Smoking and caffeine consumption: a genetic analysis of their association

Jorien L. Treur1,2, Amy E. Taylor3,4, Jennifer J. Ware4,5, Michel G. Nivard1,6, Michael C. Neale7, George McMahon4,5, Jouke-Jan Hottenga1,2,6, Bart M. L. Baselmans1,2, Dorret I. Boomsma1,2,6, Marcus R. Munafò3,4† & Jacqueline M. Vink1,2,6†

Department of Biological Psychology, VU University Amsterdam, Amsterdam, The Netherlands1, EMGO+ Institute for Health and Care Research, VU University Medical Center, Amsterdam, The Netherlands2, UK Centre for Tobacco and Alcohol Studies, School of Experimental Psychology, University of Bristol, Bristol, UK3, MRC Integrative Epidemiology Unit at the University of Bristol, Bristol, UK4, School of Social and Community Medicine, University of Bristol, Bristol, UK5, Neuroscience Campus Amsterdam, VU University Medical Center, Amsterdam, The Netherlands6 and Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA USA7

ABSTRACT

Smoking and caffeine consumption show a strong positive correlation, but the mechanism underlying this association is unclear. Explanations include shared genetic/environmental factors or causal effects. This study employed three methods to investigate the association between smoking and caffeine. First, bivariate genetic models were applied to data of 10,368 twins from the Netherlands Twin Register in order to estimate genetic and environmental correlations between smoking and caffeine use. Second, from the summary statistics of meta-analyses of genome-wide association studies on smoking and caffeine, the genetic correlation was calculated by LD-score regression. Third, causal effects were tested using Mendelian randomization analysis in 6605 Netherlands Twin Register participants and 5714 women from the Avon Longitudinal Study of Parents and Children. Through twin modelling, a genetic correlation of $r = 0.47$ and an environmental correlation of $r = 0.30$ were estimated between current smoking (yes/no) and coffee use (high/low). Between current smoking and total caffeine use, this was $r = 0.44$ and $r = 0.00$, respectively. LD-score regression also indicated sizeable genetic correlations between smoking heaviness and cups of coffee per day, $r = 0.44$ between smoking initiation and coffee use and $r = 0.25$ between smoking persistence and coffee use. Consistent with the relatively high genetic correlations and lower environmental correlations, Mendelian randomization provided no evidence for causal effects of smoking on caffeine or vice versa. Genetic factors thus explain most of the association between smoking and caffeine consumption. These findings suggest that quitting smoking may be more difficult for heavy caffeine consumers, given their genetic susceptibility.

Keywords ALSPAC, caffeine, LD-score regression, Mendelian randomization, smoking, twin modelling.

Correspondence to: Jorien L. Treur, Department of Biological Psychology, VU University Amsterdam, Van der Boechorststraat 1, 1081 BT Amsterdam, The Netherlands. E-mail: j.l.treur@vu.nl

INTRODUCTION

A recent study in two large European cohorts (one Dutch and one British) showed a strong positive correlation between smoking behaviour and caffeine consumption (Treur et al. 2016). When analyzing coffee, tea, cola and energy drinks separately, the strongest associations were found for coffee. Others have also demonstrated that smoking is associated with an increased consumption of coffee (Swanson, Lee, & Hopp 1994; Freedman et al. 2012) and of total caffeine (Hewlett & Smith 2006).

Smoking and caffeine consumption are both influenced by genetic factors. In a Dutch twin study, 44% of individual differences in smoking initiation were explained by genetic factors, while for nicotine dependence, this was 75% (Vink, Willemse, & Boomsma 2005). The heritability estimates for coffee/caffeine use range from 36% to 77% (see overview in Vink, Staphorsius, &
Strong associations between smoking and caffeine use may therefore be the result of shared genetic and/or environmental factors. The bivariate twin model estimates genetic and environmental influences on two traits and on their overlap by comparing the resemblance in monozygotic (MZ) twin pairs to the resemblance in dizygotic (DZ) twin pairs. Few studies have applied this model to smoking and caffeine consumption. In American male twin pairs from the ‘Twin Registry of White male World War II veterans’, the observational association between number of cigarettes smoked per day, number of alcoholic drinks per week and number of cups of coffee per day was completely due to genetic factors (Swan, Carmelli, & Cardon 1996), while both genetic and non-shared (unique) environmental factors contributed to the association between heavy smoking and heavy coffee drinking (Swan, Carmelli, & Cardon 1997). In men and women from the ‘Virginia Twin Registry’, number of cigarettes smoked per day, alcoholic drinks per week and total caffeine consumption were associated because of genetic and unique environmental factors (Hettema, Corey, & Kendler 1999). In male twin pairs from the same registry, the common environment that is shared by twins explained the correlation between caffeine consumption and cigarettes per day in adolescence. As participants aged, these common environmental influences gradually decreased to zero, and genetic influences increased (Kendler et al. 2008).

Recently, a novel technique to estimate genetic correlation between two traits, LD (linkage disequilibrium)-score regression (Bulik-Sullivan et al. 2015b), was developed. This method utilizes the effect size estimates of all included single nucleotide polymorphisms (SNPs) in genome-wide association (GWA) meta-analyses, to estimate genetic correlation between two traits. When calculating correlations among 25 phenotypes (ranging from schizophrenia to coronary artery disease), results were similar to genetic correlations estimated with individual genotype data (Bulik-Sullivan et al. 2015a). To our knowledge, LD-score regression has not yet been applied to GWA meta-analyses on smoking (TAG 2010) and caffeine use (Cornelis et al. 2014).

The presence of both genetic and environmental correlations is consistent with causal effects underlying the association between traits (De Moor et al. 2008). Experimental work in animals and humans has provided evidence for causal effects of smoking on caffeine use (Joeres et al. 1988; Langmann et al. 2000; Benowitz, Peng, & Jacob 2003) and of caffeine use on smoking (Shoaib et al. 1999; Gasior et al. 2002; Rezvani et al. 2013). Mendelian randomization (MR) analysis can be employed to test causality (Davey Smith & Ebrahim 2003; Palmer et al. 2012; Davey Smith & Hemani 2014). MR utilizes one or several genetic variants robustly associated with a certain trait as an ‘instrument’, or proxy, for that same trait. Because of the random nature of genetic assortment, variants that are associated with a particular trait should not be associated with confounding factors. Furthermore, outcome measures cannot affect the genes that an individual is born with, removing the possibility of reverse causation. Bidirectional MR, where the effect of a genetic variant for heaviness of smoking (TAG 2010) on caffeine consumption, and the effect of eight genetic variants for caffeine consumption (Cornelis et al. 2014) on smoking behaviour, is tested, could unravel a possible causal association between smoking and caffeine.

We used these three methods to clarify the nature of the association between smoking and caffeine use. First, in a large sample of 10 368 twins from the Netherlands Twin Register (NTR), bivariate genetic models were applied to data on smoking and caffeine consumption. Second, genetic correlation between smoking and caffeine was computed with LD-score regression, utilizing data from recent GWA meta-analyses. Third, in a sample of 6605 participants from the NTR and 5714 from the Avon Longitudinal Study of Parents and Children (ALSPAC), causal effects were tested using bidirectional MR.

**MATERIALS AND METHODS**

Figure 1 provides an overview of the three methods, with the corresponding aims, study samples/data, smoking measures, caffeine measures and statistical analyses. More detailed information on each of the approaches is provided as follows.

**Study samples/data**

**The Netherlands Twin Register**

The NTR is an ongoing longitudinal study of Dutch twins and their family members (Willemse et al. 2013). The 10th NTR survey sent in 2013–2014 contained questions on an extensive list of caffeinated and decaffeinated drinks (including coffee, tea, cola and energy drinks) and on smoking behaviour. The fifth NTR survey sent in 2000 contained questions on (caffeinated) coffee consumption and smoking. For those who did not complete the 10th survey, data from the fifth survey were included. NTR participants were included in the bivariate twin modelling analyses (method 1) and in the MR analyses (method 3).

A total of 10 368 twins with data on smoking and coffee use were available for bivariate modelling {mean age = 32.5 [standard deviation (SD) 14.5], 68.1% female, 6866 twins from complete pairs and 3502 twins from
incomplete pairs). Of this total, 1425 twins were monozygotic male (MZM), 907 dizygotic male (DZM), 3541 monozygotic female (MZF), 1948 dizygotic female (DZF) and 2547 dizygotic opposite sex (DOS). For 8060 twins, there were data on smoking and total caffeine use (mean age = 33.7 [SD 15.3], 69.0% female, 4778 twins from complete pairs and 3282 twins from incomplete pairs). In this group, there were 1089 MZM, 677 DZM, 2837 MZF, 1498 DZF and 1959 DOS twins.

For MR analysis, phenotype and genotype data were available in up to 6605 NTR participants (mean age = 42.7 [SD 16.7], 66.4% female), of which 1415 had data on smoking and coffee only.

The Avon Longitudinal Study of Parents and Children

The Avon Longitudinal Study of Parents and Children is a prospective cohort study that recruited 14,541 pregnant women who resided in the county of Avon in the UK and who had expected delivery dates ranging between 1 April 1991 and 31 December 1992. The ALSPAC Ethics and Law Committee and the Local Research Ethics Committees gave ethical approval for the study. An elaborate account of ALSPAC and its methods is given elsewhere (Boyd et al. 2012; Fraser et al. 2013). Questions on smoking and caffeinated and decaffeinated coffee, tea and cola consumption were asked in surveys sent to the mothers during pregnancy at 18 weeks gestation, 32 weeks gestation and after delivery when the child was aged 2, 47, 85, 97 and 145 months. Analyses were performed for all timepoints, but only the results from when the child was 47 months are reported. This specific timepoint was selected because the sample size was large and because during and immediately after pregnancy, smoking behaviour and caffeine use may be different. Please note that the study website contains details of all the data that are available through a fully searchable data dictionary (www.bris.ac.uk/alspac/researchers/data-access/data-dictionary).

Participants registered at ALSPAC were included in the MR analyses (method 3). Phenotype and genotype data of 5714 participants were available (mean age = 33.4 [SD 4.5]).

Genotype data

Single nucleotide polymorphism data were available from genome-wide SNP arrays. In the NTR sample, these data were collected through several projects between 2004 and 2008. Adult participants (18+) who had participated in NTR research at least once were selected for genotyping. Full details on the data collection and genotyping methods in NTR can be found elsewhere (Willemsen et al. 2010; Nivard et al. 2014). In ALSPAC, DNA was extracted from blood samples that were collected from 10,000 mothers during their normal antenatal care. More details on the collection of DNA and genotyping methods in ALSPAC are provided elsewhere (Jones et al. 2000; Hinds et al. 2013). Genotype data were utilized in the MR analyses (method 3).

Genetic risk for smoking was reflected by SNP rs1051730, located in the CHRNA3 gene at chromosome 15 and robustly and consistently associated with smoking heaviness (TAG 2010; Thorgeirsson et al. 2010). This SNP is in high LD with rs16969968 (TAG 2010). For caffeine use, a genetic risk score was created based on eight SNPs that reached genome-wide significance in their association with coffee consumption in a large meta-analysis (Cornelis et al. 2014). The number of coffee consuming increasing alleles at each locus was summed across all eight variants for each individual. Alleles were weighted according to the magnitude of the effect size (β) for coffee consumption, taken from the recent meta-analysis by Cornelis et al. (2014).

Table 1 provides more detailed information on all SNPs, including the risk alleles, frequencies of these risk alleles and effect sizes.
for men and women separately. This cut-off point was determined from coffee users (0 = low users, 1 = high users). More details on the making of these variables are included. From these questions, daily total caffeine consumption (in mg) and daily caffeine consumption in NTR and ALSPAC (Treur et al. 2016). For the current paper, variables were defined that reflect smoking initiation (0 = never smokers, 1 = former smokers, current smoking (0 = never and former smokers, 1 = current smokers) and smoking persistence (0 = former smokers, 1 = current smokers). Smoking measures

All NTR and ALSPAC participants were classified as current smokers, former smokers or never smokers. A variable on smoking heaviness (cigarettes smoked per day) was available for current smokers in both cohorts. A more detailed explanation of these variables is available in a publication on the observational associations between smoking and caffeine consumption in NTR and ALSPAC (Treur et al. 2016). For the current paper, variables were defined that reflect smoking initiation (0 = never smokers, 1 = former smokers, current smoking (0 = never and former smokers, 1 = current smokers) and smoking persistence (0 = former smokers, 1 = current smokers).

Caffeine measures

In both NTR and ALSPAC, questions were asked about the consumption of caffeinated coffee, tea and cola, while in the NTR, an additional question on energy drinks was included. From these questions, daily total caffeine consumption (in mg) and daily caffeine consumption through coffee (in mg) were calculated for all participants. More details on the making of these variables are given elsewhere (Treur et al. 2016). For bivariate twin modelling (method 1; NTR data only), a dichotomous variable was created where 1 SD above the mean was chosen as a cut-off point, distinguishing ‘low’ from ‘high’ coffee users (0 = low [≤ 1 SD above the mean, N = 8599], 1 = high [> 1 SD above the mean, N = 1769]) and ‘low’ from ‘high’ total caffeine users (N = 6863 and N = 1197, respectively). This cut-off point was determined for men and women separately.

Table 1 SNPs utilized in Mendelian randomization analysis.

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chr</th>
<th>Closest gene</th>
<th>RA/NRA</th>
<th>Effect size</th>
<th>RAF</th>
<th>Genotyped/imputed</th>
<th>Imputation quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>NTR</td>
<td>ALSPAC</td>
<td>NTR</td>
<td>ALSPAC</td>
<td>NTR</td>
<td>ALSPAC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1051730</td>
<td>15</td>
<td>CHRNA3</td>
<td>A/G</td>
<td>1.03</td>
<td>0.32</td>
<td>0.33</td>
<td>G</td>
</tr>
<tr>
<td>rs1260326</td>
<td>2</td>
<td>GCKR</td>
<td>C/T</td>
<td>0.04</td>
<td>0.63</td>
<td>0.60</td>
<td>G</td>
</tr>
<tr>
<td>rs1481012</td>
<td>4</td>
<td>ABCG2</td>
<td>A/G</td>
<td>0.06</td>
<td>0.89</td>
<td>0.90</td>
<td>G</td>
</tr>
<tr>
<td>rs6968554</td>
<td>7</td>
<td>AHR</td>
<td>G/A</td>
<td>0.13</td>
<td>0.64</td>
<td>0.64</td>
<td>G</td>
</tr>
<tr>
<td>rs7800944</td>
<td>7</td>
<td>MLXIP1</td>
<td>C/T</td>
<td>0.05</td>
<td>0.27</td>
<td>0.29</td>
<td>I</td>
</tr>
<tr>
<td>rs17685</td>
<td>7</td>
<td>POR</td>
<td>A/G</td>
<td>0.07</td>
<td>0.26</td>
<td>0.26</td>
<td>G</td>
</tr>
<tr>
<td>rs6265</td>
<td>11</td>
<td>BDNF</td>
<td>C/T</td>
<td>0.05</td>
<td>0.80</td>
<td>0.81</td>
<td>G</td>
</tr>
<tr>
<td>rs2472297</td>
<td>15</td>
<td>CYP1A1</td>
<td>T/C</td>
<td>0.15</td>
<td>0.27</td>
<td>0.27</td>
<td>G</td>
</tr>
<tr>
<td>rs9902453</td>
<td>17</td>
<td>EFCAB5</td>
<td>G/A</td>
<td>0.04</td>
<td>0.49</td>
<td>0.45</td>
<td>I</td>
</tr>
</tbody>
</table>

Effect sizes represent β coefficients and were obtained from TAG (2010) and Cornelis et al. (2014). ALSPAC, Avon Longitudinal Study of Parents and Children; Chr, chromosome; RA, risk allele (smoking or coffee consumption increasing allele); NRA, non-risk allele; NTR, Netherlands Twin Register; RAF, risk allele frequency in the total study sample of ALSPAC and NTR participants; SNP, single nucleotide polymorphism.

Statistical analyses

Bivariate genetic modelling

The bivariate twin model estimates the influence of additive genetic effects (A), common environmental effects shared by twins from the same family (C) and unique environmental effects (E) on smoking and caffeine use, as well as how much of the correlation between smoking and caffeine use is due to A, C and E. Briefly, MZ twins are 100% genetically similar, while DZ twins share ~50% of their segregating genes; both types of twins may share their environment. In the case of one trait, a higher resemblance between MZ twins than between DZ twins indicates an influence of additive genetic factors (A). If the correlation between DZ twins is greater than half the correlation between MZ twins, the common environment that is shared by both twins (C) is also of influence. When the correlation between MZ twins is lower than 1, this must be due to unique environmental factors (E). The influence of genes and environment on the correlation between smoking and caffeine (bivariate) is deduced from the correlation between smoking in twin 1 and caffeine in twin 2. When this ‘cross-correlation’ is higher in MZ than in DZ twin pairs, an influence of A is implied. When the DZ cross-correlation is greater than half the MZ cross-correlation, the influence of C is suggested. When the MZ cross-correlation is lower than the correlation between smoking and caffeine in one person, an influence of E is implied. For more elaborate descriptions and comparable bivariate twin designs, see, e.g. Kendler (2013); Poehlman et al. (2003) and Poehlman et al. (2008).
To estimate genetic and environmental influences, bivariate structural equation modelling was performed in OpenMx (Boker et al. 2011). There were two models, one with current smoking (0 = never and former smokers, 1 = current smokers) and coffee use (0 = low, 1 = high) and one with current smoking and total caffeine use (0 = low, 1 = high). In these so-called liability threshold models, an underlying liability resulting from genetic and environmental factors is assumed. A threshold divides individuals into current smokers and non-current smokers and into high and low caffeine users. The thresholds depend on the prevalence of current smoking and high caffeine use, respectively (Falconer & Mackay 1996; Wray & Visscher 2015). Age was included as a moderator on the thresholds, allowing prevalence to differ with age (categories: <20, 20–24, 25–34, 35–44, 45–54 and ≥55 years).

The first step of genetic modelling was to fit a bivariate saturated model to data from five sex-by-zygosity groups (MZM, DZM, MZF, DZF and DOS twin pairs). Next, the effects of A, C and E on smoking, coffee/caffeine and the genetic and environmental correlations were estimated in a bivariate ACE model. Several constraints were imposed during model fitting, which are described in the Results section. The fit of submodels was tested with likelihood ratio tests, following a χ² distribution where the amount of d.f. (degrees of freedom) is equal to the difference in d.f. of the two models. Constraints were retained when they did not significantly deteriorate the fit (P-value ≥0.05).

LD-score regression

Genome-wide association meta-analysis results were available for cigarettes smoked per day, smoking initiation (0 = never smokers, 1 = former and current smokers) and smoking cessation (0 = current smoking, 1 = former smoking) (TAG 2010) and for cups of coffee per day (Cornelis et al. 2014). Smoking cessation was defined as smoking persistence (0 = former smokers, 1 = current smokers) by multiplying the genetic correlation by –1. The meta-analyses on smoking and coffee included GWA studies of 16 and 28 population-based samples of European ancestry, including up to 46,481 individuals and 91,462 individuals, respectively. Findings are thus not restricted, or specific, to one single population.

Genetic correlations were estimated using LD-score regression (Bulik-Sullivan et al. 2015a). The intuitive concept behind this technique is that for highly polygenic traits, SNPs that tag many neighbouring SNPs due to strong LD have a higher chance of tagging a causal locus. In contrast, SNPs that are in relatively weak LD with their neighbours tag fewer causal loci. One can therefore formulate the expected effect size for a SNP in a GWAS as a function of the degree of LD, sample size in a GWAS, number of SNPs considered and the heritability. To estimate genetic correlation, the effect size estimates of all SNPs included in GWAS of two phenotypes of interest are utilized. First, the association between a particular SNP and phenotype 1 (represented by a Z score) is multiplied by the association between that same SNP and phenotype 2. Second, the product thereof is regressed on the LD that the SNP has with all neighbouring SNPs (i.e. the LD score). As such, it is possible to estimate genetic correlation between two traits solely based on observed summary statistics. We used precomputed LD scores based on meta-analyses of individuals of European ancestry that are publicly available (from: https://github.com/bulik/ldsc). LD-score regression can be utilized even when there is sample overlap, because effect size inflation due to sample overlap will equally impact all SNPs, regardless of their LD score, and inflation due to sample overlap inflates the intercept, not the slope.

Mendelian randomization analysis

By measuring genetic variants strongly predictive of smoking and caffeine use instead of these behaviours themselves, MR minimizes effects of confounding and reverse causation (Fig. 2). In Stata (version 9.0; StataCorp...
For current smoking and total caffeine use, similar submodels were applied (Table 2-B; models 8 till 15), resulting in a best-fitting model without any influence of C and without E for the overlap.

Table 4 depicts parameter estimates for the full ACE and best-fitting AE models. Variation in current smoking was mostly due to additive genetic factors (ACurrentSmoking = 76% [70–79%]) in a bivariate model with coffee and ACurrentSmoking = 74% [66–80%] with total caffeine) with the remainder being due to unique environmental factors (ECurrentSmoking = 24% [19–30%] and ECurrentSmoking = 26% [20–34%], respectively). About half of the variation in coffee use was due to additive genetic factors (ACoffee = 53% [48–58%]) and the other half to unique environmental factors (ECoffee = 47% [40–54%]). For total caffeine, results were similar (ATotalCaffeine = 49% [47–58%] and ETotalCaffeine = 51% [42–60%]). See Figs S1 and S2 for a graphical representation of these models, with all parameter estimates.

A genetic correlation of r = 0.47 (0.38 to 0.56) was found between smoking and coffee and of r = 0.44 (0.35 to 0.53) between smoking and total caffeine. There was a unique environmental correlation of r = 0.30 (0.15 to 0.45) between smoking and coffee. The unique environmental correlation between smoking and total caffeine was r = 0.00 (0.00 to 0.00), meaning that all correlation between the two traits was attributed to other sources (namely genetic). Phenotypic overlap between current smoking and coffee was mostly due to additive genetic factors (ACurrentSmoking-Coffee = 75% [62–87%]), with some influence of unique environmental factors (ECurrentSmoking-Coffee = 25% [13–38%]). Between current smoking and total caffeine, the overlap was completely due to genetic factors.

**LD-score regression**

The genetic correlation between cigarettes per day and cups of coffee per day, as calculated by LD-score regression, was r = 0.44 (0.14 to 0.74). This confirms the results from the bivariate twin models of the same genetic risk factors influencing current smoking and high coffee/total caffeine consumption. Between smoking initiation and cups of coffee per day, the genetic correlation was lower at r = 0.28 (0.11 to 0.45), and between smoking persistence and cups of coffee per day, it was r = 0.25 (0.04 to 0.46).

**Mendelian randomization analysis**

As expected, the caffeine use SNP score was strongly associated with a higher coffee and total caffeine consumption. For coffee, for smoking, or for the overlap, and E for the overlap resulted in a significant deterioration of the fit (models 12 till 15), and these effects were thus retained.
Table 2 Structural equation models to explore additive genetic (A), common environmental (C) and unique environmental (E) influences on current smoking and caffeine use and on their overlap.

### A: Current smoking and coffee use (N = 10 368)

<table>
<thead>
<tr>
<th>Model Description</th>
<th>Estimated parameters</th>
<th>-2LL</th>
<th>d.f.</th>
<th>Compared with</th>
<th>χ²</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Saturated five-group model</td>
<td>46</td>
<td>17 021.62</td>
<td>20</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>2. β's covariate dropped</td>
<td>42</td>
<td>17 850.72</td>
<td>20</td>
<td>1</td>
<td>829.11</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>3. Thresholds/β's constrained across sex for smoking</td>
<td>41</td>
<td>17 041.36</td>
<td>20</td>
<td>1</td>
<td>19.74</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>4. Thresholds/β's constrained across sex for coffee</td>
<td>41</td>
<td>17 126.56</td>
<td>20</td>
<td>1</td>
<td>104.95</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>5. Cross-trait-within twin correlation constrained across twin birth order</td>
<td>41</td>
<td>17 027.16</td>
<td>20</td>
<td>1</td>
<td>5.55</td>
<td>0.35</td>
</tr>
<tr>
<td>6. Cross-trait cross-twin correlation constrained across twin birth order</td>
<td>36</td>
<td>17 032.36</td>
<td>20</td>
<td>5</td>
<td>5.2</td>
<td>0.39</td>
</tr>
<tr>
<td>7. Correlations MZM = MZF + correlations DZM = DZF = DOS</td>
<td>24</td>
<td>17 043.87</td>
<td>20</td>
<td>6</td>
<td>11.5</td>
<td>0.49</td>
</tr>
<tr>
<td>8. ACE model</td>
<td>25</td>
<td>17 046.54</td>
<td>20</td>
<td>1</td>
<td>24.92</td>
<td>0.47</td>
</tr>
<tr>
<td>9. C for coffee dropped</td>
<td>24</td>
<td>17 048.9</td>
<td>20</td>
<td>8</td>
<td>2.36</td>
<td>0.12</td>
</tr>
<tr>
<td>10. C for smoking dropped</td>
<td>23</td>
<td>17 048.9</td>
<td>20</td>
<td>9</td>
<td>0</td>
<td>1.0</td>
</tr>
<tr>
<td>11. C for overlap dropped</td>
<td>22</td>
<td>17 048.99</td>
<td>20</td>
<td>10</td>
<td>0.09</td>
<td>0.76</td>
</tr>
<tr>
<td>12. A for coffee dropped</td>
<td>21</td>
<td>17 256.83</td>
<td>20</td>
<td>11</td>
<td>207.84</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>13. A for smoking dropped</td>
<td>21</td>
<td>17 148.73</td>
<td>20</td>
<td>11</td>
<td>99.74</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>14. A for overlap dropped</td>
<td>21</td>
<td>17 143.17</td>
<td>20</td>
<td>11</td>
<td>94.18</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>15. E for overlap dropped</td>
<td>21</td>
<td>17 064.68</td>
<td>20</td>
<td>11</td>
<td>15.69</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

### B: Current smoking and total caffeine use (N = 8060)

<table>
<thead>
<tr>
<th>Model Description</th>
<th>Estimated parameters</th>
<th>-2LL</th>
<th>d.f.</th>
<th>Compared with</th>
<th>X²</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Saturated five-group model</td>
<td>46</td>
<td>12 194.18</td>
<td>16</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>2. β's covariate dropped</td>
<td>42</td>
<td>12 761.54</td>
<td>16</td>
<td>1</td>
<td>567.36</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>3. Thresholds/β's constrained across sex for smoking</td>
<td>41</td>
<td>12 209.66</td>
<td>16</td>
<td>1</td>
<td>15.48</td>
<td>0.01</td>
</tr>
<tr>
<td>4. Thresholds/β's constrained across sex for total caffeine</td>
<td>41</td>
<td>12 207.36</td>
<td>16</td>
<td>1</td>
<td>13.18</td>
<td>0.02</td>
</tr>
<tr>
<td>5. Cross-trait-within twin correlation constrained across twin birth order</td>
<td>41</td>
<td>12 194.68</td>
<td>16</td>
<td>1</td>
<td>0.5</td>
<td>0.99</td>
</tr>
<tr>
<td>6. Cross-trait cross-twin correlation constrained across twin birth order</td>
<td>36</td>
<td>12 195.81</td>
<td>16</td>
<td>5</td>
<td>1.13</td>
<td>0.95</td>
</tr>
<tr>
<td>7. Correlations MZM = MZF + correlations DZM = DZF = DOS</td>
<td>24</td>
<td>12 207.23</td>
<td>16</td>
<td>6</td>
<td>11.42</td>
<td>0.49</td>
</tr>
<tr>
<td>8. ACE model</td>
<td>25</td>
<td>12 209.95</td>
<td>16</td>
<td>1</td>
<td>15.77</td>
<td>0.83</td>
</tr>
<tr>
<td>9. C for total caffeine dropped</td>
<td>24</td>
<td>12 211.23</td>
<td>16</td>
<td>8</td>
<td>1.29</td>
<td>0.98</td>
</tr>
<tr>
<td>10. C for smoking dropped</td>
<td>23</td>
<td>12 211.23</td>
<td>16</td>
<td>9</td>
<td>0</td>
<td>0.98</td>
</tr>
<tr>
<td>11. C for overlap dropped</td>
<td>22</td>
<td>12 211.34</td>
<td>16</td>
<td>10</td>
<td>0.1</td>
<td>0.75</td>
</tr>
<tr>
<td>12. A for total caffeine dropped</td>
<td>21</td>
<td>12 311.73</td>
<td>16</td>
<td>11</td>
<td>100.39</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>13. A for smoking dropped</td>
<td>21</td>
<td>12 273.89</td>
<td>16</td>
<td>11</td>
<td>62.55</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>14. A for overlap dropped</td>
<td>21</td>
<td>12 240.56</td>
<td>16</td>
<td>11</td>
<td>29.23</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>15. E for overlap dropped</td>
<td>21</td>
<td>12 212.95</td>
<td>16</td>
<td>11</td>
<td>1.62</td>
<td>0.20</td>
</tr>
</tbody>
</table>

A threshold represents the prevalence of smoking or coffee/caffeine use. β = effect of age on the prevalence (threshold) of smoking or coffee/caffeine use. The best-fitting models are depicted in bold. DOS, dizygotic opposite sex twin pairs; DZF, dizygotic female twin pairs; DZM, dizygotic male twin pairs; MZE, monozygotic female twin pairs; MZM, monozygotic male twin pairs.
### Table 3
Twin correlations for current smoking and coffee use (N = 10,368) and for current smoking and total caffeine use (N = 8,060) from the best-fitting saturated models.

<table>
<thead>
<tr>
<th></th>
<th>Cross-twin within-trait</th>
<th>Cross-twin cross-trait</th>
<th>Current smoking</th>
<th>Coffee</th>
<th>Within-twin cross-trait</th>
<th>Cross-twin cross-trait</th>
<th>Current smoking</th>
<th>Total caffeine</th>
<th>Within-twin cross-trait</th>
<th>Cross-twin cross-trait</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MZ</td>
<td>0.75 (0.74 to 0.78)</td>
<td>0.52 (0.49 to 0.59)</td>
<td>0.37 (0.36 to 0.37)</td>
<td>0.26 (0.20 to 0.27)</td>
<td>0.73 (0.64 to 0.80)</td>
<td>0.47 (0.37 to 0.49)</td>
<td>0.23 (0.15 to 0.30)</td>
<td>0.18 (0.09 to 0.20)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DZ</td>
<td>0.41 (0.32 to 0.50)</td>
<td>0.32 (0.30 to 0.35)</td>
<td>0.43 (0.41 to 0.45)</td>
<td>0.21 (0.16 to 0.29)</td>
<td>0.40 (0.27 to 0.51)</td>
<td>0.30 (0.15 to 0.43)</td>
<td>0.31 (0.24 to 0.38)</td>
<td>0.18 (0.07 to 0.29)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Cross-twin within trait = correlation between smoking twin 1 and smoking twin 2 or coffee/caffeine twin 1 and coffee/caffeine twin 2. Within-twin cross-trait = correlation between smoking and coffee/caffeine in one twin. Cross-twin cross-trait = correlation between smoking twin 1 and coffee/caffeine twin 2. DZ dizygotic; MZ monozygotic.

### Table 4
Estimates of additive genetic (A), common environmental (C) and unique environmental (E) influences on current smoking and coffee use (N = 10,368) and on current smoking and total caffeine use (N = 8,060), from the full and the best-fitting bivariate genetic models.

<table>
<thead>
<tr>
<th></th>
<th>Current smoking</th>
<th>Coffee</th>
<th>Phenotypic overlap</th>
<th>Current smoking</th>
<th>Total caffeine</th>
<th>Phenotypic overlap</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0.67 (0.51 to 0.79)</td>
<td>0.40 (0.15 to 0.58)</td>
<td>0.48 (0.17 to 0.84)</td>
<td>0.64 (0.37 to 0.79)</td>
<td>0.36 (0.03 to 0.56)</td>
<td>0.44 (0.00 to 1.00)</td>
</tr>
<tr>
<td>C</td>
<td>0.08 (0.02 to 0.25)</td>
<td>0.12 (0.00 to 0.32)</td>
<td>0.24 (0.10 to 0.55)</td>
<td>0.08 (0.00 to 0.31)</td>
<td>0.12 (0.00 to 0.40)</td>
<td>0.35 (0.00 to 0.88)</td>
</tr>
<tr>
<td>E</td>
<td>0.25 (0.21 to 0.31)</td>
<td>0.48 (0.41 to 0.54)</td>
<td>0.28 (0.24 to 0.42)</td>
<td>0.28 (0.21 to 0.36)</td>
<td>0.52 (0.43 to 0.62)</td>
<td>0.21 (0.00 to 0.50)</td>
</tr>
<tr>
<td>Best-fitting model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0.76 (0.70 to 0.79)</td>
<td>0.53 (0.48 to 0.58)</td>
<td>0.75 (0.62 to 0.87)</td>
<td>0.74 (0.66 to 0.80)</td>
<td>0.49 (0.47 to 0.58)</td>
<td>1.00 (1.00 to 1.00)</td>
</tr>
<tr>
<td>C</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>E</td>
<td>0.24 (0.19 to 0.30)</td>
<td>0.47 (0.40 to 0.54)</td>
<td>0.25 (0.13 to 0.38)</td>
<td>0.26 (0.20 to 0.34)</td>
<td>0.51 (0.42 to 0.60)</td>
<td>—</td>
</tr>
<tr>
<td>Genetic correlation</td>
<td>0.47 (0.38 to 0.56)</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Unique environmental correlation</td>
<td>0.30 (0.15 to 0.45)</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

Phenotypic overlap reflects how much of the observational association between current smoking and coffee/total caffeine is due to additive genetic (A), common environmental (C) and unique environmental (E) influences, summing up to 1.
consumption in mg of caffeine per day (Fig. S3-A; β coefficient 66.5 [45.0–87.90] and 86.7 [64.4–109.1], respectively). The smoking SNP (rs1051730) was associated with more cigarettes smoked per day (Fig. S3-B; 0.6 [0.2–1.1]), confirming previous studies.

The caffeine use SNP score was not associated with educational attainment or social class, while for the smoking SNP, there was some minor evidence for a negative association in ALSPAC only (Table S3). This association is most likely spurious, as discussed elsewhere (Taylor et al. under review). Briefly, Taylor et al. found no consistent evidence for a causal, negative effect of smoking on socio-economic status when analysing the effect of the smoking SNP on different measures of socio-economic status in ALSPAC and the Nord-Trøndelag Health Study.

There was no association between the smoking SNP and caffeine consumption (total or coffee only) in current, former or never smokers (Fig. S4-A), thus providing no support for a causal effect of smoking on caffeine. Also, none of the pooled analyses showed an association between the caffeine use SNP score and smoking behaviour, meaning that there is no evidence for a causal effect of caffeine use on smoking behaviour (Fig. S4-B). See Tables S4 and S5 for MR analyses at all timepoints in ALSPAC.

DISCUSSION

The association between smoking and caffeine consumption was investigated with three methods: bivariate twin modelling, LD-score regression and MR. All three sets of analyses pointed to a similar conclusion of shared genetic factors explaining the phenotypic overlap between smoking and caffeine use. The lack of evidence for causal effects between smoking and caffeine use could have been due to low power.

It is the first time that the association between smoking and caffeine use was investigated in Dutch twins. Our results are in line with earlier US-based twin studies that found genetic correlation between smoking and caffeine use (Swan et al. 1996, 1997; Hettema et al. 1999; Kendler et al. 2008). Apart from one (Swan et al. 1996), these studies also found unique environmental correlation between smoking and caffeine use (Swan et al. 1997; Hettema et al. 1999; Kendler et al. 2008) as we did for smoking and coffee use. For smoking and total caffeine use, we did not find unique environmental correlation. In the present study, we analysed caffeine as a continuous measure (mg per day). The results of these analyses were very similar with the exception that there was, in addition to a genetic correlation, also a unique environmental correlation for both smoking and coffee and smoking and total caffeine (Tables S3–S5). The fit of this alternative model was poor when compared with the fully saturated model, probably because the continuous measure of caffeine was severely (right) skewed. In spite of the variation in measures of smoking and caffeine use, the present study corroborates previous twin studies. The genetic correlations of r0.47 (for coffee only) and r0.44 (for total caffeine) are very similar to those found between cigarettes per day and cups of coffee per day (r0.43; Swan et al. 1996) and between heavy smoking and heavy coffee use (r0.43; Swan et al. 1997). An important strength of our study is that it involves a large sample of >10 000 twins, making it twice as large as the biggest previous twin study (Swan et al. 1997). Even though the Netherlands is seen as a typical ‘coffee drinking’ country (Fredholm 2011; Ferdman 2014), the genetic underpinnings of the overlap between smoking and caffeine were similar compared with US populations.

Genetic correlations as estimated by twin modelling were remarkably similar to the genetic correlation between cigarettes per day and cups of coffee per day based on effect size estimates of two large GW A meta-analyses. Some caution is warranted with the comparison of these two methods, however, given the difference in measures of smoking and caffeine use. A major advantage of LD-score regression is that genetic correlation is based on data of multiple (European) populations, instead of just one (Dutch) population in the case of twin modelling. There was also evidence for an overlap between SNPs associated with cups of coffee per day and smoking initiation and smoking persistence. Genetic variants that increase coffee consumption thus also increase the chance of becoming a smoker and decrease the chance of quitting smoking once started. The latter is in agreement with observational studies finding an inverse relation between quitting smoking and coffee consumption (Sorlie & Kannel 1990; Olsen 1993; Fernandez et al. 1997). The present study is the first to correlate SNPs associated with smoking behaviour with SNPs associated with coffee consumption.

In Dutch and British individuals, there was no evidence for causal effects of smoking on caffeine use, or vice versa, as also suggested from the bivariate twin data, in which the unique environmental correlation was low or zero (De Moor et al. 2008). MR analyses may have been underpowered to pick up on causal effects. Given the study’s sample size, an increase of 29 mg of caffeine with each cigarette per day would have provided evidence for a causal effect of smoking on caffeine (power
of 0.80), as opposed to the increase of 5 mg that we found. Whereas each extra mg of caffeine was associated with +0.01 cigarettes per day, an increase of 0.04 would have pointed to a causal association of caffeine on smoking (Brion, Shakhbazov, & Visscher 2013). Although the eight SNPs included in the caffeine use SNP score were taken from a study looking exclusively at coffee use (Cornells et al. 2014), the two most significant SNPs (rs6968554 and rs2472297) were recently also positively associated with total caffeine (coffee + tea + cola), coffee and tea, but not with cola alone (McMahon et al. 2014). A potential limitation of MR is pleiotropy. Under pleiotropy, one genetic variant or set of variants is associated with multiple phenotypes. It could, e.g. be the case that the caffeine use SNP score directly affects smoking (not acting through a causal effect of caffeine use on smoking), which would undermine the principle of MR. Pleiotropic effects can be minimized by selecting genetic instruments with effects that plausibly act directly on the trait in question. Also, when different (sets of) SNPs separately have the same association with the outcome of interest, it is less likely that this is due to pleiotropy (Davey Smith & Hemani 2014). In the present study, MR analyses were therefore repeated with a genetic risk score for caffeine including only rs6968554 and rs2472297. Both of these SNPs play a clear role in the metabolism of caffeine, and it is unlikely that they have a direct effect (e.g. not acting through caffeine use) on smoking. The results of these analyses were very similar to the risk score based on eight SNPs (data not shown), suggesting that pleiotropy did not affect the results.

This study was the first to combine multiple methods with the aim of unravelling the nature of the co-morbidity between smoking and caffeine use. Our findings point to shared genetic factors underlying the association between smoking and caffeine use, not ruling out that there is an additional (smaller) influence of causal effects. At least, some of the genetic risk factors for smoking overlap with genetic risk factors for caffeine use. This finding suggests that initiating smoking may be especially undesirable for heavy caffeine users, given their genetic susceptibility to smoke more heavily or to more easily become nicotine dependent. Because smoking is likely to be initiated before heavy caffeine use is manifested, a more important implication may be that smokers who are also heavy caffeine users might, on average, find it more difficult to quit because of their genetic background. To confirm our findings and to further clarify the complex association between these (addictive) behaviours, further research is required. Especially, causal effects from smoking on caffeine or vice versa need to be explored through MR analysis in larger samples.

Acknowledgements

We thank the members of the twin families registered with the NTR for participating in this study. The work described here was supported by the European Research Council (ERC) [grant number 284167: ‘Beyond the Genetics of Addiction’ (principal investigator J. M. V.) and grant number 230374: ‘Genetics of Mental Illness’ (principal investigator D. I. B.)] and grants from the Netherlands Organization for Scientific Research (NWO): ZonMW Addiction (31160008 and NWO 016-115-035), Genetic and Family Influences on Adolescent Psychopathology and Wellness (NWO 463-06-001), A twin-sib study of adolescent wellness (NWO-VENI 451-04-034), VU University’s Institute for Health and Care Research (EMGO+) and Neuroscience Campus Amsterdam (NCA). Part of the genotyping was funded by Biobanking and Biomolecular Resources Research Infrastructure (BBMRI–NL, 184.021.007), NWO/SPI 56-464-14192, the Genetic Association Information Network (GAIN) of the Foundation for the National Institutes of Health, Rutgers University Cell and DNA Repository (NIMH U24 MH068457-06), the Avera Institute, Sioux Falls, South Dakota (USA), and the National Institutes of Health (NIH R01 HD042157-01A1, MH081802, Grand Opportunity grants 1RC2 MH089951 and 1RC2 MH089995). A. E. T., M. R. M. and J. J. W. are members of the UK Centre for Tobacco and Alcohol Studies, a UK Clinical Research Council Public Health Research Centre of Excellence. Funding from British Heart Foundation, Cancer Research UK, Economic and Social Research Council, Medical Research Council, and the National Institute for Health Research, under the auspices of the UK Clinical Research Collaboration, is gratefully acknowledged. Support from the Medical Research Council (MC_UU_12013/6) is also gratefully acknowledged. We are extremely grateful to all the ALSPAC families who took part in this study, the midwives for their help in recruiting them and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. The UK Medical Research Council and the Wellcome Trust (Grant ref: 102215/2/13/2) and the University of Bristol provide core support for ALSPAC. Finally, we would like to acknowledge the work of the Tobacco, Alcohol and Genetics (TAG) consortium and the Coffee and Caffeine Genetic Consortium. This publication is the work of the authors, and J. L. T., A. E. T., J. J. W., M. G. N., M. C. N., G. M., J. J. H., B. M. L. B., D. I. B., M. R. M. and J. M. V. will serve as guarantors for the contents of this paper.

Authors Contribution

J. L. T. and J. M. V. were responsible for NTR data collection. D. I. B. supervised the data collection. J. L. T. carried
out the analyses and drafted the manuscript. J. M. V., D. I. M., M. R. M. and A. E. T. assisted with writing and interpretation of findings. M. G. N. and M. C. N. assisted with carrying out the analyses. J. J. H. and B. M. L. B. contributed to the data cleaning and checks of the NTR sample. J. J. W. and G. M. contributed to the (genotype) data cleaning and checks of the ALSPAC sample. All of the authors reviewed the content of the manuscript and approved the final version.

References


meta-analysis of self-reported allergy identifies shared and allergy-specific susceptibility loci. Nat Genet 45, 907–911. doi: 10.1038/ng.2686 http://www.nature.com/ng/journal/v45/n8/abs/ng.2686.html#supplementary-information


SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher’s web-site:

**Table S1.** Prevalences of current smoking, high coffee use and high total caffeine use as estimated in structural equation models

**Table S2.** Twin correlations from structural equation models before constraining correlations across gender

**Table S3.** Structural equation models to explore additive genetic (A), common environmental (C) and unique environmental (E) influences on current smoking and caffeine use in mg per day, and on their overlap

**Table S4.** Twin correlations for current smoking and coffee use in mg per day (N = 10,368) and for current smoking and total caffeine use in mg per day (N = 8,060) from the best-fitting saturated models

**Table S5.** Estimates of additive genetic (A), common environmental (C) and unique environmental (E) influences on current smoking and coffee use in mg per day (N = 10,368) and on current smoking and total caffeine use in mg per day (N = 8,060), from the full and the best-fitting bivariate genetic models

**Table S6.** Associations between the caffeine use SNP score and the smoking SNP and confounding variables in the Netherlands Twin Register (NTR) and the Avon Longitudinal Study of Parents and Children (ALSPAC)

**Table S7.** Mendelian Randomization analyses between the smoking SNP and smoking behaviour and daily caffeine consumption (in mg) in the Avon Longitudinal Study of Parents and Children (ALSPAC)

**Figure S1.** Path estimates for bivariate genetic models on current smoking & coffee and on current smoking & total caffeine. A = additive genetic factors, C = common genetic factors, E = unique environmental factors. Both the initial models and the best-fitting models are shown.

**Figure S2.** Path estimates for bivariate genetic models on current smoking & coffee and on current smoking & total caffeine. A = additive genetic factors, C = common genetic factors, E = unique environmental factors. Here, genetic (between A1 and A2), common environmental (between C1 and C2) and unique environmental (E1 and E2) correlations are shown. Calculation of the genetic correlation was based on the following formula: $a_{\text{CurrentSmoking}} \cdot a_{\text{CurrentSmoking-Coffee}} / (a^2_{\text{CurrentSmoking}} \cdot a^2_{\text{CurrentSmoking-Coffee}} + a^2_{\text{Coffee}})$, where $a_{\text{CurrentSmoking}}$ and $a_{\text{Coffee}}$ represent the path loadings going from A1 to ‘Current smoking’ and from A2 to ‘Coffee’, respectively and $a_{\text{CurrentSmoking-Coffee}}$ represents the path loading going from A2 to ‘Current smoking’ in Figure S1. Environmental correlations were calculated in the same way. Both the initial models and the best-fitting models are shown.

**Figure S3.** Instrumental value of the genetic risk scores. The forest plots show associations between the caffeine use SNP score and total caffeine use and caffeine from coffee in mg per day (A) and between the smoking SNP and cigarettes smoked per day (B). NTR = Netherlands Twin Register; ALSPAC = Avon Longitudinal Study of Parents and Children.

**Figure S4.** MR analyses testing causal effects. The forest plots show associations between the smoking SNP and total caffeine use and caffeine from coffee in mg per day (A) and between the caffeine use SNP score and smoking behaviour (cigarettes smoked per day, smoking persistence, smoking initiation and current smoking) (B). NTR = Netherlands Twin Register; ALSPAC = Avon Longitudinal Study of Parents and Children.