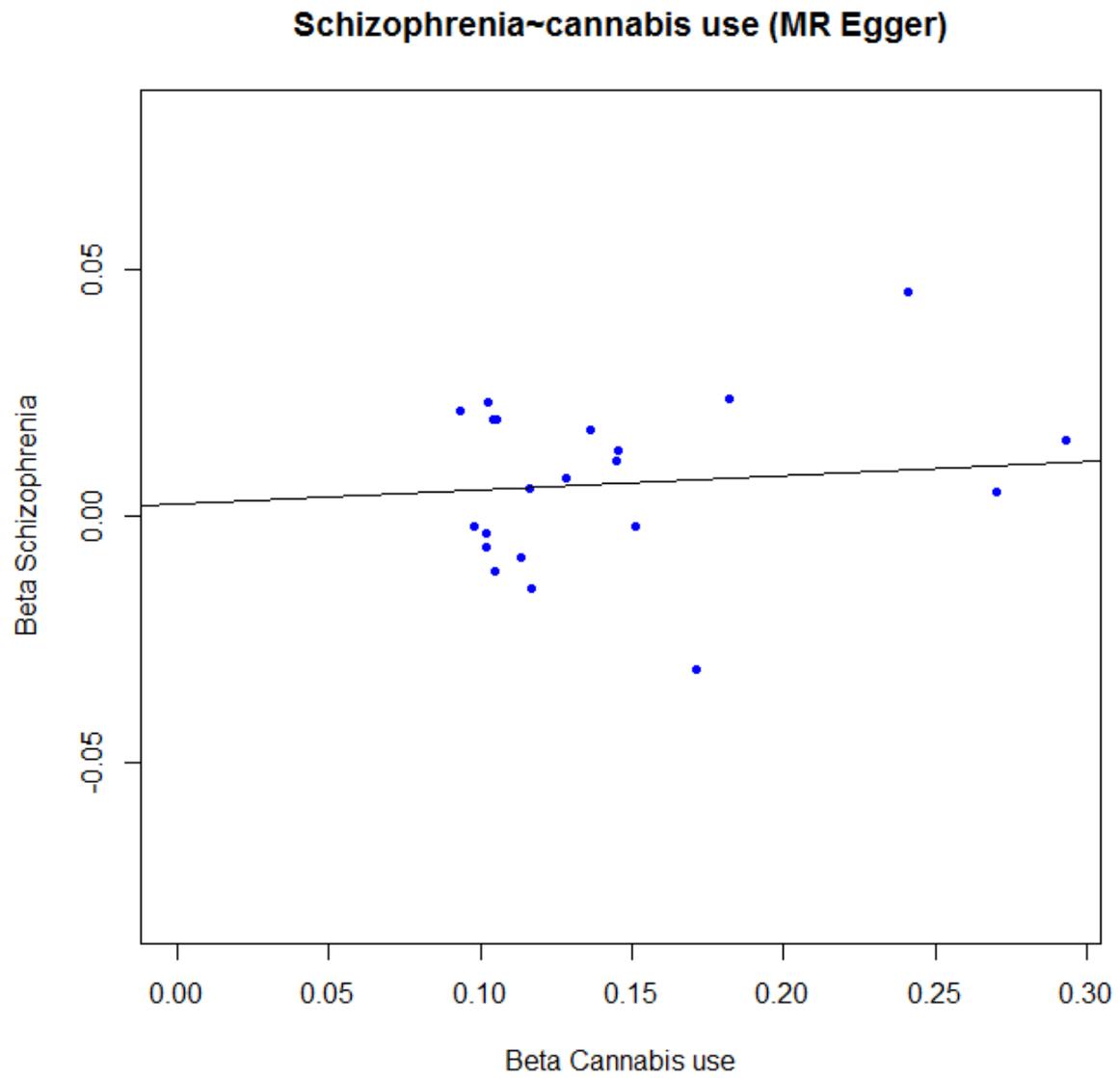
r^2 shown; all other SNPs were not correlated with each other.

Supplementary Table 2: List of SNPs associated with schizophrenia ($P < 5 \times 10^{-8}$).

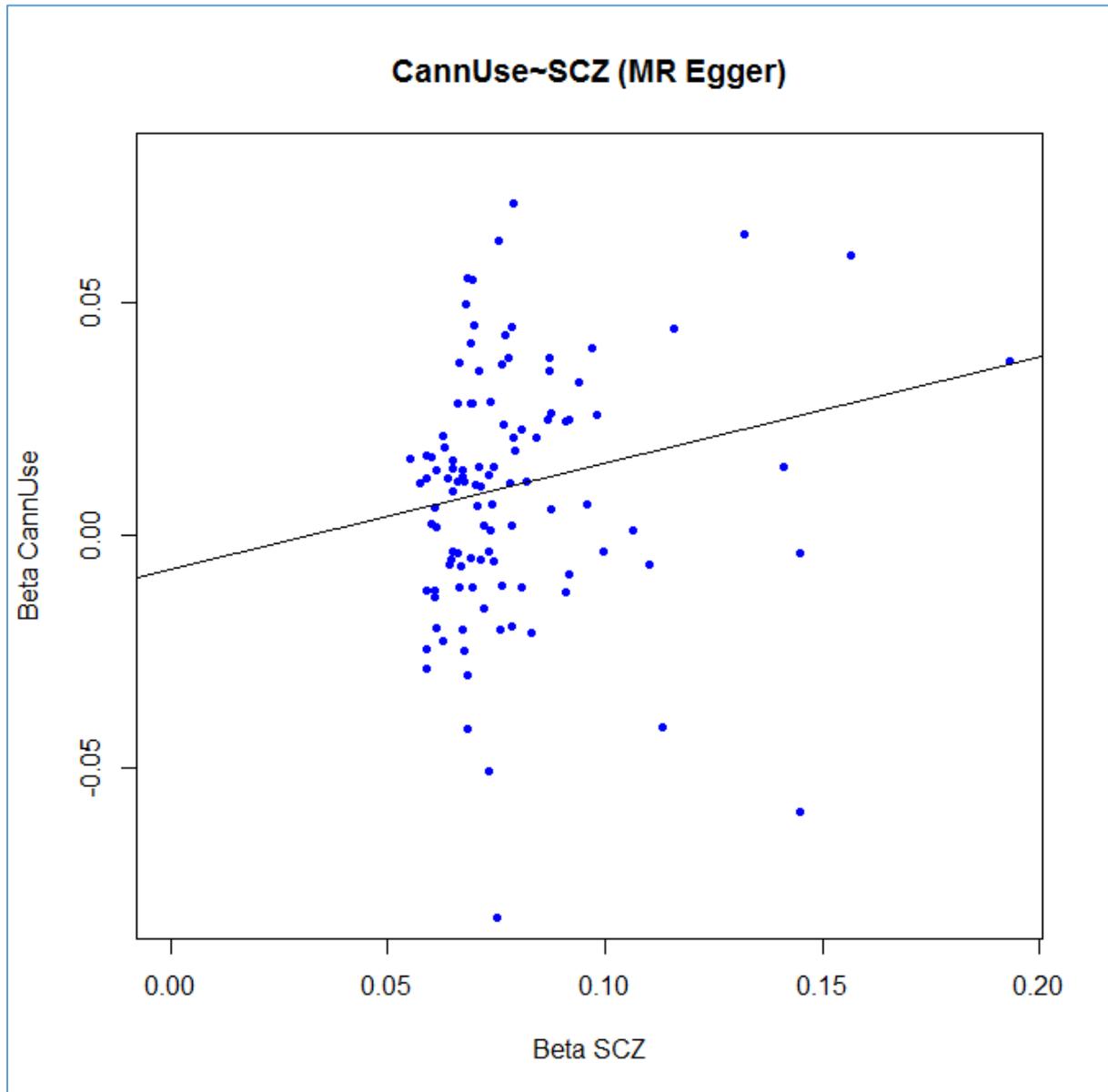
SNP ID	Chr	SNP ID	Chr	SNP ID	Chr	SNP ID	Chr
rs12129573	1	rs4330281	3	rs10503253	8	rs12887734	14
rs11210892	1	rs832187	3	rs73229090	8	rs4702	15
rs77149735*	1	rs17194490	3	rs36068923	8	rs8042374	15
rs10803138	1	rs1106568	4	rs6984242	8	rs12903146	15
rs7523273	1	rs35518360	4	rs7819570	8	rs56205728	15
rs76869799*	1	rs215411	4	rs11139497	9	rs190065944	15
chr1_8424984_D*	1	rs10520163	4	rs11191419	10	rs950169	15
chr1_243881945_I*	1	rs4391122	5	chr10_104957618_I*	10	rs12148337	15
rs1702294	1	rs111294930	5	rs7893279	10	rs12691307	16
rs140505938	1	rs16867576	5	rs7907645	10	rs9922678	16
rs4648845	1	rs11740474	5	rs55833108	10	rs8044995	16
rs1498232	1	rs12522290	5	rs55661361	11	rs12325245	16
rs14403	1	chr5_140143664_I*	5	rs75059851	11	rs7405404	16
rs6670165	1	rs2973155	5	rs9420	11	rs8082590	17
rs59979824	2	rs3849046	5	rs11027857	11	rs4523957	17
rs11685299	2	rs1501357	5	rs77502336	11	rs9636107	18
rs6704768	2	rs10043984	5	chr11_46350213_D*	11	chr18_52749216_D*	18
rs6434928	2	rs4388249	5	rs2514218	11	rs72934570	18
rs3768644	2	rs79212538	5	rs12421382	11	rs715170	18
rs6704641	2	rs115329265	6	rs10791097	11	rs78322266*	18
rs2909457	2	chr6_84280274_D*	6	rs2239063	12	rs2053079	19
rs75575209	2	rs117074560	6	rs324017	12	rs2905426	19
chr2_200825237_I*	2	rs1339227	6	rs679087	12	rs56873913	19
chr2_146436222_I*	2	rs6466055	7	rs2007044	12	rs6065094	20
chr2_149429178_D*	2	rs12704290	7	rs2851447	12	rs7267348	20
rs11693094	2	rs7801375	7	rs4240748	12	rs9607782	22
rs11682175	2	chr7_2025096_I*	7	rs4766428	12	chr22_39987017_D*	22
rs7432375	3	chr7_24747494_D*	7	rs10860964	12	rs6002655	22
rs9841616	3	rs13240464	7	rs12826178*	12	rs1023500	22
chr3_180594593_I*	3	rs3735025	7	rs2693698	14	rs12845396*	23
rs75968099	3	rs211829	7	rs2332700	14	rs1378559*	23
rs2535627	3	rs4129585	8	rs2068012	14	rs5937157*	23

* SNPs not in cannabis initiation GWAS. Note: Chr, chromosome.

Supplementary figure 1: MR Egger regression of cannabis initiation predicting risk of schizophrenia.



Supplementary Figure 2: MR Egger regression risk of schizophrenia predicting cannabis initiation.



Supplementary Figure 3: Forest plot showing the effect size and confidence intervals of Mendelian randomization associations between cannabis initiation and risk of schizophrenia, systematically removing each cannabis initiation SNP in turn

