OPEN ACCESS

University of Dundee

University of Dundee

Genetic and shared couple environmental contributions to smoking and alcohol use in the UK population

Clarke, Toni-Kim; Adams, Mark J.; Howard, David M.; Xia, Charley; Davies, Gail; Hayward, Caroline

ARTICLE

Genetic and shared couple environmental contributions to smoking and alcohol use in the UK population

Toni-Kim Clarke[1] · Mark J. Adams [1] · David M. Howard [1] · Charley Xia[2] · Gail Davies[3] · Caroline Hayward [2] · Archie Campbell[4,5] · Sandosh Padmanabhan[6] · Blair H. Smith [7] · Alison Murray[8] · David Porteous [5] · Ian J. Deary[3,9] · Andrew M. McIntosh [1,9]

Abstract

Alcohol use and smoking are leading causes of death and disability worldwide. Both genetic and environmental factors have been shown to influence individual differences in the use of these substances. In the present study we tested whether genetic factors, modelled alongside common family environment, explained phenotypic variance in alcohol use and smoking behaviour in the Generation Scotland (GS) family sample of up to 19,377 individuals. SNP and pedigree-associated effects combined explained between 18 and 41% of the variance in substance use. Shared couple effects explained a significant amount of variance across all substance use traits, particularly alcohol intake, for which 38% of the phenotypic variance was explained. We tested whether the within-couple substance use associations were due to assortative mating by testing the association between partner polygenic risk scores in 34,987 couple pairs from the UK Biobank (UKB). No significant association between partner polygenic risk scores were observed. Associations between an individual's alcohol PRS ($b = 0.05$, S.E. $= 0.006$, $p < 2 \times 10^{-16}$) and smoking status PRS ($b = 0.05$, S.E. $= 0.005$, $p < 2 \times 10^{-16}$) were found with their partner's phenotype. In support of this, G carriers of a functional *ADH1B* polymorphism (rs1229984), known to be associated with greater alcohol intake, were found to consume less alcohol if they had a partner who carried an A allele at this SNP. Together these results show that the shared couple environment contributes significantly to patterns of substance use. It is unclear whether this is due to shared environmental factors, assortative mating, or indirect genetic effects. Future studies would benefit from longitudinal data and larger sample sizes to assess this further.

Introduction

Alcohol and tobacco have been used recreationally by humans across the world for centuries. The extent to which an individual uses alcohol and tobacco, and whether they use them at all, depends on individual genetics, environment and cultural attitudes and the complex interactions

✉ Toni-Kim Clarke
   toni.clarke@ed.ac.uk

1   Division of Psychiatry, Royal Edinburgh Hospital, University of Edinburgh, Edinburgh, UK

2   MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK

3   Department of Psychology, University of Edinburgh, Edinburgh, UK

4   Usher Institute for Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK

5   Centre for Genomic and Experimental Medicine, Institute of Genetics & Molecular Medicine, University of Edinburgh, Edinburgh, UK

6   Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK

7   Division of Population and Health Genomics, University of Dundee, Dundee, UK

8   The Institute of Medical Sciences, Aberdeen Biomedical Imaging Centre, University of Aberdeen, Aberdeen, UK

9   Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK

between these factors. There has been extensive research into individual differences in alcohol and tobacco use, and the genetic component of these behaviours is well established. Heritability estimates range from 10 to 60% for alcohol use [1–3], with alcohol use disorders tending to have higher estimates than for levels of consumption [3, 4]. Similarly, smoking behaviours have a genetic component with the heritability estimates of nicotine dependence higher (60–70%) than tobacco use (ever vs never) (40–50%) [4–6].

It is clear from heritability studies that a significant proportion of the phenotypic variance in individual differences comes from environmental, or other unmeasured sources, and measurement error. Childhood trauma, parental substance dependence, parental divorce and stressful life events have all been cited as environmental risk factors [4, 7, 8]. Environmental influences on substance use are typically found to be more pronounced in adolescence and are associated with first use [8–12]. The proportion of variance in alcohol initiation explained by environmental factors has been found to be as high as 76% [13]. Indeed, twin studies have shown that the shared family environment accounts for 23% of the variance in drug use [14] but have also highlighted the importance of the environment that is unique to the individual [10]. Typically, environmental influences decrease in importance and genetic effects become more prominent moving from adolescence into early adulthood [9]. The decline in environmental effects is thought in part to reflect the waning influence of authority figures and peers as individuals gain more independence.

The role of the recent shared environment and its effect on adult alcohol and tobacco use is less well studied. Observing the correlations between couples is a useful measure as couples typically share many aspects of the recent environment, whereas their earlier exposures are distinct. Studies have shown that members of a couple have similar levels of substance use [15, 16]. The rate of alcohol use disorder in the husbands of female alcoholic probands was found to be 31% [15] and the correlation for substance dependence symptoms among mothers and fathers of youths in treatment for substance use disorder was found to be 0.4 [16]. Furthermore, females who abuse alcohol and nicotine are more likely to have married individuals who do the same [17].

One explanation for the observed similarities between couple's substance use is assortative mating. Assortative mating is a pattern of non-random mating whereby individuals with similar phenotypes are more likely to mate with one another and this leads to increased genetic similarity at loci known to be associated with substance use. Although assortative mating has been proposed to increase alcohol dependence correlations between spouses [18], it may be that indirect genetic effects contribute to phenotypic correlations.

Indirect genetic effects occur when the genotype of an individual influences the phenotype of another conspecific individual. For example, people at high genetic risk for alcohol use who drink heavily could create an environment which increases their partner's risk for alcohol use, such as increased alcohol availability or a stressful environment arising from problematic drinking. In the context of smoking, individuals at low genetic risk for smoking who do not smoke may encourage their partner to quit smoking. Furthermore, as substance use patterns are dynamic, individuals may change their behaviour over time to match that of their partners.

In the present study, we aimed to measure the genetic and environmental contributions to people's differences in alcohol use and smoking behaviour in a population-based cohort, Generation Scotland: the Scottish Family Health Study (GS) [19, 20]. Exploiting the diverse family relationships in GS, we estimate the contribution of shared family, sibling and couple effects on substance use, and estimate the proportion of phenotypic variance attributable to genetic effects in the presence of these factors. To investigate the potential role of assortative mating, we estimated spousal phenotypic associations for alcohol and smoking use phenotypes across 34,987 couple pairs in the UK Biobank (UKB). As assortative mating can lead to increased genetic similarity between individuals, we also estimated the intra-couple polygenic risk score associations for alcohol and nicotine use phenotypes. We also tested whether a spousal PRS predicted their partner's phenotype. Finally, we explored whether a functional SNP (rs1229984) in *ADH1B*, that influences alcohol metabolism and is strongly associated with alcohol intake, was associated with levels of the partner's drinking.

Methods

Sample descriptions

Generation Scotland: the Scottish Family Health Study

Generation Scotland: the Scottish Family Health Study (GS) is a family-based cohort recruited via general practitioners across Scotland. Individuals were invited to participate if they were able to recruit at least one other family member aged 18 or over. Ethical approval for GS was obtained from NHS Tayside Research Ethics Committee (REC reference number 05/S1401/89) and informed consent was obtained for all participants.

Genotyping

Genotyping was performed on 20,195 individuals using the Illumina OmniExpress BeadChip. Quality control steps

removed individuals with a genotype call rate <98%, SNPs with a call rate of <98%, SNPs with a minor allele frequency <1%, or those which deviated from Hardy–Weinberg equilibrium ($p < 5 \times 10^{-6}$). Principal component analyses were also performed to remove population outliers [21]. After quality control, 19,904 individuals remained with 561,125 autosomal SNPs.

Phenotypes

Smoking status Smoking behaviours were assessed as part of a pre-clinical questionnaire. Individuals were asked whether they were current, former or never smokers. Former and current smokers were then collapsed to create an ever/never smoking variable.

Cigarettes per day Participants who had endorsed ever smoking were asked the average amount of cigarettes they smoked daily, currently, or in the past. Cigarettes per day was recorded as an ordinal variable and treated as an interval variable in the present study with $0 = 0$, $1 =$ Less than daily, $2 = 1$–4, $3 = 5$–9, $4 = 10$–14, $5 = 15$–19, $6 = 20$–24, $7 = 25$–29, $8 = 30$–34, $9 = 35$–39, $10 = 40$–44, $11 = 45$–49, $12 = 50+$ cigarettes per day.

Smoking age of onset Lifetime smokers were asked at what age they started smoking and responses were categorised and treated as an interval variable so that $1 =$ less than 5 years, $2 = 5$–9, $3 = 10$–14, $4 = 15$–19, $5 = 20$–24, $6 = 25$–29, $7 = 30$–34, $8 = 35$–39, $9 = 40$–44, $10 = 45$–49, $11 = 50+$ years of age. Individuals who did not know what age they started smoking were set to NA for age of onset.

Alcohol consumption This was assessed using self-report as part of the pre-clinical questionnaire; participants were asked how many units of alcohol they had consumed in the previous week. A prompt was shown in the questionnaire to provide examples of the typical units of alcohol in each drink type.

Alcohol misuse As part of a GS re-contact study in 2014, 9618 members of GS completed a follow-up questionnaire as part of the Stratifying Resilience and Depression Longitudinally project [22]. These individuals completed the CAGE questionnaire [23] which can identify individuals at risk of problem drinking. The CAGE questionnaire consists of four questions and provides a total score of 0–4 depending on the number of items endorsed.

The total sample size for each of the GS phenotypes and mean and standard deviations are shown in Supplementary Table 1.

Identification of couple pairs

Using the family and genetic data in GS, couples were identified as those who shared a child. This identified 1742 genotyped couple pairs.

UK Biobank

The UKB is a prospective population-based sample of 502,629 participants recruited across 22 assessment centers in the United Kingdom from the period of 2006–2010 [24]. People were invited to participate if they were aged between 40 and 69 years, were registered with the National Health Service and lived within ~25 miles of an assessment center. Informed consent was obtained from all participants and was conducted under generic approval from the National Health Service National Research Ethics Service (Ref 11/NW/0382) and under UKB approval for project 4844.

Genotyping

Genetic data were available for 487,409 individuals in the UKB and genotyping was performed on either the Affymetrix Axiom array or the UK BiLEVE Axiom array [25]. In order to create a White British unrelated dataset, we removed 131,790 related individuals who were third degree relatives or closer (using a kinship coefficient > 0.044). We identified one individual from each group of relatives by creating a genomic relationship matrix and using a genetic-relatedness cut-off of 0.025 and added these back into the sample ($N = 55,745$). Quality control steps removed individuals with a genotype call rate <98%, SNPs with a call rate of <98%, SNPs with a minor allele frequency <1% or those which deviated from Hardy–Weinberg equilibrium ($p < 5 \times 10^{-6}$). After quality control 414,584 autosomal SNPs remained.

Phenotypes

Smoking status Smoking status was ascertained as part of a touchscreen interview. Participants were asked whether they were current, previous or never smokers; previous and current smokers were collapsed to make an 'ever smoker' phenotype. Former smokers were asked about previous smoking behaviour. Those who endorsed 'Just tried once or twice' were classed as never smokers and therefore the phenotype is an ever/never-regular smoker as defined according to the Centre for Disease Control and Prevention (CDC).

Cigarettes per day Current smokers were asked how many cigarettes they smoked on average each day and former smokers about how many cigarettes they previously

smoked. If individuals stated they smoked over 150 cigarettes per day this answer was rejected; if they endorsed 100 or over they were asked to confirm this selection.

Smoking age of onset Lifetime smokers were asked using the touchscreen questionnaire how old they were when they first started smoking on most days. Responses under age 5 were rejected and under age 12 were prompted for confirmation.

Alcohol consumption Participants were asked how many of various drink types they normally drank on a monthly and weekly basis and this was converted into a measure of units per week. The full derivation of this measure has been described previously [26].

Alcohol misuse The AUDIT questionnaire [27] was administered to a subset of the UKB who responded to an online mental health questionnaire follow-up over a 1 year period in 2017. The AUDIT is a ten-item questionnaire with scores ranging from 0 to 40 that measures both alcohol consumption (Q1–Q3) and problems with alcohol (Q4–Q10). Three AUDIT scores were created based on the score for all questions (AUDIT-T), on the questions measuring alcohol consumption and frequency (AUDIT-C [Q1–Q3]), and on those measuring problems with alcohol or alcohol abuse (AUDIT-P [Q4–Q10]). These measures have been described in greater detail previously [28].

The total sample size for each of the UKB phenotypes used in the present study and mean and standard deviations are shown in Supplementary Table 1.

Identification of couple pairs

Participants were assigned to couple pairs on the basis of a shared household identifier. Individuals who shared a household, reported living in a household with two individuals, and who reported living with a husband, wife, or partner were selected. Any couples with an age gap of >10 years were removed, as were couples whose parental ages matched for either parent. After further selecting White British unrelated individuals from this group there were 34,987 opposite-sex pairs available for analysis. A total of 407 same-sex couples were also identified using the above algorithm. Due to the lower number of same-sex pairs genetic associations were not analysed in these individuals although phenotypic associations were estimated.

Heritability analyses in Generation Scotland

Genetic and environmental effects were estimated in GCTA v1.91 using linear mixed models [29] by fitting a pedigree kinship matrix and a SNP matrix (genetic relationship matrix) alongside three matrices representing the environment shared by nuclear families (parents and children) (F), couples (identified by a shared child) (C) and siblings (S).

$$Y = Xb + G + K + F + S + C + \varepsilon,$$

where Y is a vector representing the substance use trait of interest and b is the effect of X, a matrix of values that represents the fixed effect covariates of age, sex and 20 principal components. The genetic effects are represented by G (SNP matrix) and K (pedigree kinship matrix), the three environmental components are F, S and C, and ε the residual error term. This method was first described by Xia et al. [30], and the construction of the genetic and environmental matrices are described in more detail in the Supplementary Material. Briefly, the G component captures variance explained by common SNPs, the K component captures additional genetic effects by modelling pedigree relationships (achieved by setting all entries in the SNP matrix <0.025–0). The F component represents nuclear family members by setting the relationship matrix coefficient to 1 if individuals were parent-offspring, sibling or couples. Similarly, the S component represents sibling pairs and the C component couple pairs.

The most parsimonious model was selected by performing backward stepwise selection. The initial model included all five components (GKFCS) and components were removed iteratively if they failed to meet significance in the likelihood ratio test (LRT) and Wald tests ($\alpha = 5\%$) and among the components satisfying this condition it had the highest (least significant) P value in the Wald test. This process was repeated until all the remaining components were significant in either the LRT or Wald test. The population prevalence for smoking status was 48% and used to convert the estimates for this trait from the observed scale to the liability scale.

Polygenic risk scores

In order to ensure no overlap between training and test datasets we took an unrelated white British subset of the UKB and then removed all of the identified couple pairs used in the present study. Genome-wide association was then performed on the remaining individuals for alcohol consumption (units per week) ($N = 279,418$ [range = 0–102 units, mean = 15.1, SD = 16.1]), cigarettes per day ($N = 91,762$ [range 1–140, mean = 18.3, SD = 10.2]), age of smoking onset ($N = 96,430$ [range 5–69 years, mean = 17.4 years, SD = 4.3]) and smoking status ($N = 136,999$ cases and 163209 controls) using BGENIE, version 1.1 [25], with age, sex, genotyping array and the first 20 genetics principal components as covariates.

PRS were created in PRCise-2 [31] using raw QC'd genotype data and a MAF cut-off of 0.01. The parameters of

$r^2 = 0.1$ and window $= 250$ kb were used to create independent SNPs and the scores for p value thresholds from 0.00005 to 0.5 created in increments of 0.00005. The score which explained the most variance in the trait of interest was then used for downstream analyses (Supplementary Figs. 1–4). PRS were regressed onto the first four principal components to correct for population stratification and the residuals taken for analyses.

Statistical analyses

Phenotypic associations were tested in R using linear regression models. Baseline models tested the phenotypic association between substance use phenotypes without controlling for any covariates. Phenotypes were regressed onto age and sex, and then age, sex and test-center (categorical) and the residuals from these used for regression analyses. Association between PRS (residualized for principal components) were also performed in R using linear models. Variables were scaled to have a mean of 0 and a standard deviation of 1 and therefore the reported beta are standardised. Permutation tests were carried out to test the independence of couple phenotypes using the *coin* package in R and 10,000 Monte Carlo re-samplings [32].

Results

The total sample size for each of the substance abuse phenotypes available in Generation Scotland and UKB is shown in Supplementary Table 1, along with the mean and standard deviations for each trait. The variance explained by each genetic and environmental component in the Generation Scotland cohort is shown in Table 1 and Fig. 1. Significant genetic effects were detected for all traits. Units per week of alcohol has the lowest SNP-based genetic estimate, accounting for 6% of the variance (S.E. $= 0.02$). The CAGE score's SNP-based genetic contribution was 19% (S.E. $= 0.03$). Smoking status had the highest estimate ($G = 0.22$, S.E. $= 0.03$). With the exception of CAGE score, the kinship component significantly contributed to the variance explained for all traits, with values between 12 and 20%. This suggests a role for additional genetic effects such as

rare variants or epistatic effects that are detectable when analysing close relatives. The sum of the G and K components is comparable with narrow-sense heritability estimates [30] and therefore the total genetic contribution to units per week and CAGE score was 18% and 19% respectively. For smoking status, smoking age of onset and cigarettes per day the narrow-sense heritability estimates were 41%, 26% and 41% respectively.

The most significant environmental contribution across all traits was the couple component (C). The contribution was 29% (S.E. $= 0.04$) for smoking status and 9% (S.E. $= 0.05$) for cigarettes per day. The largest contribution, of 38% (S.E. $= 0.03$), was to the phenotypic variance explained for units of alcohol consumed per week. Modest early environmental effects were also detected for units per week and smoking status, as follows: the nuclear family component (F) explained 7% of the variance (S.E. $= 0.03$) in units per week and the shared sibling environment (S) explained 10% of the variance in smoking status (S.E. $= 0.03$) (Table 1) (Fig. 1).

The results of the full backward stepwise model selections are shown in Supplementary Table 2. For smoking status, 80% of the variance was explained suggesting that only 20% of the variance can be apportioned to other environmental effects or sampling error (Fig. 1). The total variance explained in units per week was 63% and for cigarettes per day and CAGE score the total variance explained was 50%. For some traits the majority of phenotypic variance was unexplained: only 35% of the variance in age of smoking onset was explained, suggesting that the majority of the variance in this trait is influenced by unique environmental factors or shared factors that are not captured by the current model. The unexplained variance could also be attributed to measurement error. Substance use can be difficult to measure as it relies on accurate recall of behaviours which change across the lifespan.

All traits in GS showed a significant amount of variance explained by the couple environment (C, Table 1). The within-couple phenotypic associations in GS are shown in Supplementary Table 3. Smoking status ($b = 0.19$ (S.E. $= 0.02$)), alcohol consumption ($b = 0.26$ (S.E. $= 0.03$)) and CAGE score ($b = 0.22$ (S.E. $= 0.05$)) were all significantly associated within-couple pairs. After residualizing the traits for age, sex and recruitment area a nominally significant

Table 1 Variance component analysis in Generation Scotland showing the significant environmental and genetic components for each trait when fitted simultaneously

Trait	G (Genetic—SNP) (S.E.)	K (Genetic—Kinship) (S.E.)	F (Family) (S.E.)	C (Couple) (S.E.)	S (Sibling) (S.E.)
Units per week	0.06 (0.02)	0.12 (0.06)	0.07 (0.03)	0.38 (0.03)	–
CAGE score	0.19 (0.03)	–	–	0.31 (0.04)	–
Smoking status	0.22 (0.03)	0.19 (0.05)	–	0.29 (0.04)	0.10 (0.03)
Cigarettes per day	0.21 (0.05)	0.20 (0.06)	–	0.09 (0.05)	–
Smoking age onset	0.14 (0.05)	0.12 (0.05)	–	0.09 (0.05)	–

Fig. 1 Proportion of variance in substance use traits explained by genetic and environmental components in Generation Scotland. G genetic, K kinship, F nuclear family, C couple, S sibling
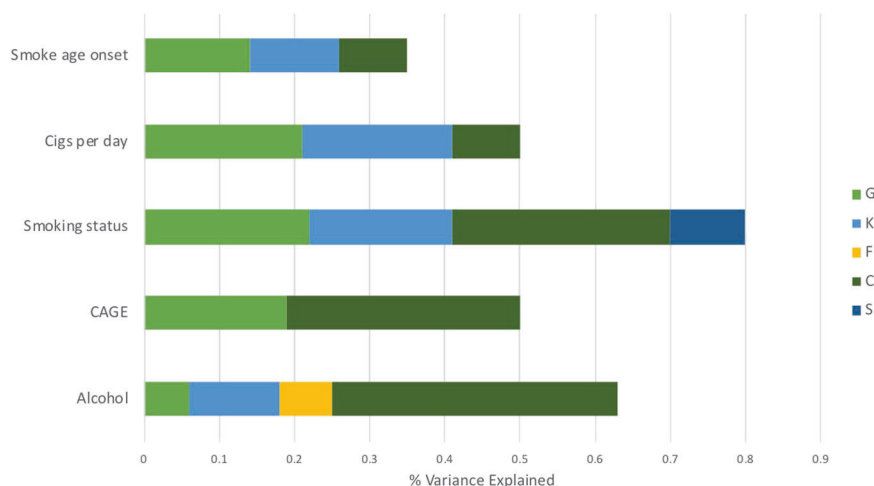

Table 2 Phenotypic correlations between substance use phenotypes in opposite-sex couples in UKB ($N = 34,987$ pairs). N shown in trait column reflect the N where both members of the couple had available phenotype data. Reported betas are standardised

Trait	Baseline			Age + Sex			Age + Sex + Test centre		
	Beta	SE	P	Beta	SE	P	Beta	SE	P
Smoking status $N = 34,732$	0.22	0.005	$<2 \times 10^{-16}$	0.21	0.005	$<2 \times 10^{-16}$	0.21	0.005	$<2 \times 10^{-16}$
Smoking age onset $N = 5226$	0.09	0.013	9×10^{-12}	0.09	0.01	3×10^{-12}	0.09	0.01	1×10^{-10}
Smoking cigs per day $N = 4749$	0.13	0.01	$<2 \times 10^{-16}$	0.13	0.01	$<2 \times 10^{-16}$	0.12	0.01	$<2 \times 10^{-16}$
Alcohol units per week $N = 31,263$	0.52	0.005	$<2 \times 10^{-16}$	0.52	0.005	$<2 \times 10^{-16}$	0.52	0.005	$<2 \times 10^{-16}$
AUDIT score $N = 6787$	0.48	0.011	$<2 \times 10^{-16}$	0.47	0.01	$<2 \times 10^{-16}$	0.47	0.01	$<2 \times 10^{-16}$
AUDIT-C $N = 6787$	0.51	0.011	$<2 \times 10^{-16}$	0.50	0.01	$<2 \times 10^{-16}$	0.50	0.01	$<2 \times 10^{-16}$
AUDIT-P $N = 6787$	0.12	0.009	$<2 \times 10^{-16}$	0.11	0.009	$<2 \times 10^{-16}$	0.11	0.009	$<2 \times 10^{-16}$

Table 3 Association between couple polygenic risk scores (PRS) in UKB for substance use traits

PRS trait	Association between couples PRS UKB	
	PRS beta (S.E.)	P value
Alcohol consumption	0.001 (0.005)	0.82
Smoking status	0.008 (0.005)	0.12
Cigarettes per day	0.008 (0.005)	0.14
Age smoking onset	−0.001 (0.005)	0.85

couple association was also observed for cigarettes per day ($b = 0.10$, (S.E. $= 0.05$)) (Supplementary Table 3) and the couple association for alcohol consumption became stronger ($b = 0.41$, (S.E. $= 0.02$)).

Similar phenotypic associations were observed between members of couple pairs in the UKB. Smoking status, cigarettes per day, age of smoking onset, units of alcohol per week and AUDIT scores were all significantly associated within-couple pairs (Table 2). Controlling for age, sex and recruitment center did not significantly alter the observed associations. AUDIT scores were strongly associated between

members of a couple ($b = 0.48$ (S.E. $= 0.01$), $p < 2 \times 10^{-16}$) as were units per week ($b = 0.52$ (S.E. $= 0.005$), $p < 2 \times 10^{-16}$). Smoking status, cigarettes per day and age of smoking onset were more modestly associated within couples ($b = 0.09$–0.22, $p < 9 \times 10^{-12}$).

There were 407 same-sex pairs identified in the UKB and the phenotypic associations between these individuals are presented in Supplementary Table 4. Although there were very few individuals for some phenotypes, the associations for the phenotypes with larger sample sizes (smoking status, units per week) appear similar to those observed in opposite-sex pairs.

Couple correlations can arise because of assortative mating, whereby individuals with similar phenotypes mate, potentially resulting in greater genetic similarity between members of a couple. In order to test this, polygenic risk scores (PRS) were created for each substance use trait using GWAS summary statistics from independent samples. The associations between couples PRS were then tested in the UKB as there were more couple pairs available ($N = 34,987$ vs $N = 1742$ in GS). No significant associations were observed between partners' PRS in the UKB sample (Table 3).

Table 4 Association between male PRS and female PRS and partner and own phenotype in the UKB. All significant p values (<0.05) have permutation p values <0.05

PRS association (standardised β, standard error, p value and r^2)	
Males—PRS with partner phenotype	
Alcohol PRS	β = 0.054 (S.E. = 0.006) $p < 2 \times 10^{-16}$, $r^2 = 0.29\%$
Age smoking onset PRS	β = 0.022 (S.E. = 0.01) $p = 0.034$, $r^2 = 0.04\%$
Smoking status PRS	β = 0.051 (S.E. = 0.005) p 2×10^{-16}, $r^2 = 0.25\%$
Cigarettes per day PRS	β = 0.015 (S.E. = 0.01) $p = 0.15$, $r^2 = 0.01\%$
Females—PRS with partner phenotype	
Alcohol PRS	β = 0.043 (S.E. = 0.005) $p = 1.7 \times 10^{-15}$, $r^2 = 0.19\%$
Age smoking onset PRS	β = 0.025 (S.E. = 0.008) $p = 0.002$, $r^2 = 0.06\%$
Smoking status PRS	β = 0.040 (S.E. = 0.005) $p = 1.5 \times 10^{-13}$, $r^2 = 0.15\%$
Cigarettes per day PRS	β = 0.008 (S.E. = 0.009) $p = 0.39$, $r^2 = 0\%$
Own—PRS with own phenotype	
Alcohol PRS	β = 0.099 (S.E. = 0.004) $p < 2 \times 10^{-16}$, $r^2 = 0.98\%$
Age smoking onset PRS	β = 0.039 (S.E. = 0.007) $p = 2.9 \times 10^{-9}$, $r^2 = 0.15\%$
Smoking status PRS	β = 0.062 (S.E. = 0.002) $p < 2 \times 10^{-16}$, $r^2 = 1.56\%$
Cigarettes per day PRS	β = 0.087 (S.E. = 0.007) $p < 2 \times 10^{-16}$, $r^2 = 0.76\%$

Individuals' PRSs were tested for association with the partner's substance use phenotypes. Male alcohol consumption PRS was significantly positively associated with female partner's alcohol consumption in the UKB ($b = 0.054$, S.E. = 0.006, $p < 2 \times 10^{-16}$, $r^2 = 0.29\%$) (permutation p value $< 2 \times 10^{-16}$). The same was observed for female alcohol consumption PRS—a significant association with male partner phenotype was found ($b = 0.043$, S.E. = 0.005, $p = 1.7 \times 10^{-15}$, $r^2 = 0.19\%$) (permutation p value $< 2 \times 10^{-16}$) (Table 4). The association between alcohol consumption PRS and partner consumption is weaker and explains less of the variance than the association with an individual's own alcohol consumption ($b = 0.099$, S.E. = 0.004, $p < 2 \times 10^{-16}$, $r^2 = 0.98\%$) (Table 4). Significant associations were also observed for smoking status PRS, age of smoking onset PRS and partner phenotype in the UKB. The PRSs for cigarettes per day were not associated with corresponding partner phenotypes in UKB (Table 4).

The association between the rs1229984 *ADH1B* SNP and units per week was also tested in the UKB. rs1229984 is a non-synonymous SNP in the alcohol dehydrogenase 1B gene (*ADH1B);* the minor allele (A) carriers have a version of ADH1B that oxidises alcohol more rapidly, and as such A carriers are at a reduced risk for alcohol use disorder [33, 34]. In the present sample of UKB individuals, those with the AA/AG genotype at rs1229984 drank 12.3 units per week on average (S.E. = 0.22) compared with GG individuals who drank 16.1 units (S.E. = 0.06) ($p < 2 \times 10^{-16}$). Units per week of alcohol for each genotype in the UKB couples sample is shown in Supplementary Fig. 5. This finding was replicated in the GS sample, those with the AA/AG genotype drank 8.1 (S.E = 0.92) units per week on average, compared with

11.5 (S.E = 0.22) units per week for those carrying the GG genotype ($p = 0.007$).

We next took all the individuals with a GG genotype in the UKB and split them according to whether they had a partner with the GG genotype (GG-G) or an A carrier partner (AG or AA) (GG-A). GG-G individuals consumed on average 16.1 units per week (S.E. = 0.06). GG-A individuals consumed significantly less than their GG-G equivalents, 14.7 units per week (S.E. = 0.26) ($p < 9 \times 10^{-7}$). Having a partner who carries an rs1229984 A allele was, therefore, associated with a mean of 1.4 units less alcohol per week intake amongst G carrier individuals. A similar pattern was observed in the GS sample. GG-G individuals consumed on average 11.5 units per week (S.E. = 0.22). GG individuals with a partner carrying an A genotype at rs1229984 (GG-A) drank, on average, 2 units less per week (mean = 9.4 units, S.E = 1.01). Due to the small number of GG-A individuals ($N = 90$) this was not statistically significant ($p = 0.09$).

Discussion

In the present study, using genotyping and family relationships data, we show that there are significant genetic and environmental contributions to substance use in a general population sample, Generation Scotland. The effect of the shared couple environment was particularly pronounced and contributed significantly to the variance in each trait. In support of this, we report significant phenotypic association within couples for all of the substance use traits in the UKB. In order to test whether this was due to assortative mating we analysed the association between partners' substance use PRS in the UKB. Whereas there was

no significant association between alcohol consumption PRS within couples, an individual's alcohol consumption PRS associated with their partner phenotype. Furthermore, the presence of the rs1229984 A allele in a partner was associated with reduced alcohol intake in individuals with GG genotypes at this locus.

The narrow-sense heritability of alcohol use phenotypes reported in this study (sum of G and K) are lower than those generally reported in the literature. The narrow-sense heritability of alcohol consumption and CAGE score was estimated to be 18% and 19%, respectively. Broad sense heritability estimates of 25–61% for alcohol consumption have previously been reported from studies of twins [1, 2, 5]. The SNP effects for alcohol consumption were estimated at 6%, which again are lower than, but closer to, estimates reported in the UKB for alcohol consumption (13%) and AUDIT scores in the 23andMe sample (12%) [28, 35]. Previous studies have suggested that genetic interactions and improper modelling of the environment can inflate heritability estimates [36, 37]. The narrow-sense heritability estimates for cigarette smoking were higher ranging from 26% for age of onset to 41% for smoking status and cigarettes per day. These are somewhat lower than heritability estimates from twin studies (typically 45–80%). We report, the SNP heritability of smoking status (22%), age of smoking onset (14%) and cigarettes per day (21%). The SNP heritability of smoking status has previously been estimated at 17% in the UKB, similar to our estimate in GS [38].

Early environmental factors, such as those shared by families and siblings, did not appear to explain large amounts of variance in adult substance use in this sample. Shared family environment was estimated to explain 7% of the variance in units per week of alcohol consumed. Given that the age range of the Generation Scotland sample is 18–99 years, it is likely that most members of a nuclear family no longer share a household and so the family component should represent early shared environment in this study. Parental expectations, attitudes and alcohol use have all been shown to influence adolescent alcohol use. Fewer studies have examined the effect of familial influences into adulthood; however, a family history of alcohol abuse and [39] age of first drink [40] are associated with alcohol abuse in later life, although it is unclear whether these represent genetic or environmental factors. A family study from the Netherlands found non-shared environmental factors to explain the majority of the variance in alcohol consumption and found no evidence of cultural transmission influencing adult alcohol use [41]. The findings from our study suggest that there may be a small contribution of family environment on drinking patterns in later life; these discrepant findings may be due to cultural differences between the samples.

A significant shared sibling effect was detected for smoking status, explaining 10% of the variance in this trait. Sibling effects can represent genetic or environmental effects; however, as we model genetic effects simultaneously in our model the component captures the effect of the early shared environment. Previous studies have shown sibling concordance in smoking status and this is greater when siblings report a high degree of social connectedness [42]. This suggests we may be detecting shared peer effects or the influence of one sibling's smoking status on the other [43]. No shared early environmental influences were detected for cigarettes per day or age of smoking onset. This is in contrast to twin studies which often report a significant contribution to smoking initiation from the shared environment. In a large meta-analysis Li et al. found 24–49% of the variance in smoking initiation was attributed to the shared environment [44]. These differences may be due to twins having a more similar shared environment than the family members modelled in this study (parents and siblings). It should be noted that other twin studies have found little influence of the shared environment on age of smoking onset [45], similar to the findings we report in this study.

The proportion of phenotypic variance in alcohol consumption explained by the couple environment in GS was substantial at 38%. The phenotypic associations for all alcohol use phenotypes in both GS and the UKB were high. The associations between partner alcohol consumption, AUDIT score and AUDIT-C (consumption) were 0.47–0.52 (standardised beta) in the UKB; however, the AUDIT-P (problems) association was smaller (0.12, S.E. = 0.009). Similarly, in GS the alcohol consumption association in GS was high (0.4) whereas the phenotypic association between partner CAGE scores was lower (0.22, S.E. = 0.05), demonstrating that alcohol consumption is more strongly correlated between partners than patterns of alcohol abuse in these samples. Correlations between partners can be driven by assortative mating. Partner alcohol consumption PRS were not associated with one another in the UKB; however, alcohol consumption PRS did predict partner phenotype in the UKB for both males and females. Also, having a partner who carries an A allele at the rs1229984 locus was associated with lower alcohol intake among G carriers of this SNP. PRS typically explain very little of the variance in the traits they predict (<1%) and therefore the lack of association between couples PRS does not rule out assortative mating as an explanation for the couple similarities.

Indirect genetic effects occur when the genotype of one individual influences the phenotype of another. The influence of genotype on partners' substance use may be via the contribution of that genotype to the environment, such as creating high or low exposure to alcohol. This is similar to the genetic nurture effect described by Kong et al. [46].

Using PRS for educational attainment, they show that the offspring of parents with higher PRS have greater educational attainment themselves, even when they do not inherit the 'education-associated' alleles. The nurturing environment provided by the parents with higher PRS is proposed to increase educational attainment of the offspring. In the case of alcohol consumption, partner genotype may lead to higher or lower alcohol exposure, or different attitudes towards alcohol use, which could lead to changes in partner substance use. It is difficult to distinguish between indirect genetic effects and assortative mating from our results alone, and it is possible that both are occurring. Furthermore, levels of alcohol consumption between members of a couple may become more similar over time, potentially in response to shared environmental factors such as life stress or social deprivation. Longitudinal samples or samples with more couple pairs are required to tease apart the potential contributions of each of these factors to couple substance use behaviour.

For the smoking phenotypes the variance explained by the couple environment ranged from 9 to 29%. As age of smoking onset and smoking status are typically determined during adolescence or early adulthood, behaviour convergence is less likely to explain the significant shared couple environment effect observed for these traits. Significant couple associations were observed for age of smoking onset or cigarettes per day; however, it should be noted that the PRS for smoking initiation only weakly predicted age of onset in the UKB (Table 4) and therefore may be a poor instrument to test for assortative mating. Assortative mating can also be measured by assessing the gametic phase disequilibrium (GPD) of trait-increasing alleles across the genome [47]. GPD, as a consequence of assortative mating, manifests as an increased likelihood of carrying trait-increasing alleles across the genome, independent of linkage disequilibrium. Deriving PRS from odd numbered chromosomes and analysing the correlation with PRS derived from even numbered chromosomes can quantify GPD. A recent study, which also used UKB individuals, found no evidence of GPD for alcohol use or smoking behaviour providing additional evidence that assortative mating does not significantly contribute to the phenotypic couple correlations reported in the literature [47].

There are a number of limitations to this study. The presence of assortative mating has implications for the heritability estimates of substance use phenotypes. By incorrectly modelling the couple effect as an environmental effect we reduce the residual error term in the model and may inflate the heritability estimates; however, in the absence of longitudinal data it is difficult to determine whether assortative mating or shared couple environment is responsible for the association between substance use phenotypes. Another limitation is that the substance use phenotypes are based on self-report, and for the initiation of smoking and cigarettes per day, rely on retrospective accounts which can be unreliable. Also, the definition of a never smoker according to the CDC is someone who has smoked <100 cigarettes in their lifetime. We were able to create a phenotype similar to this in the UKB, but for GS we had to dichotomise smokers into never versus ever smokers and therefore these phenotypes are not directly comparable. Finally, assigning individuals to couples was done differently in GS and UKB. Genetic data was used to identify couples who shared a child in GS, but it is possible that these individuals did not share a household at the time of recruitment. Given that GS was recruited through family participation this is less likely but cannot be ruled out. Similarly for UKB, couple data was not linked in the database but using strict exclusion criteria we were able to generate couples from the household data provided. It is more likely that we excluded potential couples from the UKB.

In conclusion, we find that the shared couple environment explains a large amount of the variance in substance use phenotypes, particularly for alcohol consumption. It is unclear whether this is due to shared environmental factors, assortative mating or indirect genetic effects. Future studies analysing the contribution of couple effects to substance use would benefit from using longitudinal data to better understand how behaviours change as individuals enter relationships and larger family samples with more couple pairs are needed to model the effect of couple's genotype alongside an individual's own genetic effects. It is important to understand the effect of assortative mating on substance use as it increases the likelihood of children inheriting any genetic risk for substance use disorders alongside the additional impact of an adverse early environment from two parents with substance use problems [18]. If substance use behaviours converge to cause spousal similarities then this is a potential modifiable risk factor to consider when addressing substance abuse as targeted interventions can be developed for vulnerable individuals. Given the magnitude of the couple associations reported here, it might be worthwhile to consider the substance use of someone's partner when any interventions to reduce intake are implemented.

Author contributions AC, AM, BHS, CH, DP, GD, MJA and SP contributed to the data acquisition, quality control and processing of the samples for this study. Manipulation of genetic data and quality control was performed by TKC, CH, MJA, DMH, CX and GD. TKC and CX were responsible for the study concept and design. Data analysis and interpretation of findings were performed by TKC, CX and AMM. The manuscript was drafted by TKC. IJD, AMM and CX provided critical revision of the manuscript for important intellectual content. All authors critically approved the content of the manuscript and approved the final version for publication.

Compliance with ethical standards

References

1. Heath AC, Martin NG. Genetic influences on alcohol consumption patterns and problem drinking: results from the Australian NH&MRC twin panel follow-up survey. Ann N Y Acad Sci. 1994;708:72–85.
2. Mbarek H, Milaneschi Y, Fedko IO, Hottenga JJ, de Moor MH, Jansen R, et al. The genetics of alcohol dependence: twin and SNP-based heritability, and genome-wide association study based on AUDIT scores. Am J Med Genet B Neuropsychiatr Genet. 2015;168:739–48.
3. Prescott CA, Kendler KS. Genetic and environmental contributions to alcohol abuse and dependence in a population-based sample of male twins. Am J Psychiatry. 1999;156:34–40.
4. Samek DR, Keyes MA, Hicks BM, Bailey J, McGue M, Iacono WG. General and specific predictors of nicotine and alcohol dependence in early adulthood: genetic and environmental influences. J Stud Alcohol Drugs. 2014;75:623–34.
5. Lessov-Schlaggar CN, Pang Z, Swan GE, Guo Q, Wang S, Cao W, et al. Heritability of cigarette smoking and alcohol use in Chinese male twins: the Qingdao twin registry. Int J Epidemiol. 2006;35:1278–85.
6. Vink JM, Willemsen G, Boomsma DI. Heritability of smoking initiation and nicotine dependence. Behav Genet. 2005;35:397–406.
7. Dawson DA, Grant BF, Ruan WJ. The association between stress and drinking: modifying effects of gender and vulnerability. Alcohol Alcohol. 2005;40:453–60.
8. Lloyd DA, Turner RJ. Cumulative lifetime adversities and alcohol dependence in adolescence and young adulthood. Drug Alcohol Depend. 2008;93:217–26.
9. Kendler KS, Schmitt E, Aggen SH, Prescott CA. Genetic and environmental influences on alcohol, caffeine, cannabis, and nicotine use from early adolescence to middle adulthood. Arch Gen Psychiatry. 2008;65:674–82.
10. Long EC, Verhulst B, Aggen SH, Kendler KS, Gillespie NA. Contributions of genes and environment to developmental change in alcohol use. Behav Genet. 2017;47:498–506.
11. Samek DR, Keyes MA, Iacono WG, McGue M. Peer deviance, alcohol expectancies, and adolescent alcohol use: explaining shared and nonshared environmental effects using an adoptive sibling pair design. Behav Genet. 2013;43:286–96.
12. Viken RJ, Kaprio J, Koskenvuo M, Rose RJ. Longitudinal analyses of the determinants of drinking and of drinking to intoxication in adolescent twins. Behav Genet. 1999;29:455–61.
13. Rose RJ, Dick DM, Viken R, Pulkkinen L, Kaprio J. Drinking or abstaining at age 14? A genetic epidemiological study. Alcohol Clin Exp Res. 2001;25:1594–604.
14. Kendler KS, Maes HH, Sundquist K, Ohlsson H, Sundquist J. Genetic and family and community environmental effects on drug abuse in adolescence: a Swedish national twin and sibling study. Am J Psychiatry. 2014;171:209–17.
15. Hall RL, Hesselbrock VM, Stabenau JR. Familial distribution of alcohol use: II. Assortative mating of alcoholic probands. Behav Genet. 1983;13:373–82.
16. Sakai JT, Stallings MC, Mikulich-Gilbertson SK, Corley RP, Young SE, Hopfer CJ, et al. Mate similarity for substance dependence and antisocial personality disorder symptoms among parents of patients and controls. Drug Alcohol Depend. 2004;75:165–75.
17. Agrawal A, Heath AC, Grant JD, Pergadia ML, Statham DJ, Bucholz KK, et al. Assortative mating for cigarette smoking and for alcohol consumption in female Australian twins and their spouses. Behav Genet. 2006;36:553–66.
18. Grant JD, Heath AC, Bucholz KK, Madden PA, Agrawal A, Statham DJ, et al. Spousal concordance for alcohol dependence: evidence for assortative mating or spousal interaction effects? Alcohol Clin Exp Res. 2007;31:717–28.
19. Smith BH, Campbell A, Linksted P, Fitzpatrick B, Jackson C, Kerr SM, et al. Cohort Profile: generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. Int J Epidemiol. 2013;42:689–700.
20. Smith BH, Campbell H, Blackwood D, Connell J, Connor M, Deary IJ, et al. Generation Scotland: the Scottish Family Health Study; a new resource for researching genes and heritability. BMC Med Genet. 2006;7:74.
21. Nagy R, Boutin TS, Marten J, Huffman JE, Kerr SM, Campbell A, et al. Exploration of haplotype research consortium imputation for genome-wide association studies in 20,032 generation Scotland participants. Genome Med. 2017;9:23.
22. Navrady LB, Wolters MK, MacIntyre DJ, Clarke TK, Campbell AI, Murray AD, et al. Cohort Profile: Stratifying Resilience and Depression Longitudinally (STRADL): a questionnaire follow-up

of Generation Scotland: Scottish Family Health Study (GS:SFHS). Int J Epidemiol. 2018;47:13–14g.

23. Ewing JA. Detecting alcoholism. The CAGE questionnaire. JAMA. 1984;252:1905–7.

24. Allen N, Sudlow C, Downey P, Peakman T, Danesh J, Elliott P, et al. UK Biobank: current status and what it means for epidemiology. Health Policy Technol. 2012;1:123–6.

25. Bycroft C, Freeman C, Petkova D, et al.: Genome-wide genetic data on ~500,000 UK Biobank participants. bioRxiv 166298. https://www.biorxiv.org/content/early/2017/07/20/166298.

26. Clarke TK, Adams MJ, Davies G, Howard DM, Hall LS, Padmanabhan S, et al. Genome-wide association study of alcohol consumption and genetic overlap with other health-related traits in UK Biobank (N=112 117). Mol Psychiatry. 2017;22:1376–84.

27. Babor TF. The alcohol use disorders identification test—guidelines for use in primary care. Geneva, Switzerland: World Health Organisation–Department of Mental Health and Substance Dependence; 2001. 2nd edn.

28. Sanchez-Roige S, Palmer AA, Fontanillas P, Elson SL, Adams MJ, Howard DM et al. Genome-wide association study meta-analysis of the Alcohol Use Disorder Identification Test (AUDIT) in two population-based cohorts (*N* = 141,958). 2018. https://doi.org/10.1101/275917.

29. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 2011;88:76–82.

30. Xia C, Amador C, Huffman J, Trochet H, Campbell A, Porteous D, et al. Pedigree- and SNP-associated genetics and recent environment are the major contributors to anthropometric and cardiometabolic trait variation. PLoS Genet. 2016;12:e1005804.

31. Euesden J, Lewis CM, O'Reilly PF. PRSice: polygenic risk score software. Bioinformatics. 2015;31:1466–8.

32. Hothorn T, Hornik K, van de Wiel MAV, Zeileis A. Implementing a class of permutation tests: the coin package. J Stat Softw. 2008;28:1–23.

33. Edenberg HJ, McClintick JN. Alcohol dehydrogenases, aldehyde dehydrogenases, and alcohol use disorders: a critical review. Alcohol Clin Exp Res. 2018;42:2281–97.

34. Walters R. Trans-ancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. 2018. https://doi.org/10.1101/257311.

35. Sanchez-Roige S, Fontanillas P, Elson SL, and Me Research T, Gray JC, de Wit H et al. Genome-wide association study of alcohol use disorder identification test (AUDIT) scores in 20 328 research participants of European ancestry. Addict Biol. 2019;24:121–31.

36. Zaitlen N, Kraft P, Patterson N, Pasaniuc B, Bhatia G, Pollack S, et al. Using extended genealogy to estimate components of heritability for 23 quantitative and dichotomous traits. PLoS Genet. 2013;9:e1003520.

37. Zuk O, Hechter E, Sunyaev SR, Lander ES. The mystery of missing heritability: genetic interactions create phantom heritability. Proc Natl Acad Sci USA. 2012;109:1193–8.

38. Ge T, Chen CY, Neale BM, Sabuncu MR, Smoller JW. Phenome-wide heritability analysis of the UK Biobank. PLoS Genet. 2017;13:e1006711.

39. Schuckit MA, Smith TL. The relationships of a family history of alcohol dependence, a low level of response to alcohol and six domains of life functioning to the development of alcohol use disorders. J Stud Alcohol. 2000;61:827–35.

40. Grant BF, Stinson FS, Harford TC. Age at onset of alcohol use and DSM-IV alcohol abuse and dependence: a 12-year follow-up. J Subst Abus. 2001;13:493–504.

41. van Beek JH, de Moor MH, Geels LM, Willemsen G, Boomsma DI. Explaining individual differences in alcohol intake in adults: evidence for genetic and cultural transmission? J Stud Alcohol Drugs. 2014;75:201–10.

42. Slomkowski C, Rende R, Novak S, Lloyd-Richardson E, Niaura R. Sibling effects on smoking in adolescence: evidence for social influence from a genetically informative design. Addiction. 2005;100:430–8.

43. Leonardi-Bee J, Jere ML, Britton J. Exposure to parental and sibling smoking and the risk of smoking uptake in childhood and adolescence: a systematic review and meta-analysis. Thorax. 2011;66:847–55.

44. Li MD, Cheng R, Ma JZ, Swan GE. A meta-analysis of estimated genetic and environmental effects on smoking behavior in male and female adult twins. Addiction. 2003;98: 23–31.

45. Heath AC, Kirk KM, Meyer JM, Martin NG. Genetic and social determinants of initiation and age at onset of smoking in Australian twins. Behav Genet. 1999;29:395–407.

46. Kong A, Thorleifsson G, Frigge ML, Vilhjalmsson BJ, Young AI, Thorgeirsson TE, et al. The nature of nurture: effects of parental genotypes. Science. 2018;359:424–8.

47. Yengo L, Robinson MR, Keller MC, Kemper KE, Yang Y, Trzaskowski M, et al. Imprint of assortative mating on the human genome. Nat Human Behaviour. 2018;2:948–54.