DOI:
10.1017/S0033291718000478

Document Version
Peer reviewed version

Overlap of Heritable Influences between Cannabis Use Disorder, Frequency of Use and Opportunity to Use Cannabis: Trivariate Twin Modelling and Implications for Genetic Design

Lindsey A. Hines[1,2], Katherine I. Morley[1,3], Fruhling Rijsdijk[4], John Strang[1], Arpana Agrawal[5], Elliot C. Nelson[5], Dixie Statham[6], Nicholas G. Martin[7] and Michael T. Lynskey[1]

[1]Addictions Department, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, England

[2]Centre for Adolescent Health, Royal Children's Hospital, Murdoch Children Research Institute, Parkville, Victoria, Australia

[3]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Australia.

[4] Social Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, England

[5]Department of Psychiatry, Washington University School of Medicine, St Louis, MO, USA

[6]School of Social Sciences, University of the Sunshine Coast, Queensland, Australia

[7]QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia

Word count: 4499

Abstract

Background

The genetic component of Cannabis Use Disorder (CUD) may overlap with influences acting more generally on early stages of cannabis use. This paper aims to determine the extent to which genetic influences on the development of cannabis abuse/dependence are correlated with those acting on opportunity to use cannabis and frequency of use.

Methods

Cross-sectional study of 3303 Australian twins, measuring age of onset of cannabis use opportunity, lifetime frequency of cannabis use and lifetime DSM-IV cannabis abuse/dependence. A trivariate Cholesky decomposition estimated additive genetic (A), shared environment (C) and unique environment (E) contributions to opportunity to use cannabis, frequency of cannabis use, cannabis abuse/dependence, and the extent of overlap between genetic and environmental factors associated with each phenotype.

Results

Variance components estimates were A=0.64 (95% CI 0.58 – 0.70) and E=0.36 (95% CI 0.29 – 0.42) for age of opportunity to use cannabis, A=0.74 (95% CI 0.66 – 0.80) and E=0.26 (95% CI 0.20 – 0.34) for cannabis use frequency, and A=0.78 (95% CI 0.65 – 0.88) and E=0.22 (95% CI 0.12 – 0.35) for cannabis abuse/dependence. Opportunity shares 45% of genetic influences with frequency of use, and only 17% of additive genetic influences are unique to abuse/dependence from those acting on opportunity and frequency.

Conclusions

There are significant genetic contributions to lifetime cannabis abuse/dependence, but a large proportion of this overlaps with influences acting on opportunity and frequency of use. Individuals without drug use opportunity are uninformative, and studies of drug use disorders must incorporate individual exposure to accurately identify aetiology.

Introduction

As the legislative landscape regarding cannabis alters, potentially altering patterns of use (Hopfer 2014; Hasin *et al.* 2015; Shi *et al.* 2015), a greater understanding of environmental and genetic influences on progression to harmful or disordered cannabis use is needed. Cannabis Use Disorder (CUD) is included in the DSM-5 (American Psychiatric Association & DSM-5 Task Force 2013), an amalgamated update of DSM-IV cannabis abuse and cannabis dependence (American Psychiatric Association 2000) characterised by loss of control over use, failure to fulfil social roles, recurrent use in hazardous situations, and use despite worsening of health problems. An estimated 10% - 16% of individuals who have ever used cannabis will develop dependence (Anthony 2006) and globally 13.1 million individuals meet criteria for cannabis dependence contributing 10.3% of the illicit drug use global burden of disease (Degenhardt *et al.* 2014).

Individuals with drug dependence pass through several intermediate stages before developing a clinical condition, and many non-clinical individuals will reach earlier stages of drug use involvement without progressing to disorder. The earliest stage of involvement is having the opportunity to use (regardless of whether the individual uses the drug or not). Opportunity is required for use to occur, and forms an individual's earliest necessary condition from which they are at risk of developing dependence (Wagner & Anthony 2002). Once initiation of use has occurred, individuals will vary in frequency of cannabis use, with increased frequency associated with increased likelihood for the development of cannabis dependence (Chen *et al.* 1997). Considering the sources of variation in progression through the stages of cannabis use, and the extent to which influences are consist across different stages, can provide insight into the aetiology of CUD (Hines *et al.* 2015a, 2016).

Twin modelling has identified a strong genetic contribution to CUD, with a review of 6 studies in the area concluding heritability estimates range from 45% – 78% (Agrawal & Lynskey 2006). Meta-analysis estimated heritability of problematic cannabis use (having one or more of the symptoms of cannabis abuse or dependence) at 51.4 (95% CI 37.9–64.9) in males and 58.5 (95% CI 44.2–72.9) in females (Verweij *et al.* 2010). However, the

magnitude of these influences may differ across stages of drug use. Early stages may be genetically influenced through personality traits such as novelty seeking (Laucht *et al.* 2007), whereas at subsequent stages, such as drug dependence and development of withdrawal, genetic influences on drug metabolism, may be more influential (Dick *et al.* 2014).

Common genetic influences may act on multiple stages. The majority of research into the correlation of influences between initiation of use and disordered use comes from the alcohol and tobacco literature, where a genetic correlation (0.15 – 0.88) has been consistently demonstrated between the earlier and later stages of drug use (Broms *et al.* 2006; Pagan *et al.* 2006; Morley *et al.* 2007). Similarly, studies of alcohol use disorder have identified a strong genetic correlation between age of alcohol initiation and alcohol use disorder (Sartor *et al.* 2009; Ystrom *et al.* 2014). Similar mechanisms may be acting on CUD. Only 34% of the variance in cannabis abuse/dependence is unique to this phenotype, with the rest shared with genetic influences on initiation (Agrawal *et al.* 2005), and cannabis availability explains almost all the shared environmental risks in cannabis initiation and abuse (Gillespie *et al.* 2009b).

To date, research has not explored the extent to which genetic influences may correlate across more than 2 stages of drug use. Additionally, the heritability of the earliest stage of drug use - having opportunity to use a drug (Wagner & Anthony 2002) – has been somewhat overlooked. This is despite evidence of the importance of this phenotype for design of genetic research (Nelson *et al.* 2013): individuals who do not have opportunity to use a substance are unable to express their genetic vulnerability to later stages, including use and use disorders. Not only are such individuals structurally missing in analytic terms, but excluding individuals who have no drug use opportunity to use from genetic association studies can provide superior control for environmental background and related covariates.

Opportunity may be regarded as a putative environmental factor, likely subject to broader environmental modifications, such as changes in national policy, but also to individual-specific factors, including peer provision of drugs. Despite these underpinnings, such

"environmental" factors have been shown to have heritable variation (Kendler & Baker 2007; Gillespie *et al.* 2009b). Considering this phenotype in the context of later stages of drug transitions, such as escalation to frequent use and the development of abuse/dependence will provide insight into the pathways to the development of dependence.

By applying trivariate twin models to the phenotypes age of cannabis opportunity, frequency of cannabis use, and abuse/dependence, this paper aims to determine the extent to which genetic influences on the development of cannabis abuse/dependence are unique to the phenotype, and the extent to which they correlate with influences on opportunity to use cannabis and the frequency of cannabis use.

Methods

Sample

The sample was drawn from the Australian Twin Registry. From a pool of pairs born 1972-1979, 3348 MZ and DZ twins completed the interview component of a study of cannabis and other drug misuse. A full description of the study methodology and of the characteristics of participants has been published previously (Lynskey *et al.* 2012). The 3303 twins who provided information on whether or not they had ever had the opportunity to use cannabis, and who had complete zygosity information, form the analysis sample for this paper. This sample consisted of 975 MZ males, 481 DZ males, 734 MZ females, 371 DZ females, and 742 opposite sex DZ twins. Of these, 808 were singletons. Mean age was 31.8 (range 27 – 40 years, median 32.0).

Assessment

Participants were assessed through computer-assisted telephone interviews which collected information on socio-demographics, childhood experiences, drug use and common mental health disorders, including cannabis and other drug use disorders, assessed using the Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA-OZ) interview (Bucholz *et al.*

1994; Heath *et al.* 1997). The SSAGA-OZ is a validated measure of mental health using DSM-IV criteria, and includes assessment of cannabis and other drug abuse and dependence. Specific measures used in the current analyses are described below.

Measures

Opportunity to use cannabis

Participants were asked “have you ever been offered, or had the opportunity to use cannabis, even if you didn't use it at the time? How old were you the first time?” Of 3348 twins interviewed, 3325 provided information on whether or not they had ever had the opportunity to use cannabis. Of these twins, information on zygosity was missing for 22, resulting in an analysis sample of 3303.

For analysis, participants were categorised as having never had the opportunity to use cannabis (N = 356, 10.8%), having had later opportunity to use cannabis (first opportunity reported as happening at age 16 and over, N = 2264, 68.5%), or having had early opportunity to use cannabis (first opportunity reported as occurring at age 15 or earlier, N = 670, 20.3%). As there is no precedent in the literature for what age represents an “early” opportunity to use cannabis, sensitivity analyses were conducted on the cut-off age. The correlations obtained by different cut-off points indicated results were not affected by the choice of age 15 as age cut-off for early opportunity (see supplementary material).

Cannabis use frequency

Participants were asked about lifetime frequency of use through the item “have you used marijuana 40 or more times, 21-39 times, 11-20 times, 7-10 times, 1-6 times?”, then estimated number of times used. Participants were categorised as having used cannabis infrequently, at a level that precluded being asked about cannabis abuse/dependence (0 - 11 times, N = 1913), moderately (12 – 50 times, N = 476), or high frequency (50+ times, N=554).

Cannabis abuse/dependence

Participants were classified as meeting DSM-IV criteria for lifetime cannabis abuse if they reported one or more of the following: often using cannabis in a situation where they might get hurt; arrested more than twice within a 12 month period as a result of their cannabis use; cannabis use having caused difficulty with work, study or household responsibilities; cannabis having caused social and interpersonal problems more than 3 times within a 12 month period.

Participants were classified as meeting lifetime criteria for DSM-IV cannabis dependence if they reported 3 or more of the following symptoms occurring within the same 12 month period: using cannabis a greater number of times/greater amount than was intended, tolerance, wanting to cut down/stop use, spending so much time obtaining/using/recovering from the effects of cannabis the participant had little time for anything else, reducing important activities as a result of cannabis use, continuing use despite it worsening health/emotional problems. In the sample used in this analysis, 16.4% (N=543) reported cannabis abuse and/or dependence.

Individual characteristics

Sex

Sex was determined through self-report (76.9% female, N=2540).

Zygosity

Zygosity of twin pairs was measured through standard questions about physical similarity and the extent to which twin identity was confused by parents, teachers and strangers; methods found to give better than 95% agreement with results of genotyping (Cederlof *et al.* 1961; Kasriel & Eaves 1976; Sarna *et al.* 1978).

Statistical analyses

All analyses were conducted using OpenMX v2.5.2 (Boker *et al.* 2011) for the statistical software R v3.1.2 (R Core Team 2013). Analyses used full information maximum-likelihood

estimation with raw data, and the optimiser SLSQP was applied to analyses. Analyses were adjusted for sex.

Staged Trivariate Twin Model

Classical twin modelling estimates the extent to which additive genetic (A), common environment (C) and unique environment (E) influence a phenotype (Neale & Cardon 1992). Approaches using twins reared together can be used to determine the heritability of, and environmental contribution to, a phenotype or trait. Identical – or monozygotic (MZ) – twin pairs share 100% of their genetic material. Fraternal – or dizygotic (DZ) – twin pairs share only 50%, on average, of the same genetic material. This means they are no more genetically alike than full siblings. However, unlike siblings DZ twins will grow up in the same environment. Using this knowledge we can calculate the extent to which the variance in a phenotype is due to genetic effects, and the extent to which it is due to environmental effects (Plomin *et al.* 2013). If the MZ correlation is twice the DZ correlation then all twin-pair similarity can be attributed to A, whereas if the MZ correlation is greater than the DZ correlation, but not twice the DZ correlation, there is also evidence of some shared environmental influences. The extent to which the MZ twin correlation is less than 1.0 indicates the magnitude of non-shared environmental influences. Dominant genetic effects (D), which are non-additive interaction effects between genes, cannot be assessed simultaneously with C (Neale & Cardon 1992). Structural equation modelling of twin data is used to obtain precise estimates of A, C and E and allows for the comparison of models and generation of confidence intervals around estimates (Neale & Cardon 1992).

A staged twin model was fitted to assess contributions of A, C, and E to variance in age of opportunity to use cannabis, frequency of cannabis use, and lifetime cannabis abuse/dependence, and to estimate the extent to which the influences of A, C and E on the three phenotypes were correlated (Heath *et al.* 2002). The staged model is appropriate for situations where early-stage phenotypes, such as cannabis use opportunity, are *necessary* for the expression of later behaviours, such as the development of dependence, and is a

variation of the classic bivariate model appropriate for analysis of variables with data missing at random (data are missing as a result of observations on a previous variable, as opposed to data missing completely at random) (Kendler *et al.* 1999; Heath *et al.* 2002; Neale *et al.* 2006). See Heath et al (2002) for full details. Explicitly modelling such structurally missing data also has the advantage of estimating the extent of covariation between these contingent stages of use (i.e., opportunity, frequency, abuse/dependence) while not excluding those who do not provide information on a prior stage (e.g., opportunity) from analyses of later stages (e.g., abuse/dependence).

A Cholesky decomposition model was used to parse the phenotypic correlations between the three stages of cannabis use and misuse into A, C and E sources, including those specific to each of the latter stages of frequency and abuse/dependence as well as the magnitude of overlapping influences across the 3 stages.

Assumption testing

The analysis assumes each threshold-selected trait has an underlying bivariate/multivariate normal liability distribution. Exploring this methodological issue falls beyond the scope of this paper, but such modelling techniques have been shown to be robust to breaches of this assumption (Reinartz *et al.* 2009). Thresholds represent cut-off points along this unobserved continuous distribution of liability.

In order to test whether thresholds could be equated between MZ and DZ twins, nested models were compared against a saturated twin model. Differences in the fit of more parsimonious models compared to the saturated or ACE model were assessed via the Akaike Information Criterion (AIC) and the change in -2loglikelihood (Δ-2LL), which can be approximated by a chi square distribution with degrees of freedom (DF) equal to the difference in degrees of freedom of the nested models. Where these measures lead to different conclusions on parsimony, the p value has been prioritised. Significance of thresholds (and equality between thresholds) was determined by Δ-2LL and change in DF

(ΔDF) and associated chi-square distribution. Significance of variance and covariance paths was similarly determined through likelihood ratio testing.

Results

Prevalence of, and Correlations between, Opportunity to use Cannabis, Frequency of Cannabis Use and Abuse/dependence

Of those who reported opportunity to use cannabis by age 15 (N=683), 35.8% (N=244) reported high frequency cannabis use (lifetime use 50+ times), compared to 13.7% (N= 310) of those who reported cannabis use opportunity at age 16 or older (N=2264). Of those who reported high frequency cannabis use (50+ times, N=554), 75.6% (N =418) met criteria for lifetime cannabis abuse/dependence compared to 26.3% (N =125) of those who reported lower frequency cannabis use (12 – 50 times, N = 476).

A saturated twin model was used to estimate tetrachoric correlations for the categorically-defined traits of age of opportunity, frequency of cannabis use and lifetime cannabis abuse/dependence (see Table 1). The relative magnitude of MZ within-trait correlations indicate heritable influences on all of these traits. The across twin/across trait correlations and confidence intervals indicate genetic factors contribute to all correlations. MZ within trait and across trait correlations are not twice the DZ correlations, suggesting some influence of C. All correlations are less than 1.0, suggesting moderate to low effects of E.

Assumption Testing

MZ and DZ thresholds could not be equated (Δ-2LL=15.0, ΔDF=5, P=0.01), and were estimated separately in all further models.

Trivariate Cholesky Model Fitting

A saturated model provided fit statistics, estimates for each component of the variance for all three phenotypes, and estimates for the covariance between phenotypes. The fit statistics for this model were -2LL=11029.68 DF=7249, AIC=--3468.32.

Nesting Models to Develop Parsimonious Model Fit

In order to identify the most parsimonious model, nested models constrained individual variance and covariance components to zero, when confidence intervals on the estimate from the saturated model included 0. It was possible to drop all C parameters (Δ-2LL=6.07, ΔDF=6, P value=0.41) without a significant decrement in fit. In addition, there was no statistically significant covariance between opportunity and either frequency or abuse/dependence attributable to E (Δ-2LL=0.58, ΔDF=2, P value=0.75).

Final Model

The final most parsimonious model was an AE model (Δ-2LL=7.22, ΔDF=8, P value=0.51). Variance component estimates are presented in Table 2. Approximately 64-78% of the variance in each phenotype was due to additive genetic influences, with confidence intervals indicating both frequency and abuse/dependence were modestly, but significantly, more heritable than opportunity to use. A proportion of these genetic influences were shared across the three stages. As shown in Table 2, genetic correlations across stages ranged from 0.37 (opportunity and abuse/dependence) to 0.68 (frequency and abuse/dependence). For frequency, about 55% of the genetic influences were unique from those acting on opportunity, while for cannabis abuse/dependence, 17% of the genetic influences were unique from those acting on opportunity and frequency of use. In addition, cannabis abuse/dependence shared individual-specific environmental influences with frequency (but not opportunity) with 27% specific to this stage.

Discussion

Additive genetic influences determine the majority of variance in age of opportunity to use cannabis (0.64, 95% CI 0.58 – 0.70), frequency of cannabis use (0.74, 95% 0.66 – 0.80), and cannabis abuse/dependence (0.78, 95% 0.65 – 0.88). Of these influences, 55% of additive genetic influences acting on frequency of cannabis use are unique from those acting on age of opportunity to use cannabis, and 17% of additive genetic influences acting on cannabis abuse/dependence are unique from those acting on opportunity and frequency. No significant effect of the shared environment was observed, but there were unique environmental influences on all phenotypes. The only correlated unique environmental influences were between cannabis use frequency and abuse/dependence.

Previous research has not explored the correlation between influences on cannabis use opportunity and cannabis abuse or dependence, although existing studies focusing on cannabis initiation observed overlapping liabilities between cannabis initiation and progression to heavy use (0.88; 33% due to genetic factors) (Fowler *et al.* 2007). This is a similar genetic contribution to the overlap in liabilities to that presently observed between cannabis opportunity and frequency of use. This demonstrates the present findings are in line with existing research showing genetic correlation between the early stages of cannabis use and later substance use disorders.

Opportunity to use cannabis is the necessary first step in progression towards problematic use, and this phenotype could be expected to be subject only to environmental influence. However, 64% of the variance in cannabis age at opportunity was due to genetic factors. Although it may be surprising that an apparently environmental phenotype is influenced by heritable factors, this result is consistent with previous findings that cannabis use availability (Gillespie *et al.* 2009b) and other putative measures of 'environment' (Kendler & Baker 2007) are, in fact, influenced by genetic factors. Environmental measures can be heritable if there is a bidirectional relationship between an individual's behaviour and their environment, if aspects of behaviour are subject to genetic influences (Kendler & Baker 2007; Lynskey & Agrawal 2009). A review of this area identified positive and negative life

events, divorce and social support all have heritable influences (Kendler & Baker 2007). The additive genetic correlation may also indicate evocative or active interactions taking place (Plomin *et al.* 2013), with genes influencing earlier age of cannabis use opportunity contributing to individuals selecting into environments and behaviours that facilitate the development of cannabis dependence.

Alternatively, genetic influences associated with other behaviours may be influencing progression through the stages of cannabis use. Previous research has identified conduct disorder influences transitions to cannabis use opportunity, and from opportunity to dependence (Hines *et al.* 2016). This is in line with existing research demonstrating the consistent influence of conduct disorder on drug use (Lynskey *et al.* 2002; Storr *et al.* 2011; Reboussin *et al.* 2015), and genes relating to conduct disorder and involvement with deviant peers (Gillespie *et al.* 2009a) are plausible candidates for the shared genetic liability between age of opportunity and the development of cannabis abuse/dependence. Additionally, personality factors associated with drug use (Malmberg *et al.* 2010), such as sensation seeking, may underlie this shared genetic liability.

Cannabis opportunity, frequency of use, and abuse/dependence show a moderate effect of the unique environment (0.35, 0.26 and 0.22, respectively), but the correlation between unique environmental influences on opportunity and the later stages of drug use was non-significant. This may reflect measurement error (Plomin *et al.* 2013), but is in line with existing research demonstrating the pattern of environmental factors associated with progression between specific stages of drug use differs between transitions (Sartor *et al.* 2007; Belsky DW *et al.* 2013; Hines *et al.* 2016). For example, childhood and early adolescent factors have been shown to be uniquely associated with cannabis opportunity, whereas escalating other drug use factors is uniquely associated with development of cannabis dependence (Hines *et al.* 2016).

The present analysis indicated none of the observed variance in opportunity to use cannabis, frequency of use or abuse/dependence in males was attributable to the shared environment in this sample. The shared environment is usually found to be more important at earlier stages than later (Fowler *et al.* 2007), and these findings contradict findings of a high shared environmental correlation between cannabis availability and cannabis abuse (Gillespie *et al.* 2009b). The samples differ, with the Gillespie *et al.* findings based on an all-male population, but these contradictory findings indicate cannabis availability (the perceived ease of obtaining cannabis) and opportunity (having been offered cannabis, or being around cannabis use) represent different phenotypes.

Previous research has not tested the extent to which genetic influences on cannabis initiation and cannabis abuse overlap, so comparisons cannot be made to the present findings for opportunity and abuse/dependence. However, when considered in light of findings that variation in progression to subsequent use of cannabis is almost entirely attributable to the unique environment (Hines *et al.* 2015b), a picture is beginning to emerge of how different factors influence progression from the very earliest stages of cannabis to the development of dependence.

Implications

The potential for opportunity to use cannabis to be a marker for intervention has previously been discussed (Neumark *et al.* 2012), and the overlap in genetic influences between age of opportunity and both frequency of cannabis use and cannabis abuse/dependence indicates there is potential to use this measure to indicate those at greatest risk of developing later frequent and/ or problematic use. It has previously been suggested that prevention strategies focused on modifying beliefs, norms and behavioural patterns within close social networks may be effective at reducing drug use opportunity, and consequently drug use (Neumark *et al.* 2012). The identified moderate influence of unique environmental factors on all phenotypes indicates there is scope to determine further influences which may be amenable to target within intervention efforts.

The findings of this paper have important implications for future studies of gene variants and heritability of problematic cannabis use, and in the choice of controls in case-control studies. These results indicate only a moderate proportion of genetic influences on cannabis abuse/dependence are unique from those acting on age of opportunity to use cannabis. These findings reflect previous research demonstrating the importance of considering drug use opportunity when looking at the genetics of opiate use (Nelson *et al.* 2013). Comparison of participants in treatment for opiate dependence with nondependent neighbourhood controls (high exposure to illicit drugs, either via use or from residing in environments with widespread drug availability) identified SNPs in *ANKK1* and *TTC12* as associated with heroin dependence, whereas comparison with controls sourced from the ATR (individuals not dependent on alcohol or illicit drugs, with significantly lower illicit drug exposure) found no association with these SNPs (Nelson *et al.* 2013). Until now the importance of considering cannabis use opportunity in genetic studies has not been explored, although some studies remove those who have not initiated use. Removing those who have not initiated cannabis use can reduce sample size and power, and the present results indicate excluding those without opportunity may avoid conflating genetic influences whilst retaining a greater proportion of a sample. A further advantage of incorporating opportunity to use may arise in meta-analyses of genomewide association studies (GWAS) of cannabis use and misuse. Marked regional variation in opportunity to use across different samples may comprise an international meta-analytic effort. Exclusion of, or accounting for, variability in exposure opportunity, even using crude indices of national policy or cannabis-related law, might reduce heterogeneity in the extent to which genetic vulnerability to later stages of cannabis problems have been adequately expressed.

Consequently, a key implication of the current findings is the necessity of taking into consideration the stage of drug use reached amongst the controls for genomic analyses. Existing research has utilised information on the extent of cannabis use in controls (e.g. excluding those who had used cannabis fewer than 6 times) (Hartman *et al.* 2009), but such

issues are not always taken into consideration (Benyamina *et al.* 2009). This may be especially important in studies of cannabis; a drug with high prevalence of use, but relatively low prevalence of dependence amongst lifetime users. As the legal status of cannabis changes (Shi *et al.* 2015) availability may become to be comparable to that of alcohol, but individual opportunity to use may remain variable. Depending on the research question, and on the development of research identifying genetic overlap between progression to other stages of cannabis use and problematic cannabis use, screening controls not only for opportunity or initiation of cannabis use, but also for frequency of use may have utility in improving cannabis dependence SNP identification in the future.

These findings have further implications for the overlap of genetic influences across drug classes. Existing research has suggested a proportion of the genetic factors underlying SUDs are not specific to individual drugs, and environmental influences determine the drug of misuse (Kendler *et al.* 2003) However, previous research in this area has not incorporated consideration of the stage sequential nature of drug dependence into their analyses. Much of the non-specificity of genetic influences on SUDs likely results from shared influences on the earlier stages of drug use, with more specific influences (such as those related to drug metabolism, for example) associated with later stages of use.

Limitations

Certain limitations must be taken into account when interpreting these results. The data are based on retrospective self-report. Retrospective recall of age onset of drug use behaviours has been shown to be reliable (Shillington *et al.* 1995; Johnson & Mott 2001; Parra *et al.* 2003; Ensminger *et al.* 2007), but the analyses would benefit from replication in prospective longitudinal cohorts. Self-report has been shown to be a valid measure of data collection relating to drug use (Darke 1998), and has been described as the gold standard for collecting data on phenotypes such as initiation and opportunity (Wagner & Anthony 2002). Given use of cannabis was illegal at time of data collection, some participants in this study

may have misreported their drug use. However, the high prevalence of self-reported lifetime cannabis use (68.5%) suggest it's unlikely this was an issue.

The results are based on a twin population. Research has demonstrated twin and non-twin populations do not differ in incidence of psychiatric illness (Kendler *et al.* 1996), and no association has been found between twin environmental similarity and mental health outcomes (Kendler *et al.* 1993).

Conclusions

There are significant genetic contributions to lifetime cannabis abuse/dependence, but a proportion of this overlaps with genetic influences acting on the opportunity to use cannabis and the frequency of cannabis use. Individuals without drug use opportunity are uninformative, and studies of drug use disorder and frequency of use, whether focused on identifying gene variants or environmental factors, must incorporate consideration of drug use exposure use amongst controls in order to accurately identify aetiological factors.

Acknowledgements

This research was funded by National Institute on Drug Abuse (NIDA) grants DA18267, DA23668 & DA032573 and facilitated through access to the Australian Twin Registry. Twins Research Australia receives support from the National Health and Medical Research Council through a Centre of Research Excellence Grant, which is administered by the University of Melbourne.

Declarations of interest

AA has previously received peer-reviewed funding from ABMRF/Foundation for Alcohol Research which receives partial support from the brewing industry.

JS is a researcher and clinician and has worked with a range of types of treatment and rehabilitation service-providers. He has also worked with pharmaceutical companies to seek to identify new or improved treatments, and also with a range of governmental and non-

governmental organisations. His employer (King's College London) is registering intellectual property on an innovative medication development with which JS is involved (not relevant to cannabis), and JS has been named in a patent registration by a Pharma company as inventor of a potential novel overdose resuscitation product (not relevant to cannabis). A fuller account of JS's interests is on his personal web-page of the Addictions Department at http://www.kcl.ac.uk/ioppn/depts/addictions/people/hod.aspx . JS is also supported by the National Institute for Health Research (NIHR) Biomedical Research Centre for Mental Health at South London and Maudsley NHS Foundation Trust and King's College London.

There are no other declarations of interest from authors of this paper.

References

Agrawal A, Lynskey MT (2006). The genetic epidemiology of cannabis use, abuse and dependence. *Addiction* **101**, 801–812.

Agrawal A, Neale MC, Jacobson KC, Prescott CA, Kendler KS (2005). Illicit drug use and abuse/dependence: modeling of two-stage variables using the CCC approach. *Addictive Behaviors* **30**, 1043–1048.

American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders: DSM-IV-TR*. 4th ed., text revision. American Psychiatric Association: Washington, DC.

American Psychiatric Association, DSM-5 Task Force (2013). *Diagnostic and statistical manual of mental disorders: DSM-5.*

Anthony JC (2006). The Epidemiology of Cannabis Dependence. In *Cannabis Dependence: Its Nature, Consequences, and Treatment* Eds RA Roffman & RS Stephens, pp58–105. Cambridge University Press: New York.

Belsky DW, Moffitt TE, Baker TB, et al (2013). Polygenic risk and the developmental progression to heavy, persistent smoking and nicotine dependence: Evidence from a 4-decade longitudinal study. *JAMA Psychiatry* **70**, 534–542.

Benyamina A, Bonhomme-Faivre L, Picard V, Sabbagh A, Richard D, Blecha L, Rahioui H, Karila L, Lukasiewicz M, Picard V, Marill C, Reynaud M (2009). Association between ABCB1 C3435T polymorphism and increased risk of cannabis dependence. *Progress in Neuro-Psychopharmacology and Biological Psychiatry* **33**, 1270–1274.

Boker SM, Neale MC, Maes HH, Wilde MJ, Spiegel M, Brick TR, Spies J, Estabrook R, Kenny S, Bates TC, Mehta P, Fox J (2011). OpenMx: An Open Source Extended Structural Equation Modeling Framework. *Psychometrika* **76**, 306–317.

Broms U, Silventoinen K, Madden PAF, Heath AC, Kaprio J (2006). Genetic Architecture of Smoking Behavior: A Study of Finnish Adult Twins. *Twin Research and Human Genetics* **9**, 64–72.

Bucholz KK, Cadoret R, Cloninger CR, Dinwiddie SH, Hesselbrock VM, Nurnberger J, Reich T, Schmidt I, Schuckit MA (1994). A New, Semi-Structured Psychiatric Interview for Use in Genetic Linkage Studies: A Report on the Reliability of the SSAGA. *Journal of Studies on Alcohol and Drugs* **55**, 149.

Cederlof R, Friberg L, Jonsson E, Kaij L (1961). Studies on similarity diagnosis in twins with the aid of mailed questionnaires. *Acta Genetica Et Statistica Medica* **11**, 338–362.

Chen K, Kandel DB, Davies M (1997). Relationships between frequency and quantity of marijuana use and last year proxy dependence among adolescents and adults in the United States1. *Drug and Alcohol Dependence* **46**, 53–67.

Darke S (1998). Self-report among injecting drug users: A review. *Drug and Alcohol Dependence* **51**, 253–263.

Degenhardt L, Whiteford H, Hall WD (2014). The Global Burden of Disease projects: What have we learned about illicit drug use and dependence and their contribution to the global burden of disease? *Drug and Alcohol Review* **33**, 4–12.

Dick DM, Cho SB, Latendresse SJ, Aliev F, Nurnberger JI, Edenberg HJ, Schuckit M, Hesselbrock VM, Porjesz B, Bucholz K, Wang J-C, Goate A, Kramer JR, Kuperman S (2014). Genetic influences on alcohol use across stages of development: GABRA2 and longitudinal trajectories of drunkenness from adolescence to young adulthood. *Addiction Biology* **19**, 1055–1064.

Ensminger ME, Juon H-S, Green KM (2007). Consistency between adolescent reports and adult retrospective reports of adolescent marijuana use: Explanations of inconsistent reporting among an African American population. *Drug & Alcohol Dependence* **89**, 13–23.

Fowler T, Lifford K, Shelton K, Rice F, Thapar A, Neale MC, McBride A, van den Bree MBM (2007). Exploring the relationship between genetic and environmental influences on initiation and progression of substance use. *Addiction* **102**, 413–422.

Gillespie NA, Neale MC, Jacobson K, Kendler KS (2009a). Modeling the genetic and environmental association between peer group deviance and cannabis use in male twins. *Addiction* **104**, 420–429.

Gillespie NA, Neale MC, Kendler KS (2009b). Pathways to cannabis abuse: a multi-stage model from cannabis availability, cannabis initiation and progression to abuse. *Addiction* **104**, 430–438.

Hartman CA, Hopfer CJ, Haberstick B, Rhee SH, Crowley TJ, Corley RP, Hewitt JK, Ehringer MA (2009). The association between cannabinoid receptor 1 gene (CNR1) and cannabis dependence symptoms in adolescents and young adults. *Drug and Alcohol Dependence* **104**, 11–16.

Hasin DS, Saha TD, Kerridge BT, Goldstein RB, Chou SP, Zhang H, Jung J, Pickering RP, Ruan WJ, Smith SM, Huang B, Grant BF (2015). Prevalence of Marijuana Use Disorders in the United States Between 2001-2002 and 2012-2013. *JAMA psychiatry* **72**, 1235–1242.

Heath AC, Bucholz KK, Madden PA, Dinwiddie SH, Slutske WS, Bierut LJ, Statham DJ, Dunne MP, Whitfield JB, Martin NG (1997). Genetic and environmental contributions to alcohol dependence risk in a national twin sample: consistency of findings in women and men. *Psychological Medicine* **27**, 1381–1396.

Heath AC, Martin NG, Lynskey MT, Todorov AA, Madden PAF (2002). Estimating two-stage models for genetic influences on alcohol, tobacco or drug use initiation and dependence vulnerability in twin and family data. *Twin Research: The Official Journal of the International Society for Twin Studies* **5**, 113–124.

Hines LA, Morley KI, Mackie C, Lynskey M (2015a). Genetic and Environmental Interplay in Adolescent Substance Use Disorders. *Current Addiction Reports* **2**, 122–129.

Hines LA, Morley KI, Strang J, Agrawal A, Nelson EC, Statham D, Martin NG, Lynskey MT (2015b). The association between speed of transition from initiation to subsequent use of cannabis and later problematic cannabis use, abuse and dependence. *Addiction* **110**, 1311–1320.

Hines LA, Morley KI, Strang J, Agrawal A, Nelson EC, Statham D, Martin NG, Lynskey MT (2016). Onset of opportunity to use cannabis and progression from opportunity to dependence: Are influences consistent across transitions? *Drug and Alcohol Dependence* **160**, 57–64.

Hopfer C (2014). Implications of Marijuana Legalization for Adolescent Substance Use. *Substance Abuse* **35**, 331–335.

Johnson TP, Mott JA (2001). The reliability of self-reported age of onset of tobacco, alcohol and illicit drug use. *Addiction (Abingdon, England)* **96**, 1187–1198.

Kasriel J, Eaves L (1976). The zygosity of twins: further evidence on the agreement between diagnosis by blood groups and written questionnaires. *Journal of Biosocial Science* **8**, 263–266.

Kendler KS, Baker JH (2007). Genetic influences on measures of the environment: a systematic review. *Psychological Medicine* **37**, 615–626.

Kendler KS, Jacobson KC, Prescott CA, Neale MC (2003). Specificity of Genetic and Environmental Risk Factors for Use and Abuse/Dependence of Cannabis, Cocaine, Hallucinogens, Sedatives, Stimulants, and Opiates in Male Twins. *American Journal of Psychiatry* **160**, 687–695.

Kendler KS, Neale MC, Kessler RC, Heath AC, Eaves LJ (1993). A test of the equal-environment assumption in twin studies of psychiatric illness. *Behavior Genetics* **23**, 21–27.

Kendler KS, Neale MC, Sullivan P, Corey LA, Gardner CO, Prescott CA (1999). A population-based twin study in women of smoking initiation and nicotine dependence. *Psychol Med* **29**, 299–308.

Kendler KS, Pedersen NL, Farahmand BY, Persson PG (1996). The treated incidence of psychotic and affective illness in twins compared with population expectation: a study in the Swedish Twin and Psychiatric Registries. *Psychological Medicine* **26**, 1135–1144.

Laucht M, Becker K, Blomeyer D, Schmidt MH (2007). Novelty seeking involved in mediating the association between the dopamine D4 receptor gene exon III polymorphism and heavy drinking in male adolescents: results from a high-risk community sample. *Biological Psychiatry* **61**, 87–92.

Lynskey MT, Agrawal A (2009). Genetically Informative Studies Of 'Environment'''. *Addiction* **104**, 439–440.

Lynskey MT, Agrawal A, Henders A, Nelson EC, Madden PAF, Martin NG (2012). An Australian Twin Study of Cannabis and Other Illicit Drug Use and Misuse, and Other Psychopathology. *Twin Research and Human Genetics* **15**, 631–641.

Lynskey MT, Heath AC, Nelson EC, Bucholz KK, Madden PAF, Slutske WS, Statham DJ, Martin NG (2002). Genetic and environmental contributions to cannabis dependence in a national young adult twin sample. *Psychological Medicine* **32**, 195–207.

Malmberg M, Overbeek G, Monshouwer K, Lammers J, Vollebergh WAM, Engels RCME (2010). Substance use risk profiles and associations with early substance use in adolescence. *Journal of Behavioral Medicine* **33**, 474–485.

Morley KI, Lynskey MT, Madden PAF, Treloar SA, Heath AC, Martin NG (2007). Exploring the inter-relationship of smoking age-at-onset, cigarette consumption and smoking persistence: genes or environment? *Psychological Medicine* **37**, 1357–1367.

Neale M, Cardon L (1992). *Methodology for genetic studies of twins and families*. Kluwer Academic Publications: Dordrecht, The Netherlands.

Neale MC, Røysamb E, Jacobson K (2006). Multivariate Genetic Analysis of Sex Limitation and G × E Interaction. *Twin research and human genetics : the official journal of the International Society for Twin Studies* **9**, 481–489.

Nelson EC, Lynskey MT, Heath AC, Wray N, Agrawal A, Shand FL, Henders AK, Wallace L, Todorov AA, Schrage AJ, Saccone NL, Madden PA, Degenhardt L, Martin NG, Montgomery GW (2013). ANKK1, TTC12, and NCAM1 polymorphisms and heroin dependence: importance of considering drug exposure. *JAMA Psychiatry* **70**, 325–33.

Neumark Y, Lopez-Quintero C, Bobashev G (2012). Drug use opportunities as opportunities for drug use prevention: Bogota, Colombia a case in point. *Drug and Alcohol Dependence* **122**, 127–134.

Pagan JL, Rose RJ, Viken RJ, Pulkkinen L, Kaprio J, Dick DM (2006). Genetic and Environmental Influences on Stages of Alcohol Use Across Adolescence and into Young Adulthood. *Behavior Genetics* **36**, 483–497.

Parra GR, O'Neill SE, Sher KJ (2003). Reliability of self-reported age of substance involvement onset. *Psychology of Addictive Behaviors* **17**, 211–218.

Plomin R, DeFries JC, Knopik VS, Neiderheiser J (2013). *Behavioral Genetics*. 6th edn. Worth Publishers: New York.

R Core Team (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing: Vienna, Austria.

Reboussin BA, Ialongo NS, Green KM (2015). Influences of behavior and academic problems at school entry on marijuana use transitions during adolescence in an African American sample. *Addictive behaviors* **41**, 51–57.

Reinartz WJ, Haenlein M, Henseler J (2009). *An Empirical Comparison of the Efficacy of Covariance-Based and Variance-Based SEM*. SSRN Scholarly Paper 27 August no. ID 1462666. Social Science Research Network: Rochester, NY.

Sarna S, Kaprio J, Sistonen P, Koskenvuo M (1978). Diagnosis of twin zygosity by mailed questionnaire. *Human Heredity* **28**, 241–254.

Sartor CE, Lynskey MT, Bucholz KK, Madden PAF, Martin NG, Heath AC (2009). Timing of first alcohol use and alcohol dependence: evidence of common genetic influences. *Addiction* **104**, 1512–1518.

Sartor CE, Lynskey MT, Heath AC, Jacob T, True W (2007). The role of childhood risk factors in initiation of alcohol use and progression to alcohol dependence. *Addiction* **102**, 216–225.

Shi Y, Lenzi M, An R (2015). Cannabis Liberalization and Adolescent Cannabis Use: A Cross-National Study in 38 Countries. *PloS One* **10**, e0143562.

Shillington AM, Cottler LB, Mager DE, Compton III WM (1995). Self-report stability for substance use over 10 years: data from the St. Louis Epidemiologic Catchment Study. *Drug and Alcohol Dependence* **40**, 103–109.

Storr CL, Wagner FA, Chen CY, Anthony JC (2011). Childhood predictors of first chance to use and use of cannabis by young adulthood. *Drug and alcohol dependence* **117**, 7–15.

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

Wagner FA, Anthony JC (2002). Into the world of illegal drug use: Exposure opportunity and other mechanisms linking the use of alcohol, tobacco, marijuana, and cocaine. *Journal of Epidemiology*, 918–925.

Ystrom E, Reichborn-Kjennerud T, Neale MC, Kendler KS (2014). Genetic and Environmental Risk Factors for Illicit Substance Use and Use Disorders: Joint Analysis of Self and Co-twin Ratings. *Behavior Genetics* **44**, 1–13.

Table 1: Tetrachoric correlations (95% confidence intervals) between age of opportunity to use cannabis and cannabis abuse/dependence in MZ and DZ twin pairs

	Within trait, across twin correlation			Across trait, across twin correlation		
	Age of Opportunity twin 1/twin 2	Frequency cannabis use twin 1/twin 2	Abuse/ Dependence Twin 1/twin 2	Age of Opportunity / Frequency cannabis use	Age of Opportunity / Abuse/ Dependence	Frequency cannabis use/ Abuse/ Dependence
MZ N = 1709	0.65 (0.57 – 0.71)	0.72 (0.63 – 0.75)	0.79 (0.66 – 0.82)	0.48 (0.40 – 0.55)	0.37 (0.26 – 0.48)	0.67 (0.65 - 0.75)
DZ N = 1594	0.36 (0.26 – 0.45)	0.48 (0.38 – 0.58)	0.37 (0.26 – 0.48)	0.31 (0.22 – 0.38)	0.22 (0.12 – 0.33)	0.41 (0.29 – 0.52)

Table 2: Proportion of variance (95% CI) attributable to additive genetic (A), shared environment (C) and unique environment (E) factors in the fully estimated and in the most parsimonious model

		Opportunity	Frequency	Dependence	Correlation Opportunity – Frequency	Correlation Opportunity – Dependence	Correlation Frequency – Dependence
Fully estimated ACE model	A	0.57 (0.34 – 0.69)	0.46 (0.22 – 0.70)	0.64 (0.33 – 0.84)	0.35 (0.18 – 0.54)	0.27 (0.08 – 0.46)	0.49 (0.24 – 0.70)
	C	0.07 (0.00 – 0.27)	0.25 (0.04 – 0.45)	0.13 (0.00 – 0.38)	0.13 (-0.03 – 027)	0.09 (-0.05 – 0.23)	0.18 (0.02 – 0.38)
	E	0.35 (0.28 – 0.43)	0.28 (0.23 – 0.36)	0.23 (0.13 – 0.36)	-0.02 (-0.08 – 0.05)	0.01 (-0.08 - 0.10)	0.22 (0.14 – 0.31)
Parsimonious AE model	A	0.65 (0.58 – 0.72)	0.74 (0.66 – 0.80)	0.78 (0.65 – 0.88)	0.47 (0.41 – 0.52)	0.37 (0.30 – 0.44)	0.68 (0.59 – 0.75)
	E	0.35 (0.29 – 0.42)	0.26 (0.20 – 0.34)	0.22 (0.12 – 0.35)	-	-	0.21 (0.14 – 0.29)