



De Vocht, F., Suderman, M., Tilling, K., Heron, J., Howe, L., Campbell, R., Hickman, M., & Relton, C. (2018). DNA methylation from birth to late adolescence and development of multiple-risk behaviours. *Journal of Affective Disorders*, 227, 588-594.  
<https://doi.org/10.1016/j.jad.2017.11.055>

Publisher's PDF, also known as Version of record

License (if available):  
CC BY

Link to published version (if available):  
[10.1016/j.jad.2017.11.055](https://doi.org/10.1016/j.jad.2017.11.055)

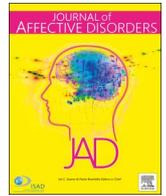
[Link to publication record in Explore Bristol Research](#)  
PDF-document

This is the final published version of the article (version of record). It first appeared online via Elsevier at <http://www.sciencedirect.com/science/article/pii/S0165032717313666> . Please refer to any applicable terms of use of the publisher.

## University of Bristol - Explore Bristol Research

### General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:  
<http://www.bristol.ac.uk/pure/user-guides/explore-bristol-research/ebr-terms/>



## Research paper

## DNA methylation from birth to late adolescence and development of multiple-risk behaviours



F. de Vocht<sup>a,\*</sup>, M. Suderman<sup>a,b,1</sup>, K. Tilling<sup>a,b</sup>, J. Heron<sup>a</sup>, L.D. Howe<sup>a,b</sup>, R. Campbell<sup>a</sup>, M. Hickman<sup>a</sup>, C. Relton<sup>a,b</sup>

<sup>a</sup> School of Social and Community Medicine, University of Bristol, Bristol, UK

<sup>b</sup> MRC Integrative Epidemiology Unit, UK

## ARTICLE INFO

## Keywords:

ALSPAC  
DNA methylation  
Epigenetics  
Risk behaviour  
Multiple risk behaviour  
Adolescents

## ABSTRACT

**Background:** Risk behaviours in adolescence are linked to poor educational attainment and health and other outcomes in young adulthood. We explored whether there are molecular mechanisms associated with the development, or the result, of multiple risk behaviours (MRBs).

**Methods:** MRBs (antisocial behaviour and delinquency, traffic-related risk behaviour, risky sexual behaviour, lack of exercise) and their sumscore were characterized based on self-reported questions at age 7 and 17 within the ARIES subsample of the ALSPAC birth cohort, and were linked to DNA methylation at over 485,000 CpG sites at ages 0,7 and 17. Associations were determined for participants with complete data (n = 227–575).

**Results:** There was weak evidence of associations between cumulative MRBs and methylation at cg01783492 and cg16720578 at age 17. DNA methylation at age 17 was associated with risky sexual behaviour (cg22883332), lack of exercise (cg03152353, cg20056908, cg20571116) and substance use (cg02188400, cg13906377). No associations between DNA methylation and individual risk behaviours at age 7 were observed. DNA methylation at age 7 might predispose for traffic-related risk behaviour (cg24683561) and substance use (cg08761410) at age 17.

**Limitations:** Main limitations are absence of information on directly measured blood cell type proportions and tissue specificity, and a modest sample size.

**Conclusions:** Cumulative MRB in late adolescence was associated with effects on DNA methylation. More specifically, risky sexual behaviour and sedentary behaviour are associated with changes in DNA methylation, while DNA methylation in childhood may predict later traffic-related risky behaviour. For substance use effects in both temporal directions were observed.

## 1. Introduction

Risk behaviours in adolescence, such as substance use, alcohol consumption, poor diet, physical inactivity, unprotected sex and antisocial behaviour are common (Kipping et al., 2015), and associated with adverse health and other outcomes in later adolescence (Hale and Viner, 2016). Analysis of a UK birth cohort (the Avon Longitudinal Study of Parents and Children cohort study) showed that multiple risk behaviours (MRBs) in adolescence are common, with 74% of 15–16 year olds being physically inactive, 34% consuming alcohol at hazardous levels, and 42% being engaged in antisocial and criminal behaviour (MacArthur et al., 2012). Initiation of most risk behaviours generally occurs at ages 14–16 (Hale and Viner, 2016), while it has further been shown that those who initiate risk behaviours early were

more likely to be multiple risk-takers (DuRant et al., 1999).

These risky behaviours are linked to poor educational attainment and a range of morbidity and mortality outcomes in later life (Galambos and Tilton-Weaver, 1998; Rohde et al., 2001; Sandfort et al., 2008; Viner and Taylor, 2007), which has made the prevention of health risk behaviours by adolescents a focus for policy in the UK and internationally (Hale and Viner, 2012). It has been shown that risk behaviours, such as for example substance use, sexual risk and delinquency in adolescence (Hair et al., 2009; Hale and Viner, 2016; Jackson et al., 2012; Meader et al., 2016; Wiefferink et al., 2006) often cluster. Two broad theories have been proposed that may explain the correlations between adolescent risk behaviours: [1] ‘Gateway theory’ (Pudney, 2003), which suggests that engagement in one form of risk leads to others through a decrease in perceiving dangers and/or through

\* Correspondence to: Population Health Sciences, University of Bristol, Canynge Hall, 39 Whatley Road, Bristol BS8 2PS, UK.

E-mail address: [frank.devocht@bristol.ac.uk](mailto:frank.devocht@bristol.ac.uk) (F. de Vocht).

<sup>1</sup> Joint first authors.

**Table 1**  
(Multiple) Risk Behaviour (MRB) categories at time point 2 (age ~ 17 years).

Sum scores	Description
MRB sum score (continuous)	A + B + C + D + E
MRB sum score (binary)	Low ( $\leq$ median) vs high ( $>$ median) MRB risk score
MRB sum score (25th vs 75th)	Low ( $\leq$ 25th percentile) vs high ( $\geq$ 75th percentile) MRB risk score (25th–75th percentile removed from analysis)
<b>Strata/individual risk behaviours</b>	
A) Antisocial behaviour and delinquency	1) At least once in the past year: theft from vehicle, carrying a weapon, being rude/rowdy in public place, breaking into a vehicle, taking a car that is not their own, stealing from a shop, buying stolen goods, stealing money from someone else, burgling a house, kicked/hurt someone on purpose, used a cheque book/card not their own, damaged property not their own on purpose. 2) Has hurt themselves on purpose at least once in the last year 3) One or more forms of gambling over past 12 months
B) Traffic-related risk behaviour	1) Has been car passenger at least once in their lifetime where the driver (a) has had alcohol or (b) does not have valid licence or (c) chose not to wear a seat belt last time travelled in a car, van or taxi. 2) Has driven a scooter off road or w/o licence (or both) at least one time in lifetime 3) Has driven on / off road w/o licence at least once in lifetime 4) Did not use a helmet last time they rode a bike (within the recent four week period)
C) Risky sexual behaviour	1) Has had sexual intercourse and not used a condom on last occasion they had sex in the last 12 months 2) Multiple sexual partners (3+) in past year
D) Lack of physical exercise	1) Has typically over the past year exercised $<$ 5 times/wk
E) Substance use	1) Has used cannabis 2–4 times a month or more frequently in the last 12 months 2) Typically smokes at least one cigarette per week over the last 12 months 3) Alcohol Use Disorders Identification Test (AUDIT) score of 8 or more 4) Illicit drug and inhalant use (excluding cannabis) over past twelve months.

increased exposure to other risk behaviours, and ‘*Problem Behaviour theory*’ which purports that sets of behaviours that are defined as problematic and/or unconventional are enacted as a manifestation of disregard of social conventions or, depending on age, of maturity (Jessor et al., 1998). The implications of either theory throughout adolescence are that whereas Gateway theory predicts accumulating associations between risk behaviours with age, Problem Behaviour theory may result in a stable risk behaviour in late adolescence. This pattern has, for example, been observed for trajectories of substance abuse from early adolescence to adulthood which described how correlations between risk behaviours decrease with age, and similar patterns have been observed for drug use and sexual behaviour, but not for alcohol use (Hale and Viner, 2016). This decrease can be ascribed to a transition from general risk behaviour aetiology in adolescence to risk-specific influences in adulthood (Vrieze et al., 2012). Like most complex phenotypes, in addition to shared social/environmental factors, there may also be also be shared biological factors that influence the development of these multiple risk behaviours, and there is evidence that the generalized risk may be directly influenced by modifiable molecular factors in childhood or adolescence (Vrieze et al., 2012). In support of this, epigenetic processes have been shown to be associated with substance use in adolescents (Cecil et al., 2016).

Epigenetics, the study of heritable changes in gene expression not due to changes in DNA sequence, offers the potential to identify molecular mechanisms by which environmental and lifestyle exposures may affect health (Florath et al., 2014; Hannum et al., 2013). Epigenetic mechanisms include DNA methylation, histone modifications and microRNA, all of which act in concert to regulate gene expression (Groom et al., 2011). DNA methylation, the addition of methyl groups to nucleotide bases, is the most stable and most readily quantifiable epigenetic marker and has thus become the most widely studied. DNA methylation is affected by genetic variation (Gaunt et al., 2016), is sensitive to pre- and postnatal exogenous influences (de Vocht et al., 2015), and has also been linked to physical and psychiatric disorders, including addiction (Cecil et al., 2015). Recent technological advances have allowed the application of genomic technologies to epigenetics, facilitating the large scale generation of quantitative DNA methylation data across the genome (Mensaert et al., 2014).

In this study, we explored, for the first time, associations between multiple risk behaviours and DNA methylation measured at multiple time points across childhood, at birth, age 7 and age 17. Critically, two of the DNA methylation measurements were taken prior to the

behaviour which enables investigation of the temporal direction of the associations to address the question; are epigenetic changes observed at different time points risk factors for behaviour or are they consequences of the behaviour itself?

## 2. Methods

This study used DNA methylation data generated under the auspices of the Avon Longitudinal Study of Parents and Children (ALSPAC) (Boyd et al., 2013; Fraser et al., 2013). ALSPAC recruited 14,541 pregnant women with expected delivery dates between April 1991 and December 1992. Of these initial pregnancies, there were 14,062 live births and 13,988 children who were alive at 1 year of age. The study website contains details of all the data that are available through a fully searchable data dictionary (<http://www.bris.ac.uk/alspac/researchers/data-access/data-dictionary>). As part of the ARIES (Relton et al., 2015) project (<http://www.ariesepigenomics.org.uk>), a sub-sample of 1,018 ALSPAC mother–child pairs had DNA methylation measured using the Infinium HumanMethylation450 BeadChip (Illumina, Inc.) (Dedeurwaerder et al., 2011). Here, we use DNA methylation data generated from cord blood and venous blood samples at age 7 and again at age 15 or 17 years, leading to three measurements of DNA methylation per child. All DNA methylation wet-lab and preprocessing analyses were performed at the University of Bristol as part of the ARIES project and has been described in detail previously (Relton et al., 2015).

Informed consent was obtained from all ALSPAC participants and ethical approval was obtained from the ALSPAC Law and Ethics Committee as well as Local Research Committees.

### 2.1. Outcome

Multiple risk behaviour (MRB) was characterized based on self-reported questions broadly divided into five domains: antisocial behaviour and delinquency (2 questions), traffic-related risk behaviour (4 questions), risky sexual behaviour (2 questions), sedentary behaviour (1 question), and substance use (4 questions). Where questions were not binary already, behaviours were recoded to present (1) or absent (0) based on cut-off points informed by the literature (described in Kipping et al. (2015)). Dichotomised scores for each behaviour were summed per domain to obtain a summary domain score for ages ~ 7 and ~ 17, and for age 17 these in turn were added to obtain the MRB sum score; i.e. the primary outcome of this study (range 0–13 in theory, 0–9 in this

population). An overview of the questions for time point 2 (age ~ 17 years), the domains, and the different MRB scores is provided in Table 1.

## 2.2. DNA methylation profile generation

DNA was bisulphite converted using the Zymo EZ DNA Methylation™ kit (Zymo, Irvine, CA). Infinium HumanMethylation450 BeadChips (Illumina, Inc.) were used to measure genome-wide DNA methylation levels at over 485,000 CpG sites. The arrays were scanned using an Illumina iScan, with initial quality review using GenomeStudio. This assay detects methylation of cytosine at CpG islands using two site-specific probes – one to detect the methylated (M) locus and one to detect the unmethylated (U) locus. Single-base extension of the probes incorporates a labelled chain-terminating ddNTP, which is then stained with a fluorescence reagent. The ratio of fluorescent signals from the methylated site versus the unmethylated site determines the level of methylation at the locus. The level of methylation is expressed as a “Beta” value ( $\beta$ -value), ranging from 0 (no cytosine methylation) to 1 (complete cytosine methylation).  $\beta$ -values are reported as percentages.

## 2.3. Quality control

During the data generation process a wide range of batch variables were recorded in a purpose-built laboratory information management system (LIMS). The LIMS also reported quality control metrics from the standard control probes on the 450 K BeadChip. Samples failing quality (samples with > 20% probes with  $p$ -value  $\geq 0.01$ ) were repeated. Samples from all three time points in ARIES were randomized across arrays to minimise the potential for batch effects. As an additional quality control step, genotype probes on the 450 K BeadChip were compared between samples from the same individual and against SNP-chip data to identify and remove any sample mismatches.

## 2.4. Methylation profile normalisation

Raw  $\beta$ -values were pre-processed using R (version 3.0.1) with background correction and sub-set quantile normalisation performed using the pipeline described by Touleimat and Tost (Touleimat and Tost, 2012) and implemented in the *watermelon* R package (Pidsley et al., 2013). Finally, to reduce influence of outliers in regression models, normalized  $\beta$ -values were 90%-Winsorized.

## 2.5. Cell type heterogeneity

Blood is composed of many cell types and composition ratios can vary over time within a given individual as well as between individuals. DNA methylation differs between blood cell types so it is necessary to adjust for cell type variance in methylation analyses to avoid confounding. Cell type proportions for peripheral blood per individual were estimated from DNA methylation profiles using the method described by Houseman et al. (2012) using an adult peripheral blood cell type reference (Reinius et al., 2012). Cell type proportions for cord blood per individual were estimated identically but using a cord blood cell type reference (Gerwin et al., 2016).

## 2.6. Statistical analyses

Associations were tested using linear models implemented by the *limma* R package (Smyth, 2004). For these analyses we only use participants with complete data. MRB sum scores were included as dependent variables as a continuous measure, as a binary score based on the median score (low ( $\leq 2$ ) vs. high ( $> 2$ )), and as a comparison of extremes (25th (0–1) vs. 75th (3–9) percentiles). All models were adjusted for confounders: cell counts estimates (see above), sex, parity,

birthweight, gestational age, prenatal tobacco exposure, early pregnancy alcohol exposure, maternal age, maternal BMI, maternal highest education, maternal smoking history, and paternal occupation. In addition, confounders not covered by the above were adjusted for in Surrogate Variable Analysis (discussed in detail in Leek and Storey (2007)). We conducted three cross-sectional analyses at birth, age 7 and age 17, and to explore causal directions we additionally assessed associations between MRBs at ages 0 and 7 with methylation at age 17 and conversely methylation at ages 0 and 7 with MRBs at age 17. None of the association  $p$ -values were below the 0.05 Bonferroni adjusted threshold for significance ( $p = 1.07 \times 10^{-7}$ ). We therefore report associations at false discovery rate (FDR) less than 10% calculated using the  $q$  method (Storey and Tibshirani, 2003). Where associations between MRB and methylation was positive, indicating that methylation at that specific CpG site was statistically significantly more present in those individuals with MRBs compared to those without, we defined these sites as “hypermethylated” and conversely when this association was inverse, we defined these sites as “hypomethylated”.

## 3. Results

The demographics of the samples used for the different analyses are shown in Table 2. Because only participants with complete data were used, the sample sizes differed from 227 for the MRB sumscore to 575 for the analyses of ‘substance use’. The samples include slightly more women than men, and are mainly ( $\geq 97\%$ ) white. Participants are more often from higher socio-economic groups than the general ALSPAC population, based on information on paternal occupation and maternal and paternal education levels. About 10% of women drank alcohol during pregnancy.

Results of the primary analyses, association between MRB sum score and DNA methylation at age 17, are shown in Table 3 and indicate weak evidence of association with hypomethylation of cg01783492, located near *TEKT5*, and hypermethylation of cg16720578, located near *BMP4*. No associations between MRB sum score and DNA methylation at age 7 with an FDR < 10% were observed.

Stratification of the MRB score into five domains (Table 4) suggests that DNA methylation may be involved in risky sexual behaviour, lack of exercise and substance use, but is not involved in antisocial behaviour and traffic-related risk behaviour. There is weak evidence of hypomethylation of CpG site cg22883332 located on the *CALML3* gene. Sedentary behaviour at age 17 is more strongly associated with hypermethylation of three CpG sites (cg03152353, cg20056908, cg20571116) located on the *NOTCH1*, *VAMP1* and *PLCD4* genes, respectively. Substance use is associated with hypomethylation of cg02188400, located upstream of *RRAS2*, and hypermethylation of cg13906377, located upstream of microRNA 6068, respectively. Comparable analyses of DNA methylation at age 7 with accumulated risk scores for the individual domains at the same age did not result in any association.

Table 5 shows associations between DNA methylation at birth and age 7 and MRB scores, stratified by domain, at age 17 to explore whether methylation may be a risk factor for later behaviour instead of the result of behaviour. There is no evidence of association with antisocial behaviour, delinquency, risky sexual behaviour and sedentary behaviour. For traffic-related risk behaviour at age 17 in contrast, hypermethylation of cg24683561 on the *PYGO2* gene at birth and age 7 was observed, suggesting that DNA methylation predates this type of risk behaviour. Similarly, for substance use also indicate a mechanism through which cg08761410, in the first intron of *PISD*, hypermethylation may predispose for substance abuse risk.

## 4. Discussion

These exploratory analyses indicate some evidence that cumulative MRB behaviour in adolescence is associated with differences in DNA

**Table 2**  
Demographics of ARIES participants by MRB category.

		Sum score	Personal risk	Traffic	Sex	Exercise	Substance
<b>N</b>		<b>227</b>	<b>441</b>	<b>319</b>	<b>524</b>	<b>364</b>	<b>575</b>
Sex (%)	Female	57.3	50.8	58.6	51.3	59.9	52.2
	Male	42.7	49.2	41.4	48.7	40.1	47.8
Parity (%)	0	54.6	50.6	49.5	49.0	48.4	48.0
	1	31.7	35.1	37.0	36.3	37.4	37.4
	2	10.6	10.7	10.3	11.3	10.7	11.1
	3	2.6	2.9	2.8	2.9	3.0	3.0
	4	0.4	0.7	0.3	0.6	0.5	0.5
Maternal ethnicity (%)	Non-white	2.6	2.9	3.1	2.5	3.0	2.8
	White	97.4	97.1	96.9	97.5	97.0	97.2
Paternal occupation (%)	Professional	23.8	20.0	22.3	19.3	21.4	19.1
	Managerial and technical	39.6	34.9	37.6	34.9	37.1	34.3
	Skilled manual	17.2	23.4	20.1	24.2	20.3	24.7
	Skilled non-manual	13.7	13.6	13.8	12.8	13.5	12.9
	Partly-skilled	4.4	5.7	5.3	6.3	6.0	6.8
	Unskilled occupation	1.3	2.5	0.9	2.5	1.6	2.3
Maternal education (%)	A level	33.5	29.0	32.9	29.2	31.6	29.2
	CSE	5.7	5.4	5.6	5.9	5.2	5.7
	Degree	30.4	24.7	29.5	24.2	27.7	24.5
	O level	27.3	34.5	27.9	33.8	29.9	33.6
	vocational	3.1	6.3	4.1	6.9	5.5	7.0
Paternal education (%)	A level	29.5	28.6	29.2	29.0	30.2	28.7
	CSE	8.4	11.8	9.7	11.8	10.7	12.2
	Degree	40.1	33.3	37.0	30.5	35.4	30.4
	O level	18.1	20.2	19.7	21.4	18.4	21.2
	vocational	4.0	6.1	4.4	7.3	5.2	7.5
Maternal alcohol Consumption (%) (during pregnancy)	Yes	8.8	11.3	10.0	11.5	11.3	11.7
	No	89.0	84.8	87.1	84.5	85.4	84.7

methylation. In relation to specific behaviours, which we analysed to investigate whether effects in individual risk behaviours could have been concealed because our main hypothesis concerned cumulative MRB, and the direction of effects, our evidence suggests that risky sexual behaviour and sedentary behaviour may induce DNA methylation changes whereas DNA methylation in childhood may predispose for later traffic-related risky behaviour. For substance use associations in both temporal directions were observed. Risky sexual behaviour may result in hypomethylation in the *CALML3* gene, which has been linked to certain behavioural/neurological phenotypes through its involvement in *MYO10* translation. Sedentary behaviour is associated with hypermethylation of CpG sites in the *NOTCH1*, *VAMP8* and *PLCD4* genes, all linked to developmental processes. *PYGO2*, which in these analyses is suggestive of a predisposition for subsequent risky behaviour in traffic, can influence the Wnt signalling pathway and therefore embryonic development.

There is very little data on associations between DNA methylation and risk behaviours, and as such the associations between DNA methylation at age 17 and MRB in mid-adolescence observed in our study cannot yet be compared to other studies. A previous study, conducted in the same population, assessed DNA methylation in relation to substance

abuse, and concluded that birth was a potentially sensitive window of biological vulnerability (Cecil et al., 2016). Although we similarly found some evidence that DNA methylation may predispose for substance use, we observed hypermethylation of cg08761410 at birth in relation to substance use at age 17, which was a different CpG site to those observed in Cecil et al. (2016). In addition, we also observed associations between MRB at age 17 and DNA methylation at age 17, but not age 7 or at birth, indicating effects in adolescence and the possibility that behaviour may affect DNA methylation as well. Other possible explanations for the observed differences are that different subsets of the ARIES cohort were used in both studies, depending on the completeness of information for the analyses. Additionally, whereas the previous study used confirmatory factor analysis to extract three first-order factors of tobacco, cannabis and alcohol use, as well as a second-order factor for substance use, in our study instead we added up dichotomized (yes/no) risk behaviours because our main focus was on the total MRB score.

The causes underlying the development of MRB remain uncertain, but social processes, personality traits, and other individual factors, including biological processes, may be involved. It has been suggested that MRBs in adolescents may be the result of deficits in impulse-

**Table 3**  
Associations between Multiple Risk Behaviour summary score at age 17 and DNA methylation at age 17.<sup>a</sup>

CpG site	Chromosome	Gene	Genomic location	Position	Beta <sup>b</sup>	P-value	FDR <sup>c</sup>
<i>Cumulative risk score</i>							
cg01783492	16	Downstream of <i>TEKT5</i>	–	10721112	–0.0181	3.54E–07	0.08
cg16720578	14	Downstream of <i>BMP4</i>	–	54410717	0.0146	1.11E–07	0.05
<i>Binary score (low/high)</i>							
None identified							
<i>Extreme scores (25% vs 75%)</i>							
None identified							

<sup>a</sup> Only CpG sites with FDR ≤ 10% reported.

<sup>b</sup> Adjusted for sex, parity, birthweight, gestational age, prenatal tobacco exposure, early pregnancy alcohol exposure, maternal age, maternal BMI, maternal highest education, maternal smoking history, paternal occupation, and surrogate variable.

<sup>c</sup> False Discovery Rate (FDR).

**Table 4**  
Associations between Individual Risk Behaviours and DNA methylation at different ages<sup>a</sup>.

Risk behaviour	Age	CpG site	Chromosome	Gene	Genomic location	Position	Beta <sup>b</sup>	P-value	FDR
Antisocial behaviour and delinquency	17	None							
	7	None							
	0	NA							
Traffic-related risk behaviour	17	None							
	7	None							
	0	NA							
Risky sexual behaviour	17	cg22883332	10	CALML3	TSS1500	5566260	-0.0335	2.18E-07	0.10
	7	NA							
	0	NA							
Lack of exercise	17	cg03152353	9	NOTCH1	Body	139417194	0.0245	2.80E-08	0.01
		cg20056908	2	VAMP8	3'UTR	85808945	0.0271	5.43E-08	0.01
		cg20571116	2	PLCD4	Body	219485946	0.0147	1.88E-07	0.03
	7	None							
Substance use	0	NA							
	17	cg02188400	11	Upstream of RRAS2	-	14406326	-0.0041	5.96E-09	0.00
		cg13906377	1	Upstream of microRNA 6068	-	63792695	0.0014	4.25E-07	0.10
	7	None							
	0	NA							

<sup>a</sup> Only CpG sites with FDR ≤ 10% reported

<sup>b</sup> Adjusted for sex, parity, birthweight, gestational age, prenatal tobacco exposure, early pregnancy alcohol exposure, maternal age, maternal BMI, maternal highest education, maternal smoking history, paternal occupation, and surrogate variable

control, which has a neurological basis (Casey et al., 2008), and in agreement with that here we do observe some evidence of DNA methylation predisposing for traffic-related risk behaviour and substance use.

Similarly, DNA methylation has been associated with other lifestyle and environmental exposures as well, and although we adjusted our statistical models for what we believe to be the most important confounding factors, we cannot exclude residual confounding from other, non-included factors.

This study has several limitations. This study had reduced statistical power due to the relatively limited sample size of the study, which was further diminished as a result of missing values in the variables used to construct the risk scores as well as in the covariates, while also the study, uniquely, made use of (self-reported) behaviours rather than metrics with less measurement error. However, dichotomisation of risk behaviours was based on the literature. Although there are approximately 900 DNA methylation profiles at each of the three time points, for some association tests, the number with complete information is as low as 227. Moreover, missingness of information on risk behaviour is socially patterned and not “completely at random” (Relton et al., 2015). Missingness certainly reduces power, however, in other analyses of risk behaviour that have imputed for missing observations the direction of associations have not changed and often strengthened (for example: Gage et al., 2015; Mahedy et al., 2017). Nonetheless, even with the diminished sample size of this study we were still able to detect

associations. Although multiple imputation could have been applied to increase sample size, current approaches are not feasible for genomic datasets including hundreds of thousands of measured variables, in our case variables corresponding to DNA methylation levels at over 480,000 CpG sites. In future, when feasible approaches have been developed, we plan to revisit these analyses.

A further limitation of the current analysis is the lack of directly measured blood cell type proportions. We therefore adjusted for cell count heterogeneity using estimates obtained using the method of Houseman et al. (2012). The method relies on having a reference dataset consisting of DNA methylation profiles of purified cell types. For peripheral blood a reference based on adult peripheral blood is available (Reinius et al., 2012), but not for childhood peripheral blood. However, for cord blood analysis, we were able to use a reference derived from cord blood (Gervin et al., 2016). Additional DNA methylation variation due to cell count heterogeneity would most likely have been represented in the surrogate variables also included as model covariates (McGregor et al., 2016).

Tissue specificity may also limit inferences from this study because levels of methylation vary between tissue types and may relate differently to traits and exposures. In the current study we have blood sample methylation but it may have been more informative had we been able to test associations in a more relevant cell type such as neurons in the brain.

DNA methylation is only one epigenetic mechanism among a

**Table 5**  
Assessing reverse causality: associations between DNA methylation at birth and age 7 and Risk behaviour at age 17.<sup>a</sup>

Risk behaviour (Age 17)	Methylation Age	CpG site	Chromosome	Gene	Genomic location	Position	Beta <sup>b</sup>	P-value	FDR
Antisocial behaviour and delinquency	7	None							
	0	None							
Traffic-related risk behaviour	7	cg24683561	1	PYGO2	TSS1500	154934703	0.0038	1.75E-09	0.00
	0	cg24683561	1	PYGO2	TSS1500	154934703	0.0037	9.49E-09	0.00
Risky sexual behaviour	7	None							
	0	None							
Lack of exercise	7	None							
	0	None							
Substance use	7	cg08761410	22	first intron PISD	-	32058119	0.0007	6.64E-08	0.03
	0	cg08761410	22	first intron PISD	-	32058119	0.0007	1.21E-07	0.06

<sup>a</sup> Only CpG sites with FDR ≤ 10% reported.

<sup>b</sup> Adjusted for sex, parity, birthweight, gestational age, prenatal tobacco exposure, early pregnancy alcohol exposure, maternal age, maternal BMI, maternal highest education, maternal smoking history, paternal occupation, and surrogate variable.

number of other interacting mechanisms including histone modifications and non-coding RNAs. It is possible that one or more of these other mechanisms is more strongly associated with risk behaviour. However, DNA methylation is a reasonable mechanism to consider in this initial study because it is the most stable and most readily available quantifiable epigenetic mark and has thus become the most widely studied (Bintu et al., 2016).

And finally, MRB was classified by aggregating self-reported individual risk behaviours identified by yes/no only, which will have impacted on measurement error. Previous work on substance use for example (Cecil et al., 2016), used confirmatory factor analysis to identify first and second order factors to define exposure, but this was for one risk behaviour only.

The main strength of this study is the unique resource which allowed for the assessment of genome-wide methylation profiles at different time points from birth to late adolescence linked to a detailed phenotypic characterization, including information on a distinct set of risk behaviours representing the same time period, which enables assessment of the temporality of the associations. In these analyses we used three cross-sectional models to compare methylation patterns at birth, age 7 and in adolescence, but with better characterization of the dynamic elements of the human methylome (Ziller et al., 2013), longitudinal analyses will help to better elucidate persistent and reversible effects of (environmental) exposures as well as critical periods of effect (Mishra et al., 2009).

A further strength of this study is that, given the evidence that behaviours are known to cluster in individuals (MacArthur et al., 2012), in this study we could assess the cumulative effect of antisocial behaviour and delinquency, traffic-related risk behaviour, risky sexual behaviour, sedentary behaviour and substance use (MRB) together, in addition to the individual behaviours, on DNA methylation.

To our knowledge this is the first study to assess associations between DNA methylation and multiple risk behaviours in adolescents. This exploratory study provides the first evidence that MRB may result in changes to DNA methylation, but also provides evidence that DNA methylation may predispose for certain risk behaviours. As such, the results of this study indicate this topic requires further study, including a larger sample size – for example by combining different cohorts with comparable information, improved statistical methodology to take into account the longitudinal nature of the data as well as to address issues of missing data in a satisfactory manner, and which in turn will also allow for improved characterization of the temporal patterns of MRB development.

## Acknowledgments

We are extremely grateful to all the 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. Medical Research Council Wellcome 102215/2/13/2) University of Bristol This publication is the work of the authors and Frank de Vocht will serve as guarantors for the contents of this paper.

## Funding

This work was supported by the MRC Integrative Epidemiology Unit and the University of Bristol (MC\_UU\_12013\_2 and MC\_UU\_12013\_8). The Accessible Resource for Integrated Epigenomics Studies (ARIES) was funded by the UK Biotechnology and Biological Sciences Research Council (BB/I025751/1 and BB/I025263/1). This work was also supported by the BBSRC and ESRC (grant number ES/N000498/1), the UK National Institute for Health Research (NIHR) Health Protection Research Unit (HPRU) in Evaluation, NIHR Biomedical Research Centre (BRC) and NIHR School for Public Health Research (SPHR). The UK

Medical Research Council and Wellcome Trust (Grant Ref. 102215/13/2) and the University of Bristol provide core support for ALSPAC. BBSRC(BB1025751/1BB/I025263/1).102215/2/13/1)(MC\_UU\_12013/2MC\_UU\_12013/8).MRC Career Development Award(MR/M020894/1).

## References

- Bintu, L., Yong, J., Antebi, Y.E., McCue, K., Kazuki, Y., Uno, N., Oshimura, M., Elowitz, M.B., 2016. Dynamics of epigenetic regulation at the single-cell level. *Science* 351 (6274), 720–724.
- Boyd, A., Golding, J., Macleod, J., Lawlor, D.A., Fraser, A., Henderson, J., Molloy, L., Ness, A., Ring, S., Davey Smith, G., 2013. Cohort Profile: the 'children of the 90s' – the index offspring of the Avon Longitudinal Study of Parents and Children. *Int. J. Epidemiol.* 42, 111–127.
- Casey, B.J., Getz, S., Galvan, A., 2008. The adolescent brain. *Dev. Rev.* 28, 62–77.
- Cecil, C., Walton, E., Viding, E., 2015. DNA methylation, substance use and addiction: a systematic review of recent animal and human research from a developmental perspective. *Curr. Addict. Rep.* 2, 331–346.
- Cecil, C.A., Walton, E., Smith, R.G., Viding, E., McCrory, E.J., Relton, C.L., Suderman, M., Pingault, J.B., McArdle, W., Gaunt, T.R., Mill, J., Barker, E.D., 2016. DNA methylation and substance-use risk: a prospective, genome-wide study spanning gestation to adolescence. *Transl. Psychiatry* 6, e976.
- Dedeurwaerder, S., Defrance, M., Calonne, E., Denis, H., Sotiriou, C., Fuks, F., 2011. Evaluation of the Infinium methylation 450K technology. *Epigenomics* 3, 771–784.
- DuRant, R.H., Smith, J.A., Kreiter, S.R., Krowchuk, D.P., 1999. The relationship between early age of onset of initial substance use and engaging in multiple health risk behaviors among young adolescents. *Arch. Pediatr. Adolesc. Med.* 153, 286–291.
- Florath, I., Butterbach, K., Muller, H., Bewerunge-Hudler, M., Brenner, H., 2014. Cross-sectional and longitudinal changes in DNA methylation with age: an epigenome-wide analysis revealing over 60 novel age-associated CpG sites. *Hum. Mol. Genet.* 23, 1186–1201.
- Fraser, A., Macdonald-Wallis, C., Tilling, K., Boyd, A., Golding, J., Davey Smith, G., Henderson, J., Macleod, J., Molloy, L., Ness, A., Ring, S., Nelson, S.M., Lawlor, D.A., 2013. Cohort profile: the avon longitudinal study of parents and children: ALSPAC mothers cohort. *Int. J. Epidemiol.* 42, 97–110.
- Gage, S.H., Hickman, M., Heron, J., Munafò, M.R., Lewis, G., Macleod, J., Zammit, S., 2015. Associations of cannabis and cigarette use with depression and anxiety at age 18: findings from the avon longitudinal study of parents and children. *PLoS One* 10, e0122896.
- Galambos, N.L., Tilton-Weaver, L.C., 1998. Multiple-risk behaviour in adolescents and young adults. *Health Rep.* 10, 9–20 (Eng; 29-21 Fre).
- Gaunt, T.R., Shihab, H.A., Hemani, G., Min, J.L., Woodward, G., Lyttleton, O., Zheng, J., Duggirala, A., McArdle, W.L., Ho, K., Ring, S.M., Evans, D.M., Davey Smith, G., Relton, C.L., 2016. Systematic identification of genetic influences on methylation across the human life course. *Genome Biol.* 17, 61.
- Gervin, K., Page, C.M., Aas, H.C., Jansen, M.A., Fjeldstad, H.E., Andreassen, B.K., Duijts, L., van Meurs, J.B., van Zelm, M.C., Jaddoe, V.W., Nordeng, H., Knudsen, G.P., Magnus, P., Nystad, W., Staff, A.C., Felix, J.F., Lyle, R., 2016. Cell type specific DNA methylation in cord blood: a 450K-reference data set and cell count-based validation of estimated cell type composition. *Epigenetics* 11, 690–698.
- Groom, A., Elliott, H.R., Embleton, N.D., Relton, C.L., 2011. Epigenetics and child health: basic principles. *Arch. Dis. Child.* 96, 863–869.
- Hair, E.C., Park, M.J., Ling, T.J., Moore, K.A., 2009. Risky behaviors in late adolescence: co-occurrence, predictors, and consequences. *J. Adolesc. Health* 45, 253–261.
- Hale, D.R., Viner, R.M., 2012. Policy responses to multiple risk behaviours in adolescents. *J. Public Health* 34 (Suppl 1), i11–i19.
- Hale, D.R., Viner, R.M., 2016. The correlates and course of multiple health risk behaviour in adolescence. *BMC Public Health* 16, 458.
- Hannum, G., Guinney, J., Zhao, L., Zhang, L., Hughes, G., Sadda, S., Klotzle, B., Bibikova, M., Fan, J.B., Gao, Y., Deconde, R., Chen, M., Rajapakse, I., Friend, S., Ideker, T., Zhang, K., 2013. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol. Cell* 49, 359–367.
- Houseman, E.A., Accomando, W.P., Koestler, D.C., Christensen, B.C., Marsit, C.J., Nelson, H.H., Wiencke, J.K., Kelsey, K.T., 2012. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinform.* 13, 1.
- Jackson, C., Sweeting, H., Haw, S., 2012. Clustering of substance use and sexual risk behaviour in adolescence: analysis of two cohort studies. *BMJ Open* 2, e000661.
- Jessor, R., Turbin, M.S., Costa, F.M., 1998. Protective factors in adolescent health behavior. *J. Personal. Soc. Psychol.* 75, 788–800.
- Kipping, R.R., Smith, M., Heron, J., Hickman, M., Campbell, R., 2015. Multiple risk behaviour in adolescence and socio-economic status: findings from a UK birth cohort. *Eur. J. Public Health* 25, 44–49.
- Leek, J.T., Storey, J.D., 2007. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet.* 3, 1724–1735.
- MacArthur, G.J., Smith, M.C., Melotti, R., Heron, J., Macleod, J., Hickman, M., Kipping, R.R., Campbell, R., Lewis, G., 2012. Patterns of alcohol use and multiple risk behaviour by gender during early and late adolescence: the ALSPAC cohort. *J. Public Health* 34 (Suppl 1), i20–i30.
- Mahedy, L., Hammerton, G., Teyhan, A., Edwards, A.C., Kendler, K.S., Moore, S.C., Hickman, M., Macleod, J., Heron, J., 2017. Parental alcohol use and risk of behavioral and emotional problems in offspring. *PLoS One* 12, e0178862.
- McGregor, K., Bernatsky, S., Colmegna, I., Hudson, M., Pastinen, T., Labbe, A.,

- Greenwood, C.M., 2016. An evaluation of methods correcting for cell-type heterogeneity in DNA methylation studies. *Genome Biol.* 17, 84.
- Meador, N., King, K., Moe-Byrne, T., Wright, K., Graham, H., Petticrew, M., Power, C., White, M., Sowden, A.J., 2016. A systematic review on the clustering and co-occurrence of multiple risk behaviours. *BMC Public Health* 16, 657.
- Mensaert, K., Denil, S., Trooskens, G., Van Criekinge, W., Thas, O., De Meyer, T., 2014. Next-generation technologies and data analytical approaches for epigenomics. *Environ. Mol. Mutagen.* 55, 155–170.
- Mishra, G., Nitsch, D., Black, S., De Stavola, B., Kuh, D., Hardy, R., 2009. A structured approach to modelling the effects of binary exposure variables over the life course. *Int. J. Epidemiol.* 38, 528–537.
- Pidsley, R., CC, Y.W., Volta, M., Lunnon, K., Mill, J., Schalkwyk, L.C., 2013. A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genom.* 14, 293.
- Pudney, S., 2003. The road to ruin? Sequences of initiation to drug use and crime in Britain. *Econ. J.* 113, C182–C198.
- Reinius, L.E., Acevedo, N., Joerink, M., Pershagen, G., Dahlen, S.E., Greco, D., Soderhall, C., Scheynius, A., Kere, J., 2012. Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. *PLoS One* 7, e41361.
- Relton, C.L., Gaunt, T., McArdle, W., Ho, K., Duggirala, A., Shihab, H., Woodward, G., Lyttleton, O., Evans, D.M., Reik, W., Paul, Y.L., Ficz, G., Ozanne, S.E., Wipat, A., Flanagan, K., Lister, A., Heijmans, B.T., Ring, S.M., Davey Smith, G., 2015. Data resource profile: accessible resource for integrated epigenomic studies (ARIES). *Int. J. Epidemiol.*
- Rohde, P., Lewinsohn, P.M., Kahler, C.W., Seeley, J.R., Brown, R.A., 2001. Natural course of alcohol use disorders from adolescence to young adulthood. *J. Am. Acad. Child Adolesc. Psychiatry* 40, 83–90.
- Sandfort, T.G., Orr, M., Hirsch, J.S., Santelli, J., 2008. Long-term health correlates of timing of sexual debut: results from a national US study. *Am. J. Public Health* 98, 155–161.
- Smyth, G.K., 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* 3 (Article 3).
- Storey, J.D., Tibshirani, R., 2003. Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. USA* 100, 9440–9445.
- Touleimat, N., Tost, J., 2012. Complete pipeline for Infinium(R) Human Methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation. *Epigenomics* 4, 325–341.
- Viner, R.M., Taylor, B., 2007. Adult outcomes of binge drinking in adolescence: findings from a UK national birth cohort. *J. Epidemiol. Community Health* 61, 902–907.
- de Vocht, F., Simpkin, A.J., Richmond, R.C., Relton, C., Tilling, K., 2015. Assessment of offspring DNA methylation across the lifecourse associated with prenatal maternal smoking using bayesian mixture modelling. *Int. J. Environ. Res. Public Health* 12, 14461–14476.
- Vrieze, S.I., Hicks, B.M., Iacono, W.G., McGue, M., 2012. Decline in genetic influence on the co-occurrence of alcohol, marijuana, and nicotine dependence symptoms from age 14 to 29. *Am. J. Psychiatry* 169, 1073–1081.
- Wiefferink, C.H., Peters, L., Hoekstra, F., Dam, G.T., Buijs, G.J., Paulussen, T.G., 2006. Clustering of health-related behaviors and their determinants: possible consequences for school health interventions. *Prev. Sci.* 7, 127–149.
- Ziller, M.J., Gu, H., Muller, F., Donaghey, J., Tsai, L.T., Kohlbacher, O., De Jager, P.L., Rosen, E.D., Bennett, D.A., Bernstein, B.E., Gnirke, A., Meissner, A., 2013. Charting a dynamic DNA methylation landscape of the human genome. *Nature* 500, 477–481.