

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

S. H. Gage^{1,2*}, H. J. Jones^{1,4}, S. Burgess³, J. Bowden^{1,4,5}, G. Davey Smith^{1,4}, S. Zammit^{4,6}
and M. R. Munafò^{1,2}

¹MRC Integrative Epidemiology Unit (IEU), University of Bristol, Bristol, UK

²UK Centre for Tobacco and Alcohol Studies, School of Experimental Psychology, University of Bristol, Bristol, UK

³Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

⁴School of Social and Community Medicine, University of Bristol, Bristol, UK

⁵MRC Biostatistics Unit, University of Cambridge, Cambridge, UK

⁶MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Cardiff, UK

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.

Method. We performed bi-directional two-sample MR using summary-level genome-wide 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-weighted fixed-effects approach. We also used height and education genome-wide 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 [odds ratio (OR) 1.04 per doubling odds of cannabis initiation, 95% confidence interval (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 0.99–1.01, $p = 0.90$) but weaker than predicted for the positive control (years in education: OR 0.99, 95% CI 0.97–1.00, $p = 0.066$) analyses.

Conclusions. Our results provide some evidence that cannabis initiation increases the risk of schizophrenia, although 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.

Received 25 August 2016; Revised 8 November 2016; Accepted 9 November 2016; First published online 8 December 2016

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

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, 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 *et al.* (2014) found that a genetic risk score for schizophrenia predicted cannabis use in a sample of 2082 healthy individuals. Verweij *et al.* (in press) used linkage

* Address for correspondence: S. H. Gage, Ph.D., School of Experimental Psychology, University of Bristol, 12a Priory Road, Bristol BS8 1TU, UK.
(Email: suzi.gage@bristol.ac.uk)

disequilibrium (LD) score regression 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 & 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 & 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 while (subject to certain assumptions) eliminating problems of confounding or reverse causation. A recent genome-wide association study (GWAS) of cannabis initiation (Stringer *et al.* 2016) did not find any single nucleotide polymorphisms (SNPs) that reached genome-wide 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 MR-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).

Method

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 genome-wide 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. Twenty-two 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 S1).

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, eight were in the education GWAS, and nine 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 four 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 S1).

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 100000 base pairs)	β (s.e.) cannabis initiation	β (s.e.) 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 ^a	15	–	0.10 (0.02)	–0.01 (0.01)	rs12906344	1.000
rs4984460	15	–	0.12 (0.03)	–0.01 (0.01)		
rs3738226	1	RCS1	0.10 (0.02)	0.00 (0.01)		
rs74944517 ^a	2	LOC129656	0.29 (0.06)	0.02 (0.02)	rs1526674	1.000
rs2326313 ^a	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 ^a	15	–	0.10 (0.02)	0.00 (0.01)	rs12439562	0.971
rs73067624 ^a	1	KCNT2	0.18 (0.04)	0.02 (0.02)		
rs12518098 ^a	5	ZSWIM6	0.10 (0.02)	0.02 (0.01)		
rs353253 ^a	5	PCYOX1L	0.11 (0.03)	–0.01 (0.01)		
rs8102250 ^a	19	CRX	0.13 (0.03)	0.01 (0.01)		
rs113019398 ^a	20	LINC00687	0.17 (0.04)	–0.03 (0.02)		
rs2033867 ^a	2	SP9	0.24 (0.05)	0.05 (0.02)		
rs13063578 ^a	3	SETD2	0.11 (0.02)	0.02 (0.01)		
rs7107987 ^a	11	CHID1	0.27 (0.06)	0.01 (0.01)		
rs12313672 ^a	12	RNA5SP356	0.14 (0.03)	0.02 (0.01)		

^a Single nucleotide polymorphisms (SNPs) not in education/height genome-wide association study.

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 genome-wide 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 S2).

Effect of cannabis initiation on schizophrenia risk

Log odds ratios (ORs) and s.e. 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 ORs and s.e. 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 ORs. Given that schizophrenia is a binary outcome, the

resulting causal effect estimate represents the odds for schizophrenia risk per unit increase in the log OR for cannabis initiation, and is therefore somewhat hard to interpret. ORs have therefore been converted (by multiplying log ORs by 0.693 and then exponentiating) in order to represent the OR 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 ORs and s.e. 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 β coefficients and s.e. 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 MR 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 β 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 σ_{Y_j} is the s.e. 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 MR-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% confidence interval (CI) 1.01–1.07 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 1.05–1.14 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 0.97–1.00 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 0.99–1.01, $p = 0.90$) (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 0.98–1.01, $p = 0.50$). The strength of evidence for a causal association was considerably weaker in this analysis with wider CIs (OR 1.01, 95% CI 0.93–1.10, $p = 0.815$), 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 0.98–1.01, $p = 0.52$). 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 0.96–1.43, $p = 0.13$). The Egger plots for these analyses are shown in Supplementary Figs S1 and S2.

Table 2 Associations between cannabis initiation and various outcomes and schizophrenia and cannabis initiation using two-sample Mendelian randomization

Exposure	Outcome	OR ^a	95% CI	<i>p</i> value	<i>N</i> SNPs
Cannabis initiation	Schizophrenia (binary outcome)	1.04 ^b	1.01–1.07	0.019	21
Cannabis initiation	Education (years)	0.99 ^b	0.97–1.00	0.066	12
Cannabis initiation	Height (z score)	1.00 ^b	0.99–1.01	0.901	12
Cannabis initiation	Schizophrenia (binary outcome)	1.03 ^b	0.99–1.08	0.145	12 ^c
Schizophrenia	Cannabis initiation (binary outcome)	1.10	1.05–1.14	2.64×10^{-5}	107

OR, Odds ratio; CI, confidence interval; SNP, single nucleotide polymorphism.

^a Per doubling in odds of exposure.

^b Correlated likelihood method used.

^c Restricted number of SNPs to be comparable to positive/negative control analyses.

Table 3 MR-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 ^a	95% CI	<i>P</i>	<i>N</i> SNPs
Cannabis initiation	Schizophrenia (binary outcome)	Intercept	0.99 ^b	0.98–1.01	0.50	21
		Causal estimate	1.01 ^b	0.93–1.10	0.815	
Cannabis initiation	Education (years)	Intercept	0.99 ^b	0.99–1.00	0.182	12
		Causal estimate	0.94 ^b	0.88–1.00	0.062	
Cannabis initiation	Height (z score)	Intercept	0.99 ^b	0.98–1.00	0.093	12
		Causal estimate	0.95 ^b	0.88–1.01	0.095	
Cannabis initiation	Schizophrenia (binary outcome)	Intercept	1.00 ^b	0.97–1.03	0.801	12 ^c
		Causal estimate	1.01 ^b	0.88–1.17	0.854	
Schizophrenia	Cannabis initiation (binary outcome)	Intercept	1.00	0.98–1.01	0.524	107
		Causal estimate	1.17	0.96–1.43	0.128	

OR, Odds ratio; CI, confidence interval; SNP, single nucleotide polymorphism.

^a Per doubling in odds of exposure.

^b Correlated likelihood method used.

^c Restricted number of SNPs to be comparable to positive/negative control analyses.

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 0.97–1.00, $p=0.18$; height: 0.99, 95% CI 0.98–1.00, $p=0.093$). 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 0.88–1.00, $p=0.062$). 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 0.88–1.01, $p=0.095$), 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 CIs were wider (OR 1.03, 95% CI 0.99–1.08, $p=0.15$). 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 Fig. S3).

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 CI 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 & 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 (although 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. Utilizing 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. First, none of the SNPs used to predict cannabis initiation reached genome-wide 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 genome-wide significance in the cannabis initiation GWAS, they were nevertheless associated with initiation at $p < 10^{-5}$. Inferences and estimates from MR analyses 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 1500 individuals (T. 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 three SNPs at genome-wide significance in a sample of 14 754 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. Moreover, 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 most 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 & 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. While 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. While 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.

Supplementary material

The supplementary material for this article can be found at <https://doi.org/10.1017/S0033291716003172>.

Acknowledgements

S.H.G. and M.R.M. 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 thank Dr Tõnu Esko from the EGCUT for his assistance in obtaining the PGC2 data without the EGCUT sample.

Declaration of Interest

None.

References

- Bacanu SA, Chen X, Kendler KS (2013). The genetic overlap between schizophrenia and height. *Schizophrenia Research* **151**, 226–228.
- 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 NM, Thompson SG (2016). Bias due to participant overlap in two-sample Mendelian randomization. *Genetic Epidemiology* **40**, 597–608.
- Burgess S, Dudbridge F, Thompson SG (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 RA, Timpson NJ, Davey Smith G, Thompson SG, Consortium E-I (2015b). Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors. *European Journal of Epidemiology* **30**, 543–552.
- Callaghan RC, Cunningham JK, Allebeck P, Arenovich T, Sajeev G, Remington G, Boileau I, Kish SJ (2012). Methamphetamine use and schizophrenia: a population-based cohort study in California. *American Journal of Psychiatry* **169**, 389–396.
- Corbin LJ, Richmond RC, Wade KH, Burgess S, Bowden J, Smith GD, Timpson NJ (2016). BMI as a Modifiable Risk Factor for Type 2 Diabetes: refining and understanding causal estimates using Mendelian randomization. *Diabetes* **65**, 3002–3007.
- 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–R98.
- D'souza DC, Perry E, Macdougall L, Ammerman Y, Cooper T, Wu YT, Braley G, Gueorguieva R, Krystal JH (2004). The psychotomimetic effects of intravenous delta-9-tetrahydrocannabinol in healthy individuals: implications for psychosis. *Neuropsychopharmacology* **29**, 1558–1572.
- Eastwood P (2013). *Statistics on Drug Misuse: England 2013*. Health and Social Care Information Centre: Leeds.
- Gage SH, Davey Smith G, Munafò M (2015). Comment on PMID 24957864: Genetic Predisposition to Schizophrenia Associated with Increased Use of Cannabis. National Library of Medicine (<https://www.ncbi.nlm.nih.gov/pubmed/24957864>).
- Gage SH, Hickman M, Heron J, Munafò 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 SH, Hickman M, Zammit S (2016a). Association between cannabis and psychosis: epidemiologic evidence. *Biological Psychiatry* **79**, 549–556.
- Gage SH, Munafò MR, Davey Smith G (2016b). Causal inference in developmental origins of health and disease (dohad) research. *Annual Review of Psychology* **67**, 567–585.
- Gage SH, 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 MA, Henningfield JE, Cone EJ (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**, 276–282.
- Iseger TA, Bossong MG (2015). A systematic review of the antipsychotic properties of cannabidiol in humans. *Schizophrenia Research* **162**, 153–161.
- Jackson NJ, Isen JD, Khoddam R, Irons D, Tuvblad C, Iacono WG, McGue M, Raine A, Baker LA (2016). Impact of adolescent marijuana use on intelligence: results from two longitudinal twin studies. *Proceedings of the National Academy of Sciences USA* **113**, E500–E508.
- Jones HJ, Stergiakouli E, Tansey KE, Hubbard L, Heron J, Cannon M, Holmans P, Lewis G, Linden De J, Jones PB, Davey Smith G, O'Donovan MC, Owen MJ, Walter JT, Zammit S (2016). Phenotypic manifestation of genetic risk for schizophrenia during adolescence in the general population. *JAMA Psychiatry* **73**, 221–228.
- Linscott RJ, Van OSJ (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**, 1133–1149.
- Meier MH, Caspi A, Ambler A, Harrington H, Houts R, Keefe RS, McDonald K, Ward A, Poulton R, Moffitt TE (2012). Persistent cannabis users show neuropsychological

- decline from childhood to midlife. *Proceedings of the National Academy of Sciences USA* 109, E2657–E2664.
- Mokrysz C, Landy R, Gage SH, Munafo MR, Roiser JP, Curran HV (2016). Are IQ and educational outcomes in teenagers related to their cannabis use? A prospective cohort study. *Journal of Psychopharmacology* 30, 159–168.
- Moore TH, Zammit S, Lingford-Hughes A, Barnes TR, Jones PB, Burke M, Lewis G (2007). Cannabis use and risk of psychotic or affective mental health outcomes: a systematic review. *Lancet* 370, 319–328.
- Nordsletten AE, Larsson H, Crowley JJ, Almqvist C, Lichtenstein P, Mataix-Cols D (2016). Patterns of nonrandom mating within and across 11 major psychiatric disorders. *JAMA Psychiatry* 73, 354–361.
- Power RA, Verweij KJ, Zuhair M, Montgomery GW, Henders AK, Heath AC, Madden PA, Medland SE, Wray NR, Martin NG (2014). Genetic predisposition to schizophrenia associated with increased use of cannabis. *Molecular Psychiatry* 19, 1201–1204.
- Rietveld CA, Medland SE, Derringer J, Yang J, Esko T, Martin NW, Westra HJ, Shakhbazov K, Abdellaoui A, Agrawal A, Albrecht E, Alizadeh BZ, Amin N, Barnard J, Baumeister SE, Benke KS, Bielak LF, Boatman JA, Boyle PA, Davies G, De Leeuw C, Eklund N, Evans DS, Ferhmann R, Fischer K, Gieger C, Gjessing HK, Hagg S, Harris JR, Hayward C, Holzapfel C, Ibrahim-Verbaas CA, Ingelsson E, Jacobsson B, Joshi PK, Jugessur A, Kaakinen M, Kanoni S, Karjalainen J, Kolcic I, Kristiansson K, Kutalik Z, Lahti J, Lee SH, Lin P, Lind PA, Liu Y, Lohman K, Loitfelder M, McMahon G, Vidal PM, Meirelles O, Milani L, Myhre R, Nuotio ML, Oldmeadow CJ, Petrovic KE, Peyrot WJ, Polasek O, Quaye L, Reinmaa E, Rice JP, Rizzi TS, Schmidt H, Schmidt R, Smith AV, Smith JA, Tanaka T, Terracciano A, Van Der Loos MJ, Vitart V, Volzke H, Wellmann J, Yu L, Zhao W, Allik J, Attia JR, Bandinelli S, Bastardot F, Beauchamp J, Bennett DA, Berger K, Bierut LJ, Boomsma DI, Bultmann U, Campbell H, Chabris CF, Cherkas L, Chung MK, Cucca F, De Andrade M, De Jager PL, De Neve JE, Deary IJ, Dedoussis GV, Deloukas P, Dimitriou M, Eiriksdottir G, Elderson MF, Eriksson JG, Evans DM, Faul JD, Ferrucci L, Garcia ME, Grönberg H, Guðnason V, Hall P, Harris JM, Harris TB, Hastie ND, Heath AC, Hernandez DG, Hoffmann W, Hofman A, Holle R, Holliday EG, Hottenga JJ, Iacono WG, Illig T, Järvelin MR, Kähönen M, Kaprio J, Kirkpatrick RM, Kowgier M, Latvala A, Launer LJ, Lawlor DA, Lehtimäki T, Li J, Lichtenstein P, Lichtner P, Liewald DC, Madden PA, Magnusson PK, Mäkinen TE, Masala M, McGue M, Metspalu A, Mielck A, Miller MB, Montgomery GW, Mukherjee S, Nyholt DR, Oostra BA, Palmer LJ, Palotie A, Penninx BW, Perola M, Peyser PA, Preisig M, Rääkkönen K, Raitakari OT, Realo A, Ring SM, Ripatti S, Rivadeneira F, Rudan I, Rustichini A, Salomaa V, Sarin AP, Schlessinger D, Scott RJ, Snieder H, St Pourcain B, Starr JM, Sul JH, Surakka I, Svento R, Teumer A, LifeLines Cohort Study, Tiemeier H, van Rooij FJ, Van Wagoner DR, Vartiainen E, Viikari J, Vollenweider P, Vonk JM, Waeber G, Weir DR, Wichmann HE, Widen E, Willemsen G, Wilson JF, Wright AF, Conley D, Davey Smith G, Franke L, Groenen PJ, Hofman A, Johannesson M, Kardia SL, Krueger RF, Laibson D, Martin NG, Meyer MN, Posthuma D, Thurik AR, Timpson NJ, Uitterlinden AG, van Duijn CM, Visscher PM, Benjamin DJ, Cesarini D, Koellinger PD (2013). GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science* 340, 1467–1471.
- Rosler W, Hengartner MP, Angst J, Ajdacic-Gross V (2012). Linking substance use with symptoms of subclinical psychosis in a community cohort over 30 years. *Addiction* 107, 1174–1184.
- 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 LA, Gelernter J (2016). Genome-wide association study of cannabis dependence severity, novel risk variants, and shared genetic risks. *JAMA Psychiatry* 73, 472–480.
- Staley JR, Blackshaw J, Kamat MA, Ellis S, Surendran P, Sun BB, Paul DS, Freitag D, Burgess S, Danesh J, Young R, Butterworth AS (2016). PhenoScanner: a database of human genotype-phenotype associations. *Bioinformatics* 32, 3207–3209.
- Stringer S, Minica CC, Verweij KJ, Mbarek H, Walters R, Abdellaoui A, Bruinenberg M, Derringer J, Van Eijk KR (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 KJ, Abdellaoui A, Nivard MG, Sainz Cort A, Ligthart L, Draisma HHM, Minica CC, Consortium IC, Gillespie NA, Hottenga JJ, Boomsma DI, Vink JM (in press). Brief Report. Genetic association between schizophrenia and cannabis use. *Drug and Alcohol Dependence*.
- Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, Chu AY, Estrada K, Luan J, Kutalik Z, Amin N, Buchkovich ML, Croteau-Chonka DC, Day FR, Duan Y, Fall T, Fehrmann R, Ferreira T, Jackson AU, Karjalainen J, Lo KS, Locke AE, Magi R, Mihailov E, Porcu E, Randall JC, Scherag A, Vinkhuyzen AA, Westra HJ, Winkler TW, Workalemahu T, Zhao JH, Absher D, Albrecht E, Anderson D, Baron J, Beekman M, Demirkan A, Ehret GB, Feenstra B, Feitosa MF, Fischer K, Fraser RM, Goel A, Gong J, Justice AE, Kanoni S, Kleber ME, Kristiansson K, Lim U, Lotay V, Lui JC, Mangino M, Mateo Leach I, Medina-Gomez C, Nalls MA, Nyholt DR, Palmer CD, Pasko D, Pechlivanis S, Prokopenko I, Ried JS, Ripke S, Shungin D, Stancakova A, Strawbridge RJ, Sung YJ, Tanaka T, Teumer A, Trompet S, Van Der Laan SW, Van Setten J, Van Vliet-Ostaptchouk JV, Wang Z, Yengo L, Zhang W, Afzal U, Arnlov J, Arscott GM, Bandinelli S, Barrett A, Bellis C, Bennett AJ, Berne C, Bluher M, Bolton JL, Botthcher Y, Boyd HA, Bruinenberg M, Buckley BM, Buyske S, Caspersen IH, Chines PS, Clarke R, Claudi-Boehm S, Cooper M, Daw EW, De Jong PA, Deelen J, Delgado G, Denny JC, Dhonukshe-Rutten R, Dimitriou M, Doney AS, Dörr M, Eklund N, Eury E, Folkersen L, Garcia ME, Geller F, Giedraitis V, Go AS,

- Grallert H, Grammer TB, Gräßler J, Grönberg H, de Groot LC, Groves CJ, Haessler J, Hall P, Haller T, Hallmans G, Hannemann A, Hartman CA, Hassinen M, Hayward C, Heard-Costa NL, Helmer Q, Hemani G, Henders AK, Hillege HL, Hlatky MA, Hoffmann W, Hoffmann P, Holmen O, Houwing-Duistermaat JJ, Illig T, Isaacs A, James AL, Jeff J, Johansen B, Johansson Å, Jolley J, Juliusdottir T, Junttila J, Kho AN, Kinnunen L, Klopp N, Kocher T, Kratzer W, Lichtner P, Lind L, Lindström J, Lobbens S, Lorentzon M, Lu Y, Lyssenko V, Magnusson PK, Mahajan A, Maillard M, McArdle WL, McKenzie CA, McLachlan S, McLaren PJ, Menni C, Merger S, Milani L, Moayyeri A, Monda KL, Morken MA, Müller G, Müller-Nurasyid M, Musk AW, Narisu N, Nauck M, Nolte IM, Nöthen MM, Oozageer L, Pilz S, Rayner NW, Renstrom F, Robertson NR, Rose LM, Roussel R, Sanna S, Scharnagl H, Scholtens S, Schumacher FR, Schunkert H, Scott RA, Sehmi J, Seufferlein T, Shi J, Silventoinen K, Smit JH, Smith AV, Smolonska J, Stanton AV, Stirrups K, Stott DJ, Stringham HM, Sundström J, Swertz MA, Syvänen AC, Tayo BO, Thorleifsson G, Tyrer JP, van Dijk S, van Schoor NM, van der Velde N, van Heemst D, van Oort FV, Vermeulen SH, Verweij N, Vonk JM, Waite LL, Waldenberger M, Wennauer R, Wilkens LR, Willenborg C, Wilsgaard T, Wojczynski MK, Wong A, Wright AF, Zhang Q, Arveiler D, Bakker SJ, Beilby J, Bergman RN, Bergmann S, Biffar R, Blangero J, Boomsma DI, Bornstein SR, Bovet P, Brambilla P, Brown MJ, Campbell H, Caulfield MJ, Chakravarti A, Collins R, Collins FS, Crawford DC, Cupples LA, Danesh J, de Faire U, den Ruijter HM, Erbel R, Erdmann J, Eriksson JG, Farrall M, Ferrannini E, Ferrières J, Ford I, Forouhi NG, Forrester T, Gansevoort RT, Gejman PV, Gieger C, Golay A, Gottesman O, Gudnason V, Gyllenstein U, Haas DW, Hall AS, Harris TB, Hattersley AT, Heath AC, Hengstenberg C, Hicks AA, Hindorf LA, Hingorani AD, Hofman A, Hovingh GK, Humphries SE, Hunt SC, Hypponen E, Jacobs KB, Jarvelin MR, Jousilahti P, Jula AM, Kaprio J, Kastelein JJ, Kayser M, Kee F, Keinanen-Kiukaanniemi SM, Kiemeny LA, Kooner JS, Kooperberg C, Koskinen S, Kovacs P, Kraja AT, Kumari M, Kuusisto J, Lakka TA, Langenberg C, Le Marchand L, Lehtimäki T, Lupoli S, Madden PA, Männistö S, Manunta P, Marette A, Matisse TC, McKnight B, Meitinger T, Moll FL, Montgomery GW, Morris AD, Morris AP, Murray JC, Nelis M, Ohlsson C, Oldehinkel AJ, Ong KK, Ouwehand WH, Pasterkamp G, Peters A, Pramstaller PP, Price JF, Qi L, Raitakari OT, Rankinen T, Rao DC, Rice TK, Ritchie M, Rudan I, Salomaa V, Samani NJ, Saramies J, Sarzynski MA, Schwarz PE, Sebert S, Sever P, Shuldiner AR, Sinisalo J, Steinthorsdottir V, Stolk RP, Tardif JC, Tönjes A, Tremblay A, Tremoli E, Virtamo J, Vohl MC; Electronic Medical Records and Genomics (eMEMERGE) Consortium.; MIGen Consortium.; PAGEGE Consortium.; LifeLines Cohort Study., Amouyel P, Asselbergs FW, Assimes TL, Bochud M, Boehm BO, Boerwinkle E, Bottinger EP, Bouchard C, Cauchi S, Chambers JC, Chanock SJ, Cooper RS, de Bakker PI, Dedoussis G, Ferrucci L, Franks PW, Froguel P, Groop LC, Haiman CA, Hamsten A, Hayes MG, Hui J, Hunter DJ, Hveem K, Jukema JW, Kaplan RC, Kivimäki M, Kuh D, Laakso M, Liu Y, Martin NG, März W, Melbye M, Moebus S, Munroe PB, Njølstad I, Oostra BA, Palmer CN, Pedersen NL, Perola M, Pérusse L, Peters U, Powell JE, Power C, Quertermous T, Rauramaa R, Reinmaa E, Ridker PM, Rivadeneira F, Rotter JI, Saaristo TE, Saleheen D, Schlessinger D, Slagboom PE, Snieder H, Spector TD, Strauch K, Stumvoll M, Tuomilehto J, Uusitupa M, van der Harst P, Völzke H, Walker M, Wareham NJ, Watkins H, Wichmann HE, Wilson JF, Zanen P, Deloukas P, Heid IM, Lindgren CM, Mohlke KL, Speliotes EK, Thorsteinsdottir U, Barroso I, Fox CS, North KE, Strachan DP, Beckmann JS, Berndt SI, Boehnke M, Borecki IB, McCarthy MI, Metspalu A, Stefansson K, Uitterlinden AG, van Duijn CM, Franke L, Willer CJ, Price AL, Lettre G, Loos RJ, Weedon MN, Ingelsson E, O'Connell JR, Abecasis GR, Chasman DI, Goddard ME, Visscher PM, Hirschhorn JN, Frayling TM. (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genetics* 46, 1173–1186.