

Title:

Unhealthy Behaviours and Parkinson's Disease: A Mendelian Randomisation Study

Running head:

Unhealthy Behaviours and Parkinson's Disease Risk

Authors:

Authors: Karl Heilbron (PhD)^{1*}, Melanie P. Jensen (MBBS)^{2*}, Sara Bandres-Ciga (PhD)³, Pierre Fontanillas (PhD)¹, The 23andMe Research Team, Cornelis Blauwendraat (PhD)³, Mike A. Nalls (PhD)^{3,4}, Andrew B. Singleton (PhD)³, George Davey Smith (M.D.)⁵, Paul Cannon (PhD)¹, Alastair Noyce (PhD)^{2,6}

*Authors contributed equally

Affiliations:

¹23andMe, Inc., 223 N Mathilda Ave, Sunnyvale, CA 94086, USA

²Preventive Neurology Unit, Wolfson Institute of Preventive Medicine, Queen Mary University of London, London, EC1M 6BQ, UK

³Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA

⁴Data Tecnica International, Glen Echo, MD, USA

⁵MRC Integrative Epidemiology Unit, Bristol Medical School, University of Bristol, Bristol, UK, BS8 2BN, UK

⁶Department of Clinical and Movement Neurosciences, UCL Institute of Neurology, London, UK

Corresponding authors:

Name: Karl Heilbron

Address: 23andMe, Inc., 223 N Mathilda Ave, Sunnyvale, CA 94086, USA

Email: kheilbron@23andme.com

Telephone: +1-650-938-6300

Name: Alastair Noyce

Address: Preventive Neurology Unit, Wolfson Institute of Preventive Medicine, Queen Mary University of London, Charterhouse Square, London, UK, EC1M 6BQ

Email: a.noyce@qmul.ac.uk

Telephone: +44 207 882 5841

Number of characters in title: 75

Number of characters in running head: 44

Number of words in the Abstract: 247

Number of words in the Introduction: 498

Number of words in the Discussion: 902

Number of words in the body of the manuscript: 4,341

Number of figures: 4 (figure 1 is black and white, figures 2-4 are colour)

Number of tables: 3

Abstract

Objective: Tobacco smoking, alcohol intake, and high BMI have been identified in observational studies as potentially protective factors against developing Parkinson's disease (PD). Because of the possibility of residual confounding and reverse causation, it is unclear whether such epidemiological associations are causal. Mendelian randomisation (MR) uses genetic variants to explore causal effects of exposures on outcomes; minimising these sources of bias. Using MR, this study sought to determine the causal relationship between tobacco smoking, alcohol intake, and high BMI, and the risk of PD.

Methods: We performed genome-wide association studies to identify single nucleotide polymorphisms associated with the exposures. MR analysis of the relationship between each exposure and PD was undertaken using a split-sample design. The inverse variance weighted (IVW) method was used to combine SNP-specific effect estimates.

Results: Ever-smoking causally reduced risk of PD (OR 0.955; 95% confidence interval [CI] 0.921-0.991; $p=0.013$). An increase in daily alcohol intake causally increased risk of PD (OR 1.125, 95% CI 1.025-1.235; $p=0.013$) and a 1 kg/m² BMI causally reduced risk of PD (OR 0.988, 95% CI 0.979-0.997; $p=0.008$). Sensitivity analyses did not suggest bias from horizontal pleiotropy or invalid instruments.

Interpretation: Using split-sample MR in over 2.4 million participants, we observed a protective effect of smoking on risk of PD, warranting the prioritisation of related therapeutic targets, such as nicotinic agonists, in prevention trials. In contrast to observational data, alcohol consumption causally increased risk of PD. Higher BMI had a protective effect on PD, but the effect was small.

Introduction

Parkinson's disease (PD) is the second most common neurodegenerative disorder, and the number of people living with PD is set to double from 6.9 million in 2015 to 14.2 million in 2040.¹ The mechanisms driving the cardinal pathology of PD – alpha-synucleinopathy and dopaminergic nigrostriatal neurodegeneration – remain unclear. Accordingly, existing therapeutic approaches are symptomatic and have no discernible impact on disease process. Observational studies have identified putative modifiable risk factors for PD but their causal effect on aetio-pathogenesis, hence their value as therapeutic targets for neuroprotection, has not been elucidated.

Tobacco use and alcohol intake are notable as potentially protective factors against being diagnosed with PD, but have detrimental effects on other health outcomes. Observational studies suggest that ever-smokers are 40% less likely to develop PD than never-smokers.² The purported association is plausible given the observed dose-response relationship with pack-years, and replication of the protective effect for oral smokeless tobacco and dietary nicotine.²⁻⁴ Similar associations have been reported with alcohol intake: high versus low (smoking-adjusted) alcohol intake was associated with a 22% lower risk of PD in a meta-analysis involving 670,550 subjects.⁵ A recent large-scale, case-control study supported the negative observational associations with both smoking and alcohol.⁶

The role of body mass index (BMI) in risk of PD is less clear. Low BMI is apparent in patients with PD in case-control studies, however a meta-analysis of 10 prospective studies found no association between premorbid BMI and PD risk.^{7, 8} Conversely, a recent Mendelian randomisation (MR) study pointed towards a protective effect of higher BMI, which was not clearly due to survival bias.⁹

It is unclear whether such epidemiological associations are truly causal. Risk-averse individuals may be both less likely to engage in unhealthy behaviours and more likely to develop PD, leading to unmeasured confounding and/or reverse causation.¹⁰ Reverse causation is plausible given the long prodromal phase of PD.¹¹ Indeed, ease of smoking cessation and weight loss may be prodromal features of PD.^{12, 13} Such factors may explain why randomised controlled trials (RCTs) and crossover trials have failed to demonstrate any disease-modifying or symptomatic benefit of nicotine on PD.¹⁴⁻¹⁸ However, these trials relate to disease progression, rather than disease risk, and the events driving disease progression may be different to those initiating it.¹⁹

Using MR, this study sought evidence for a causal relationship between PD risk (outcome) and three unhealthy behaviours (exposures): tobacco smoking, alcohol intake, and high BMI. In MR, genetic variants associated with the exposure of interest are used as instrumental variables to estimate the effect of the exposure on the outcome.^{20, 21} The random allocation of genetic variants from parent to offspring means that alleles are generally unrelated to confounding factors (measured and unmeasured) which distort inferences from observational data.²² Moreover, the germline genotype cannot be modified by the disease process, minimising bias from reverse causality.²² MR also addresses limitations of RCTs, both ethical (when participants may be assigned to a harmful exposure, *e.g.* smoking) and pragmatic (when adherence cannot be guaranteed, *e.g.* alcohol intake, BMI modulation).²³

Methods

Participants

Participants were customers of 23andMe, Inc., a personal genomics company. Study protocols were approved by an external AAHRPP-accredited institutional review board and conducted in accordance with the Declaration of Helsinki principles. Participants gave informed consent to participate. Data was collected between November 10th, 2007 and January 1st, 2018.

Defining phenotypic traits

Parkinson's disease status

PD cases were drawn from the 23andMe participants who self-reported a diagnosis of PD. For a fuller description of recruitment see Do *et al.*²⁴ A high level of agreement between patient-reported PD diagnosis and neurologist assessment has previously been demonstrated.²⁵ Controls were drawn from the 23andMe participants who self-reported never having been diagnosed with PD. 19,924 PD cases and 2,413,087 controls were included in the analysis.

Unhealthy exposures of interest

Participants self-reported their tobacco smoking habits with response categories “ever-smoker (>100 cigarettes smoked in lifetime)”, or “never-smoker (<100 cigarettes smoked in lifetime)”. Participants self-reported the number of alcohol measures consumed per day over the past two weeks, where 1 measure corresponds to 12 oz. of beer, 5 oz. of wine, or 1.5 oz of spirits. Possible response options were grouped as follows: 0 measures, 0-1 measure, 1 measure, 2 measures, 3 measures, 4 measures, 5 or more measures. Participants self-reported their mass (in kilograms) and height (in metres) squared, from which BMI was calculated.

Genome-wide association studies

We performed new genome-wide association studies for each phenotypic exposure using unrelated individuals. We selected unrelated individuals using a segmental identity-by-descent estimation algorithm.²⁶ Individuals were defined as related if they shared 700 cM identity-by-descent, including regions where the two individuals share either one or both genomic segments identical-by-descent. This level of relatedness (roughly 20% of the genome) corresponds approximately to the minimal expected sharing between first cousins in an outbred population. Ancestry composition was performed as previously reported.²⁷ Inclusion was restricted to individuals of predominantly European ancestry to minimise confounding by ancestry.²⁸

DNA extraction and genotyping were performed on saliva samples by National Genetics Institute. Samples were genotyped on one of five Illumina-based genotyping platforms. The v1 and v2 platforms were variants of the Illumina HumanHap550+ BeadChip, including about 25,000 custom SNPs selected by 23andMe, with a total of about 560,000 SNPs. The v3 platform was based on the Illumina OmniExpress+ BeadChip, with custom content to improve the overlap with our v2 array, with a total of about 950,000 SNPs. The v4 platform was a fully customized array, including a lower redundancy subset of v2 and v3 SNPs with additional coverage of lower-frequency coding variation, and about 570,000 SNPs. The v5 platform is an Illumina Infinium Global Screening Array (~640,000 SNPs) supplemented with ~50,000 SNPs of custom content. Samples had minimum call rates of 98.5%.

We phased participant data using either an internally-developed tool, Finch (V1-V4 genotyping arrays) or Eagle2 (V5 genotyping array).²⁹ Finch implements the Beagle

haplotype graph-based phasing algorithm, modified to separate the haplotype graph construction and phasing steps.³⁰ It extends the Beagle model to accommodate genotyping error and recombination, to handle cases where there are no consistent paths through the haplotype graph for the individual being phased. We constructed haplotype graphs for European and non-European samples on each 23andMe genotyping platform from a representative sample of genotyped individuals, and then performed out-of-sample phasing of all genotyped individuals against the appropriate graph. For the X chromosome, we built separate haplotype graphs for the non-pseudoautosomal region and each pseudoautosomal region, and these regions were phased separately.

Imputation

Imputation panels created by combining multiple smaller panels have been shown to give better imputation performance than the individual constituent panels alone.³¹ To that end, we combined the May 2015 release of the 1000 Genomes Phase 3 haplotypes with the UK10K imputation reference panel to create a single unified imputation reference panel.^{32, 33} To do this, multiallelic sites with N alternate alleles were split into N separate biallelic sites. We then removed any site whose minor allele appeared in only one sample. For each chromosome, we used Minimac3 to impute the reference panels against each other, reporting the best-guess genotype at each site.³⁴ This gave us calls for all samples over a single unified set of variants. We then joined these together to get, for each chromosome, a single VCF with phased calls at every site for 6,285 samples.

In preparation for imputation we split each chromosome of the reference panel into chunks of no more than 300,000 variants, with overlaps of 10,000 variants on each side. We used a single batch of 10,000 individuals to estimate Minimac3 imputation model parameters for

each chunk.³⁴ We imputed phased participant data against the chunked merged reference panel using Minimac3, treating males as homozygous pseudo-diploids for the non-pseudoautosomal region. Throughout, we treated structural variants and small indels the same as SNPs.

Association tests

We computed association test results by regression assuming additive allelic effects (logistic regression for case-control exposures, linear regression for quantitative exposures). We included covariates for age, sex, the top five genetic principal components to account for residual population structure, and indicators for genotype platforms to account for genotype batch effects. The association test p-value we report was computed using a likelihood ratio test. For tests using imputed data, we use the imputed dosages rather than best-guess genotypes. For the X chromosome, male genotypes were coded as if they were homozygous diploid for the observed allele.

Principal component analysis

We performed the genetic principal components analysis using ~65,000 high-quality genotyped variants present in all five genotyping platforms and a random sample of one million research participants with predominantly European ancestry. Principal component scores for participants not included in the analysis were obtained by projection, combining the eigenvectors of the analysis and the SNP weights.

Quality control of genotyped GWAS results

We excluded SNPs that: 1) had a call rate <90%, 2) had a Hardy-Weinberg $P < 10^{-20}$ in people with predominantly European ancestry, 3) were only genotyped on the V1 and/or V2

platforms, 4) were found on the mitochondrial chromosome or the Y chromosome, 5) failed a test for parent-offspring transmission (specifically, we regressed the child's allele count against the mean parental allele count and excluded SNPs with fitted $\beta < 0.6$ and $P < 10^{-20}$ for a test of $\beta < 1$), 6) had an association with genotyping date ($P < 10^{-50}$ by ANOVA of SNP genotypes against a factor dividing genotyping date into 20 roughly equal-sized buckets), 7) had a large sex effect (ANOVA of SNP genotypes, $r^2 > 0.1$), or 8) had probes matching multiple genomic positions in the reference genome.

Quality control of imputed GWAS results

We excluded SNPs with imputed $r^2 < 0.3$, as well as SNPs that had strong evidence of a platform batch effect. For each SNP we identified the largest subset of the data passing other quality control criteria based on their original genotyping platform – either v2+v3+v4+v5, v4+v5, v4, or v5 only – and computed association test results for the largest passing set. The batch effect test is an F test from an ANOVA of the SNP dosages against a factor representing the V4 or V5 platform; we excluded results with $P < 10^{-50}$.

Additional quality control of GWAS results

Across both genotyped and imputed GWAS results, we excluded SNPs that had sample size of less than 20% of the total GWAS sample size. We also removed SNPs that did not converge during logistic regression, as identified by $\text{abs}(\text{effect}) > 10$ or $\text{stderr} > 10$ on the log odds scale. We removed SNPs with $\text{MAF} < 0.1\%$ from linear regressions because these SNPs are sensitive to violations of the regression assumption of normally distributed residuals. If SNPs were both genotyped and imputed, and they passed QC for both, we used results from the imputed analysis.

After quality control, we had analysed 904,040 genotyped SNPs and 25,208,208 imputed SNPs.

Instrument construction

SNPs associated with each of the exposures at the genome-wide significance level ($p < 5e^{-8}$) were included as instrumental variables. We excluded SNPs with a minor allele frequency $< 3\%$ and SNPs in the *HLA* locus (hg19, chromosome 6, 26.0Mbp - 33.7Mbp). Instrument strength was assessed using the mean F statistic, as calculated by the `system_metrics` function in the `TwoSampleMR` R package.³⁵

Split-sample MR analysis

We performed MR analysis of the relationship between the exposures of interest and PD using a split-sample design, in accordance with published methods.^{22, 36} Individuals from the 23andMe cohort were randomly allocated into two evenly sized groups (Table 1). The instrument-exposure association was measured in the first group (fold 1), and the instrument-outcome association was measured in the second group (fold 2), and MR analyses undertaken. We then repeated the MR analyses, but using fold 1 for the outcome and fold 2 for the exposure. For each MR method, this resulted in two independent MR estimates, which were combined using an inverse variance weighted (IVW) fixed-effects meta-analysis.

The effect of an exposure on PD was calculated for each SNP using the Wald ratio method.³⁷ In the IVW analysis we performed a linear regression constrained through the origin of the variant-exposure and the variant-outcome associations for each instrument, weighted by their inverse variance (Table 2).

We used three methods to assess the impact of bias on IVW estimates.³⁵ First, heterogeneity in Wald ratio estimates was assessed using the Cochran's Q test and I^2 index.³⁸ Heterogeneity can indicate the possibility of bias due to horizontal pleiotropy, whereby variants influence the outcome via pathways unrelated to the exposure. Individual variant contributions to Cochran's Q statistic was calculated, and variants were excluded in the heterogeneity filtering analysis if their contribution surpassed the Bonferroni-corrected 99.8th percentile of the χ^2 (1 df) distribution.

Second, we performed MR-Egger analysis to assess the magnitude of bias in the IVW estimate occurring due to directional horizontal pleiotropy (Table 2). A hypothesis test can be used to determine whether the intercept is different from zero and this indicates a net horizontal pleiotropic effect, but is generally underpowered. In the absence of unbalanced pleiotropy, a funnel plot of the individual variant effects plotted against the inverse of their standard error will be symmetrically distributed around the point estimate.

Third, we used the weighted median method, which reports a valid MR estimate under conditions where up to 50% of the weight of the instruments are invalid (Table 2).³⁹ The estimate is obtained by calculating the Wald ratio for each SNP and selecting the median value after weighting by inverse variance. Confidence intervals for the weighted median estimates are calculated through bootstrapping.

To minimize misallocation of variants to the exposure rather than outcome group, we applied Steiger filtering to all analyses to remove genetic variants that had a stronger correlation with the outcome than with the exposure.⁴⁰

Replication in the International Parkinson's Disease Genomics Consortium (IPDGC) dataset

Summary statistics from the largest published PD GWAS meta-analysis were used as the outcome data for replication purposes after excluding data from 23andMe (the “IPDGC dataset”).⁴¹ The outcome summary statistics used for this analysis included 15,074 cases, 18,618 proxy cases, and 449,056 controls, and there were 17,410,431 genotyped and imputed SNPs tested for association with PD. Recruitment and genotyping quality control are described in the original report.⁴¹ Only exposure SNPs from fold 1 were used in the replication analysis, and we excluded SNPs on the X chromosome and palindromic SNPs.

Sensitivity analyses

If tobacco use truly has a protective effect on PD, we would expect genetic variants that increase the amount of tobacco used should decrease PD risk in tobacco users, but should have no effect on PD risk in non-users. If these variants are protective for PD in separate analyses of both ever-smokers and never-smokers, however, this suggests that the protective effect is mediated by a pleiotropic pathway unrelated to tobacco use. Given that ever/never smoker status has a positive genetic correlation with “number of cigarettes smoked per day” ($\rho_g=0.366$, $p=4.5 \times 10^{-4}$), we re-ran MR analyses separately for ever-smokers and never-smokers.⁴² We also repeated this analysis using a single SNP (rs16969968) that leads to a missense mutation in *CHRNA5* and has been shown to affect the amount of tobacco use within tobacco users.⁴³ In both analyses, we re-computed the effect of our instrument variables on PD in the same cohort, but stratifying on tobacco use status. We employed the same estimates of the effect of the tobacco use SNPs as in the main analysis. For the *CHRNA5* SNP, we used the estimate from Millard et al. that each allele that increases tobacco use is associated with an odds ratio of 1.21 for being a “heavy smoker” (95% CI: 1.19-1.23).⁴³

Statistical analysis

Analyses were performed using R statistical software 3.3.2 (2016-10-31).

Results

Tobacco use

385 SNPs were associated with self-reported smoking status (ever- versus never-smoker) in fold 1 (422 SNPs in fold 2, Tables 1 and 2 in Supplementary Data). This instrument explained an average of 1.16% of the variance in tobacco smoking liability across folds (fold 1 = 1.12%, fold 2 = 1.19%). The F statistic ($F= 54.3$ in fold 1, $F=52.6$ in fold 2) was high; validating this instrument for these analyses.

In the IVW analysis, ever-smoking caused a reduction in PD risk with an OR 0.955 (95% CI 0.921-0.991, $p = 0.013$) (Fig 1). There was no clear evidence of heterogeneity in estimates derived from individual SNPs (fold 1-versus-fold 2: Cochran's $Q = 184.3$, $p = 1.000$; fold 2-versus-fold 1: Cochran's $Q = 182.4$, $p = 1.000$). The MR-Egger intercept was not significant ($p=0.193$) and the estimate from MR-Egger was OR 0.867 (95% CI 0.747-1.006, $p = 0.061$). Funnel plots (Fig 2 A to B) suggested that individual variants were symmetrically distributed around the point estimate. Together these findings suggest no meaningful bias through unbalanced horizontal pleiotropy. The weighted median analysis was similar to the IVW estimate, suggesting minimal bias due to invalid instruments (OR = 0.927, 95% CI 0.857-1.002, $p = 0.058$). We replicated the IVW results in the IPDGC dataset (OR = 0.865, 95% CI 0.756-0.990, $p = 0.035$) and a meta-analysis of the results from the two datasets yielded a stronger p-value than results from either dataset individually (OR = 0.949, 95% CI 0.916-0.983, $p = 0.003$) (Fig 1).

As a sensitivity analysis, we re-ran MR in the 23andMe dataset separately for ever-smokers and never-smokers. As expected, we found a similar protective effect, although less precisely estimated, in ever-smokers (OR = 0.966, 95% CI 0.885-1.055) and little evidence of a

protective effect in never-smokers (OR = 1.030, 95% CI 0.966-1.097). We found similarly imprecise estimates when performing MR using a SNP in *CHRNA5* that strongly influences the amount of tobacco use within tobacco users (ever-smokers: OR = 1.037, 95% CI 0.888-1.211; never-smokers: OR = 1.033, 95% CI 0.849-1.256).

Alcohol intake

129 SNPs were associated with self-reported alcohol in fold 1 (124 SNPs in fold 2, Tables 3 and 4 in Supplementary Data). This instrument explained an average of 0.53% of the variance in alcohol intake liability across folds (fold 1 = 0.53%, fold 2 = 0.50%). The F statistic (F=49.5 in fold 1, F=48.3 in fold 2) again validating this instrument for alcohol intake.

In the IVW analysis, alcohol intake caused an increase in PD risk with an OR 1.125 for a 1-group increase in daily alcohol intake (95% CI 1.025-1.235, $p = 0.013$) (Fig 1). There was no clear evidence of heterogeneity against estimates derived from individual SNPs (fold 1-versus-fold 2: Cochran's $Q = 75.3$, $p = 1.000$; fold 2-versus-fold 1: Cochran's $Q = 70.9$, $p = 1.000$). The MR-Egger intercept was not significant ($p=0.152$) and the estimate from MR-Egger was OR 1.438 (95% CI 1.014-2.038, $p = 0.041$). Funnel plots (Fig 3A to B) suggested that individual variants were symmetrically distributed around the point estimate, again suggesting no meaningful bias through unbalanced horizontal pleiotropy. The weighted median estimate was consistent with the IVW estimate, but with wider confidence intervals, suggesting minimal bias due to invalid instruments (OR = 1.126, 95% CI 0.943-1.345, $p = 0.189$). The point estimate in the IPDGC dataset was similar to the 23andMe result, but the confidence intervals were wide (OR = 1.163, 95% CI 0.856-1.580, $p = 0.334$). A meta-

analysis of the results from the two datasets yielded a stronger p-value than results from either dataset individually (OR = 1.128, 95% CI 1.032-1.233, p = 0.008) (Fig 1).

BMI

729 SNPs were associated with self-reported alcohol in fold 1 (693 SNPs in fold 2, Tables 5 and 6 in Supplementary Data). This instrument explained an average of 5.31% of the variance in BMI liability across folds (fold 1 = 5.31%, fold 2 = 5.15%). The F statistic (F=99.7 in fold 1, F=102.2 in fold 2) indicating the validity of this instrument for BMI.

In the IVW analysis, a genetically-estimated 1 kg/m² increase in BMI causally reduced PD risk with an OR 0.988 (95% CI 0.979-0.997, p = 0.008) (Fig 1). There was no clear evidence of heterogeneity against estimates derived from individual SNPs (fold 1-versus-fold 2: Cochran's Q = 184.3, p = 1.000; fold 2-versus-fold 1: Cochran's Q = 539.8, p = 1.000). The MR-Egger intercept was not significant (p=0.223) and the estimate from MR-Egger was OR 0.977 (95% CI 0.957-0.997, p = 0.023). Funnel plots (Fig 4A to B) suggested that individual variants were symmetrically distributed around the point estimate, again suggesting no meaningful bias through unbalanced horizontal pleiotropy. The weighted median estimate was consistent with the IVW estimate, suggesting minimal bias due to invalid instruments (OR = 0.985, 95% CI 0.965-1.005, p = 0.132). The causal OR in the IPDGC dataset was similar to the 23andMe result, but the confidence intervals were wide (OR = 0.981, 95% CI 0.958-1.004, p = 0.106). A meta-analysis of the results from the two datasets yielded a stronger p-value than results from either dataset individually (OR = 0.987, 95% CI 0.979-0.995, p = 0.002) (Fig 1).

Discussion

Using split-sample MR, we observed a protective effect of smoking on risk of PD. In a cohort of over 2.4 million participants, ever-smoking caused a 5% reduction in PD risk compared to never-smoking. Sensitivity analyses using MR-Egger and MR-median suggested that the effect was unlikely to be driven by horizontal pleiotropy, outliers, or invalid instruments. To our knowledge, this represents the largest MR study to date exploring the link between smoking and PD, and provides cautious optimism to continue to explore nicotine as a putative preventive measure.

These findings are concordant with observational studies consistently demonstrating an inverse association between smoking and PD risk.² However, observational studies have not definitively ruled out reverse causality. For example, in a study of over 220,000 individuals, the protective effect of smoking on PD risk held true among ex-smokers who had quit 9-18 years before recruitment, but not among smokers who had quit >20 years before recruitment, meaning the possibility that preclinical dopaminergic changes facilitate smoking cessation could not be excluded.⁴⁴ Similarly, a cohort study reported that parental smoking during childhood reduced future PD risk, however the possibility of a transgenerational exposure, such as a toxic agent influencing both parental smoking behaviour and PD risk in children, could not be excluded.⁴⁵ By circumventing the use of proxy measures to overcome unmeasured confounding, residual confounding and reverse causality, the findings from MR analyses more robustly support a causal effect.

Our results are similar to those from smaller MR studies (Table 3) that have tested for a causal relationship between PD and tobacco use. Grover *et al.* found a significant protective effect (OR = 0.71 per log odds of ever smoking; 95% CI 0.57-0.90), while Nalls *et al.* found

a non-significant protective effect (OR = 0.94 per log odds of ever smoking; 95% CI 0.88-1.00, $p = 0.063$).^{41,46} The stringency of our instrument selection and sample size, particularly compared with the first of those studies, makes the estimate from the present study more reliable.

Early interventional studies (case series and pilot open-label trials) demonstrated beneficial effects of nicotine on motor and cognitive deficits in PD.⁴⁷⁻⁵⁴ However placebo-controlled studies largely failed to replicate the protective association between smoking and PD.¹⁴⁻¹⁸ This may be because trials have assessed transient changes in PD symptoms post-nicotine in small cohorts. Given the heterogeneity in PD progression, demonstrating disease modification following treatment would necessitate large sample sizes and long follow-up.⁵⁵ An RCT currently in progress is incorporating prolonged treatment and wash-out phases pre-endpoint assessment, and will therefore better evaluate potential neuroprotective effects of nicotine on PD (ClinicalTrials.gov Identifier: NCT01560754). Trials may also have failed because they examined the disease-modifying effect of nicotine on established PD. However the mechanisms underlying PD risk may be different to those driving disease progression. Our study suggests that smoking influences the former, highlighting the need for prevention trials of candidate disease-modifying therapies in pre-manifest PD.

Mechanisms underlying the effect of smoking on PD risk remain speculative. However a disease-modifying effect of nicotine is biologically plausible given evidence that it mitigates MPTP- and 6-hydroxydopamine-induced nigrostriatal damage and motor dysfunction in rodent models and non-human primate models of PD.⁵⁶⁻⁶²

In this MR study we observed a causal increase in risk of PD with higher alcohol intake. This finding is in disagreement with a recent meta-analysis finding that alcohol consumption was associated with a lower risk of PD in retrospective case-control studies (OR for never vs. heavy/moderate drinking: 0.74; 95% CI 0.64-0.85).⁶³ When the same study meta-analysed prospective cohort studies, however, the difference was not significant.⁶³ It is therefore plausible that the inverse association observed in case-control studies is driven by reverse causation, whereby lower alcohol consumption in cases reflects a low dopaminergic state, suppressing addictive behaviours. Our finding is consistent with the known neurotoxic effect of excess alcohol consumption in human neuropathological studies, and with experimental studies in rodents showing that alcohol reduces dopamine in the midbrain and increases oxidative damage in nigral cells.^{64, 65} Our results suggest that alcohol consumption is unlikely to help prevent PD, and may increase risk of PD.

We have previously observed a causal protective effect of higher BMI on risk of PD, an effect that was not clearly explained by survival bias.⁹ More recently, we used a hypothesis-free approach to study causal associations between a range of exposures and PD using MR, and observed that most of the top hits related to a protective effect of increased adiposity.⁶⁶ The results of the current study, undertaken in a large case-control settings but with some overlapping samples, support these previous observations.

Limitations of our study include that we did not explore whether effects differed between sub-groups (for example between genders, as previously suggested for the association between PD and smoking).² In addition, the MR analysis assumed linearity, precluding us from identifying non-linear exposure-outcome associations. Finally, it has been argued that

survival bias may distort MR analyses. However our study used a two-sample design, which has been shown to minimise the impact of survival bias on effect estimates.⁶⁷

In conclusion, we provide evidence to support a causal protective effect of smoking and high BMI on PD risk. Conversely, we observed a causal detrimental effect of alcohol on PD risk. Although a better understanding of the underlying mechanisms, as well as development of safe delivery methods is necessary, such findings help guide the prioritisation of candidate neuroprotective approaches for RCTs in participants at higher risk of developing PD.

Acknowledgements

We thank the research participants from 23andMe and the IPDGC who contributed to this work. The work was financially supported by The Michael J. Fox Foundation for Parkinson's Research (grant MJFF12737), National Institute on Aging (NIA) part of the National Institutes of Health (NIH) and by 23andMe, Inc.

Author Contributions

KH contributed to the conception and design of the study, acquired and analysed the data, and drafted the figures. MJ contributed to the conception and design of the study, and drafted the manuscript. SB acquired and analysed the data. PF, CB, MAN, AS, GDS and PC contributed to data analysis and editing the manuscript. AJN led the study group, contributed to the conception and design of the study, and edited the manuscript.

Members of the 23andMe Research Team (23andMe, Inc., 223 N Mathilda Ave, Sunnyvale, CA 94086, USA), who contributed to data acquisition, are: Michelle Agee, Stella Aslibekyan, Adam Auton, Robert K. Bell, Katarzyna Bryc, Sarah K. Clark, Sarah L. Elson, Kipper Fletez-Brant, Nicholas A. Furlotte, Pooja M. Gandhi, Barry Hicks, David A. Hinds, Karen E. Huber, Ethan M. Jewett, Yunxuan Jiang, Aaron Kleinman, Keng-Han Lin, Nadia K. Litterman, Marie K. Luff, Matthew H. McIntyre, Kimberly F. McManus, Joanna L. Mountain, Sahar V. Mozaffari, Priyanka Nandakumar, Elizabeth S. Noblin, Carrie A.M. Northover, Jared O'Connell, Aaron A. Petrakovitz, Steven J. Pitts, G. David Poznik, J. Fah Sathirapongsasuti, Janie F. Shelton, Suyash Shringarpure, Chao Tian, Joyce Y. Tung, Robert J. Tunney, Vladimir Vacic, and Xin Wang.

Conflicts of Interest

There are no conflicts of interest to report which could bias the authors in the conduct of the reported work.

References

1. Dorsey ER, Bloem BR. The Parkinson Pandemic-A Call to Action. *JAMA Neurol.* 2018 Jan 1;75(1):9-10.
2. Li X, Li W, Liu G, Shen X, Tang Y. Association between cigarette smoking and Parkinson's disease: A meta-analysis. *Arch Gerontol Geriatr.* 2015 Nov-Dec;61(3):510-6.
3. Yang F, Pedersen NL, Ye W, et al. Moist smokeless tobacco (Snus) use and risk of Parkinson's disease. *Int J Epidemiol.* 2017 Jun 1;46(3):872-80.
4. Nielsen SS, Franklin Gm Fau - Longstreth WT, Longstreth Wt Fau - Swanson PD, Swanson Pd Fau - Checkoway H, Checkoway H. Nicotine from edible Solanaceae and risk of Parkinson disease. (1531-8249 (Electronic)).
5. Zhang D, Jiang H, Xie J. Alcohol intake and risk of Parkinson's disease: a meta-analysis of observational studies. *Movement disorders : official journal of the Movement Disorder Society.* 2014 May;29(6):819-22.
6. Heilbron K, Noyce AJ, Fontanillas P, Alipanahi B, Nalls MA, Cannon P. The Parkinson's phenome-traits associated with Parkinson's disease in a broadly phenotyped cohort. *NPJ Parkinson's disease.* 2019;5:4.
7. van der Marck MA, Dicke HC, Uc EY, et al. Body mass index in Parkinson's disease: a meta-analysis. *Parkinsonism Relat Disord.* 2012 Mar;18(3):263-7.
8. Wang YL, Wang YT, Li JF, Zhang YZ, Yin HL, Han B. Body Mass Index and Risk of Parkinson's Disease: A Dose-Response Meta-Analysis of Prospective Studies. *PLoS One.* 2015;10(6):e0131778.
9. Noyce AJ, Kia DA, Hemani G, et al. Estimating the causal influence of body mass index on risk of Parkinson disease: A Mendelian randomisation study. *PLoS Med.* 2017 Jun;14(6):e1002314.

10. Evans AH, Lawrence AD, Potts J, et al. Relationship between impulsive sensation seeking traits, smoking, alcohol and caffeine intake, and Parkinson's disease. *J Neurol Neurosurg Psychiatry*. 2006 Mar;77(3):317-21.
11. Schrag A, Horsfall L, Walters K, Noyce A, Petersen I. Prediagnostic presentations of Parkinson's disease in primary care: a case-control study. *Lancet neurol*. 2015 Jan;14(1):57-64.
12. Ritz B, Lee PC, Lassen CF, Arah OA. Parkinson disease and smoking revisited: ease of quitting is an early sign of the disease. *Neurology*. 2014 Oct 14;83(16):1396-402.
13. Durcan R, Wiblin L, Lawson RA, et al. Prevalence and duration of non-motor symptoms in prodromal Parkinson's disease. *Eur J Neurol*. 2019 Jul;26(7):979-85.
14. Clemens P, Baron JA, Coffey D, Reeves A. The short-term effect of nicotine chewing gum in patients with Parkinson's disease. *Psychopharmacology (Berl)*. 1995 Jan;117(2):253-6.
15. Randomized placebo-controlled study of the nicotinic agonist SIB-1508Y in Parkinson disease. *Neurology*. 2006 Feb 14;66(3):408-10.
16. Vieregge A, Sieberer M, Jacobs H, Hagenah JM, Vieregge P. Transdermal nicotine in PD: a randomized, double-blind, placebo-controlled study. *Neurology*. 2001 Sep 25;57(6):1032-5.
17. Villafane G, Thiriez C, Audureau E, et al. High-dose transdermal nicotine in Parkinson's disease patients: a randomized, open-label, blinded-endpoint evaluation phase 2 study. *Eur J Neurol*. 2018 Jan;25(1):120-7.
18. Ebersbach G, Stock M, Muller J, Wenning G, Wissel J, Poewe W. Worsening of motor performance in patients with Parkinson's disease following transdermal nicotine administration. *Mov Disord*. 1999 Nov;14(6):1011-3.

19. Paternoster L, Tilling K, Davey Smith G. Genetic epidemiology and Mendelian randomization for informing disease therapeutics: Conceptual and methodological challenges. *PLoS Genet*. 2017 Oct;13(10):e1006944.
20. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet*. 2014 Sep 15;23(R1):R89-98.
21. Bandres-Ciga S, Noyce AJ, Traynor BJ. Mendelian Randomization-A Journey From Obscurity to Center Stage With a Few Potholes Along the Way. *JAMA Neurol*. 2019 Oct 14.
22. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol*. 2015 Apr;44(2):512-25.
23. Noyce AJ, Nalls MA. Mendelian Randomization - the Key to Understanding Aspects of Parkinson's Disease Causation? *Mov Disord*. 2016 Apr;31(4):478-83.
24. Do CB, Tung JY, Dorfman E, et al. Web-based genome-wide association study identifies two novel loci and a substantial genetic component for Parkinson's disease. *PLoS genetics*. 2011 Jun;7(6):e1002141.
25. Dorsey ER, Darwin KC, Mohammed S, et al. Virtual research visits and direct-to-consumer genetic testing in Parkinson's disease. *Digital health*. 2015 Jan-Dec;1:2055207615592998.
26. Henn BM, Hon L, Macpherson JM, et al. Cryptic distant relatives are common in both isolated and cosmopolitan genetic samples. *PloS one*. 2012;7(4):e34267.
27. Durand EY, Do CB, Mountain JL, Macpherson JM. Ancestry Composition: A Novel, Efficient Pipeline for Ancestry Deconvolution. *bioRxiv*. 2014:010512.
28. Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *Bmj*. 2018 Jul 12;362:k601.

29. Loh PR, Danecek P, Palamara PF, et al. Reference-based phasing using the Haplotype Reference Consortium panel. *Nat Genet.* 2016 Nov;48(11):1443-8.
30. Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *American journal of human genetics.* 2007 Nov;81(5):1084-97.
31. Huang J, Howie B, McCarthy S, et al. Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. *Nature communications.* 2015 Sep 14;6:8111.
32. Auton A, Brooks LD, Durbin RM, et al. A global reference for human genetic variation. *Nature.* 2015 Oct 1;526(7571):68-74.
33. Walter K, Min JL, Huang J, et al. The UK10K project identifies rare variants in health and disease. *Nature.* 2015 Oct 1;526(7571):82-90.
34. Das S, Forer L, Schonherr S, et al. Next-generation genotype imputation service and methods. *Nature genetics.* 2016 Oct;48(10):1284-7.
35. Hemani G, Bowden J, Haycock P, et al. Automating Mendelian randomization through machine learning to construct a putative causal map of the human phenome. *bioRxiv.* 2017:173682.
36. Richmond RC, Davey Smith G, Ness AR, den Hoed M, McMahon G, Timpson NJ. Assessing causality in the association between child adiposity and physical activity levels: a Mendelian randomization analysis. *PLoS Med.* 2014 Mar;11(3):e1001618.
37. Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat Med.* 2008 Apr 15;27(8):1133-63.

38. Greco MF, Minelli C, Sheehan NA, Thompson JR. Detecting pleiotropy in Mendelian randomisation studies with summary data and a continuous outcome. *Statistics in medicine*. 2015 Sep 20;34(21):2926-40.
39. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genetic epidemiology*. 2016 May;40(4):304-14.
40. Hemani G, Tilling K, Davey Smith G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS genetics*. 2017 Nov;13(11):e1007081.
41. Nalls MA, Blauwendraat C, Vallerga CL, et al. Expanding Parkinson's disease genetics: novel risk loci, genomic context, causal insights and heritable risk. *bioRxiv*. 2019:388165.
42. Bulik-Sullivan B, Finucane HK, Anttila V, et al. An atlas of genetic correlations across human diseases and traits. *Nat Genet*. 2015 Nov;47(11):1236-41.
43. Millard LAC, Munafò MR, Tilling K, Wootton RE, Smith GD. MR-pheWAS with stratification and interaction: Searching for the causal effects of smoking heaviness identified an effect on facial aging. *bioRxiv*. 2018:441907.
44. Gallo V, Vineis P, Cancellieri M, et al. Exploring causality of the association between smoking and Parkinson's disease. *International journal of epidemiology*. 2018 Nov 20.
45. O'Reilly EJ, Chen H, Gardener H, Gao X, Schwarzschild MA, Ascherio A. Smoking and Parkinson's disease: using parental smoking as a proxy to explore causality. *American journal of epidemiology*. 2009 Mar 15;169(6):678-82.
46. Grover S, Lill CM, Kasten M, Klein C, Del Greco MF, König IR. Risky behaviors and Parkinson disease: A mendelian randomization study. *Neurology*. 2019 Sep 16.

47. Moll H. The Treatment of Post-Encephalitic Parkinsonism by Nicotine. *Br Med J*. 1926 Jun 26;1(3416):1079-81.
48. Marshall J, Schnieden H. Effect of adrenaline, noradrenaline, atropine, and nicotine on some types of human tremor. *J Neurol Neurosurg Psychiatry*. 1966 Jun;29(3):214-8.
49. Ishikawa A, Miyatake T. Effects of smoking in patients with early-onset Parkinson's disease. *J Neurol Sci*. 1993 Jul;117(1-2):28-32.
50. Mitsuoka T, Kaseda Y, Yamashita H, et al. Effects of nicotine chewing gum on UPDRS score and P300 in early-onset parkinsonism. *Hiroshima journal of medical sciences*. 2002 Mar;51(1):33-9.
51. Kelton MC, Kahn HJ, Conrath CL, Newhouse PA. The effects of nicotine on Parkinson's disease. *Brain and cognition*. 2000 Jun-Aug;43(1-3):274-82.
52. Villafane G, Cesaro P, Rialland A, et al. Chronic high dose transdermal nicotine in Parkinson's disease: an open trial. *Eur J Neurol*. 2007 Dec;14(12):1313-6.
53. Lemay S, Chouinard S, Blanchet P, et al. Lack of efficacy of a nicotine transdermal treatment on motor and cognitive deficits in Parkinson's disease. *Prog Neuropsychopharmacol Biol Psychiatry*. 2004 Jan;28(1):31-9.
54. Fagerstrom KO, Pomerleau O, Giordani B, Stelson F. Nicotine may relieve symptoms of Parkinson's disease. *Psychopharmacology (Berl)*. 1994 Sep;116(1):117-9.
55. Greenland JC, Williams-Gray CH, Barker RA. The clinical heterogeneity of Parkinson's disease and its therapeutic implications. *The European journal of neuroscience*. 2018 Jul 30.
56. Visanji NP, O'Neill MJ, Duty S. Nicotine, but neither the alpha4beta2 ligand RJR2403 nor an alpha7 nAChR subtype selective agonist, protects against a partial 6-hydroxydopamine lesion of the rat median forebrain bundle. *Neuropharmacology*. 2006 Sep;51(3):506-16.

57. Ryan RE, Ross SA, Drago J, Loiacono RE. Dose-related neuroprotective effects of chronic nicotine in 6-hydroxydopamine treated rats, and loss of neuroprotection in alpha4 nicotinic receptor subunit knockout mice. *Br J Pharmacol*. 2001 Apr;132(8):1650-6.
58. Parain K, Marchand V, Dumery B, Hirsch E. Nicotine, but not cotinine, partially protects dopaminergic neurons against MPTP-induced degeneration in mice. *Brain Res*. 2001 Feb 2;890(2):347-50.
59. Huang LZ, Parameswaran N, Bordia T, Michael McIntosh J, Quik M. Nicotine is neuroprotective when administered before but not after nigrostriatal damage in rats and monkeys. *J Neurochem*. 2009 May;109(3):826-37.
60. Costa G, Abin-Carriquiry JA, Dajas F. Nicotine prevents striatal dopamine loss produced by 6-hydroxydopamine lesion in the substantia nigra. *Brain Res*. 2001 Jan 12;888(2):336-42.
61. Quik M, Chen L, Parameswaran N, Xie X, Langston JW, McCallum SE. Chronic oral nicotine normalizes dopaminergic function and synaptic plasticity in 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine-lesioned primates. *The Journal of neuroscience : the official journal of the Society for Neuroscience*. 2006 Apr 26;26(17):4681-9.
62. Quik M, Parameswaran N, McCallum SE, et al. Chronic oral nicotine treatment protects against striatal degeneration in MPTP-treated primates. *J Neurochem*. 2006 Sep;98(6):1866-75.
63. Jimenez-Jimenez FJ, Alonso-Navarro H, Garcia-Martin E, Agundez JAG. Alcohol consumption and risk for Parkinson's disease: a systematic review and meta-analysis. *Journal of neurology*. 2019 Aug;266(8):1821-34.
64. Harper C, Kril J, Daly J. Are we drinking our neurones away? *British medical journal (Clinical research ed)*. 1987 Feb 28;294(6571):534-6.

65. Ambhore N, Antony S, Mali J, Kanhed A, Bhalerao A, Bhojraj S. Pharmacological and biochemical interventions of cigarette smoke, alcohol, and sexual mating frequency on idiopathic rat model of Parkinson's disease. *Journal of young pharmacists : JYP*. 2012 Jul;4(3):177-83.
66. Noyce AJ, Bandres-Ciga S, Kim J, et al. The Parkinson's Disease Mendelian Randomization Research Portal. *bioRxiv*. 2019:604033.
67. Smit RA, Trompet S, Dekkers OM, Jukema JW, le Cessie S. Survival bias in Mendelian randomization studies: a threat to causal inference. *Epidemiology*. 2019 Aug 1.

Figure Legends

Figure 1: Forest plot of Mendelian randomisation causal association estimates between risk of Parkinson's disease and unhealthy behaviours derived from meta-analysis of the 23andMe and IPDGC datasets. The pooled odds ratio (OR), derived from meta-analysis of the inverse variance weighted estimates, and 95% confidence intervals are shown. For smoking, the unit of exposure is never versus ever smoking. For alcohol, the unit of exposure is 1-group difference in daily alcohol intake. For BMI, the unit of exposure is 1 kg/m². OR = odds ratio; PD = Parkinson's disease.

Figure 2: Funnel plots of individual variant effects for the smoking instrument (from fold 1 (A) and fold 2 (B)) plotted against the inverse of their standard error.

Figure 3: Funnel plots of individual variant effects for the alcohol instrument (from fold 1 (A) and fold 2 (B)) plotted against the inverse of their standard error.

Figure 4: Funnel plots of individual variant effects for the BMI instrument (from fold 1 (A) and fold 2 (B)) plotted against the inverse of their standard error.

Tables

Table 1

Table legend: Participant numbers in fold 1 and fold 2.

	Exposure variables			Outcome variable
	Smoking	Alcohol	BMI	Parkinson's
Fold 1	Ever-smokers: 533,995	1,203,903	1,268,201	Cases: 9,852
	Never-smokers: 728,610			Controls: 1,206,650
Fold 2	Ever-smokers: 533,704	1,204,410	1,267,856	Cases: 10,072
	Never-smokers: 728,847			Controls: 1,206,437

Table 2

Table legend: Further detail on the characteristics of MR methods used to pool effect estimates.

Abbreviations: IVW = inverse variance weight; SNP = single nucleotide polymorphism; IV = instrumental variable.

	Practical implications	Assumptions
IVW method	Meta-analyses effect estimates from multiple independent SNPs weighted by inverse variance, that is more precise estimates contribute more to the pooled estimate.	Valid as long as the three core IV assumptions are upheld. These are: <ol style="list-style-type: none">1. The SNPs are robustly associated with the exposure.2. The SNPs are not associated with confounding factors.3. The pathway through which the SNP affects the outcome is only via the exposure and not via an alternative pathway (horizontal pleiotropy). In practice this assumption is often violated and necessitates the use of other methods to explore the pooled causal effect.

Practically, this method gives the same estimate as the slope of a line of best fit constrained to pass through the origin on a

scatter plot with the SNP-exposure on the X-axis and SNP-outcome on the Y-axis.

This method results in the highest statistical power to detect a causal effect as long as three core IV assumptions are upheld.

In theory, a violation of the IV assumptions for 1 or more SNPs in the instrument could render the pooled estimate biased.

MR Egger method

This method relaxes the third IV assumption that there is no directional pleiotropy in one, some, or all of the SNPs in the combined instrument.

Relaxation of the third IV assumption, results in the introduction of a new assumption called Instrument Strength Independent of Direct Effect (InSIDE).²²

This method provides a better causal estimate in the presence of directional, horizontal pleiotropy and can be compared to the IVW estimated to determine whether the IVW might be biased.

This assumption allows there to be horizontal pleiotropy as long as the combined pleiotropic effects of SNPs are not correlated with their effects on the exposure.

Practically, this method gives the same estimate as the slope of the line of best fit that is not constrained to pass through the origin on a SNP-exposure / SNP-outcome scatter plot. The Y intercept of the line can be tested to see if it is different to 0 (the origin).

This has much less statistical power than the IVW.

Weight median method	The above two methods pool effect estimates based on the mean effect and provide combined effect estimates which assume that either 0% SNPs (IVW) or up to 100% of them (MR Egger) violate core IV assumptions. The weight median method yields the median estimate of all the SNPs in the instrument.	The assumption here is that up to 50% of SNPs in the combined instrument may be invalid, but no more than 50%.
-----------------------------	--	--

Again, this method has less statistical power than the IVW.

Table 3

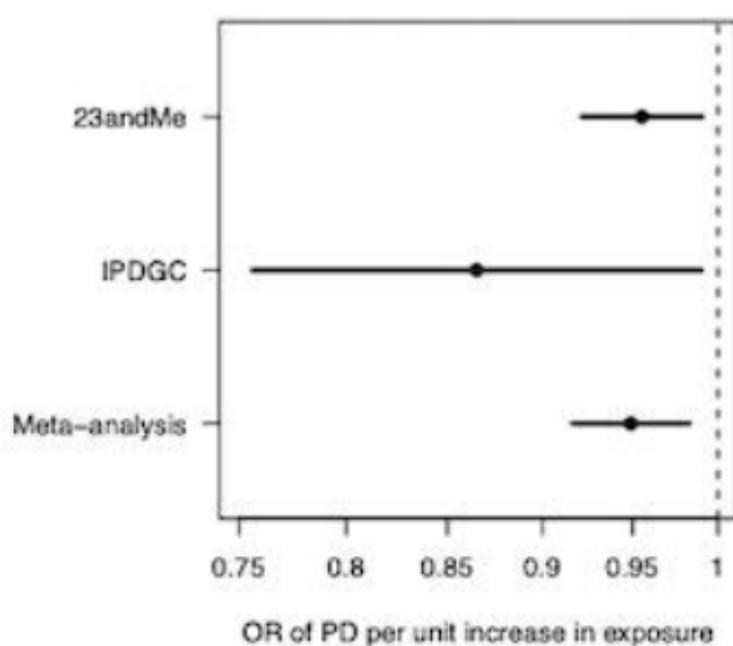
Table legend: Exposure and outcome sample sizes in the present study (23andMe discovery cohort and IPDGC replication cohort), and the two other Mendelian randomisation studies exploring the association between smoking and PD.

Abbreviations: PD = Parkinson’s disease; IPDGC = International Parkinson’s disease Genomics Consortium.

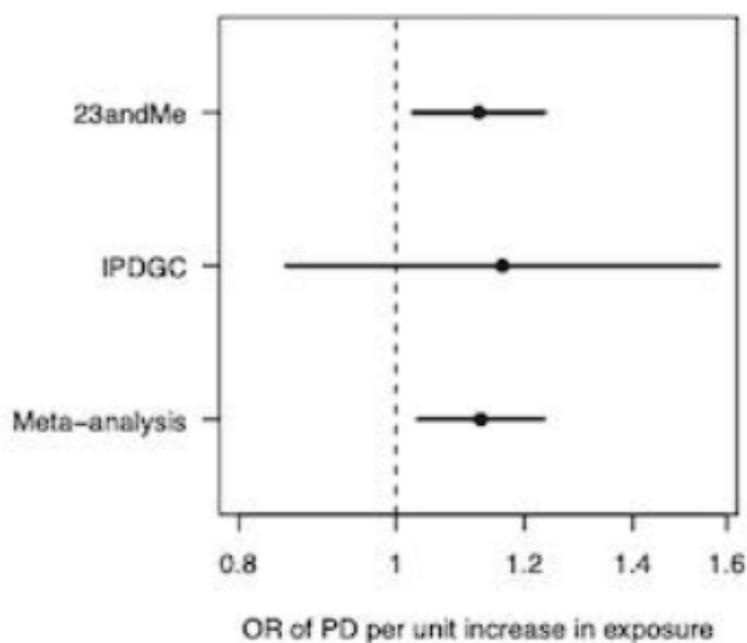
	Smokers	Non-smokers	PD cases	PD proxy cases	Controls
Grover <i>et al.</i>	246,715	271,918	9,581	..	33,245
Nalls <i>et al.</i>	208,988	244,705	37,688	18,618	1,417,791
23andMe discovery cohort	1,067,699	1,457,457	19,924	..	2,413,087
IPDGC replication cohort	15,074	18,618	449,056

Figure 1

Tobacco use



Alcohol consumption



Body mass index

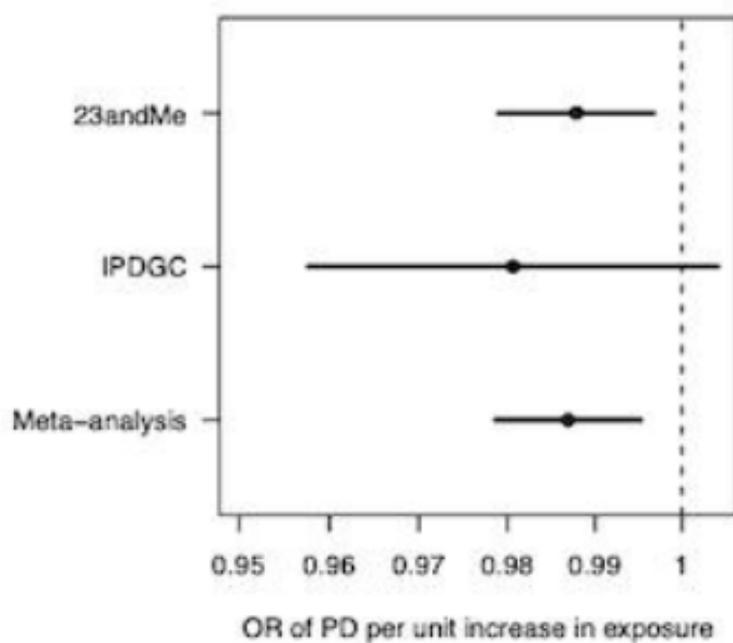
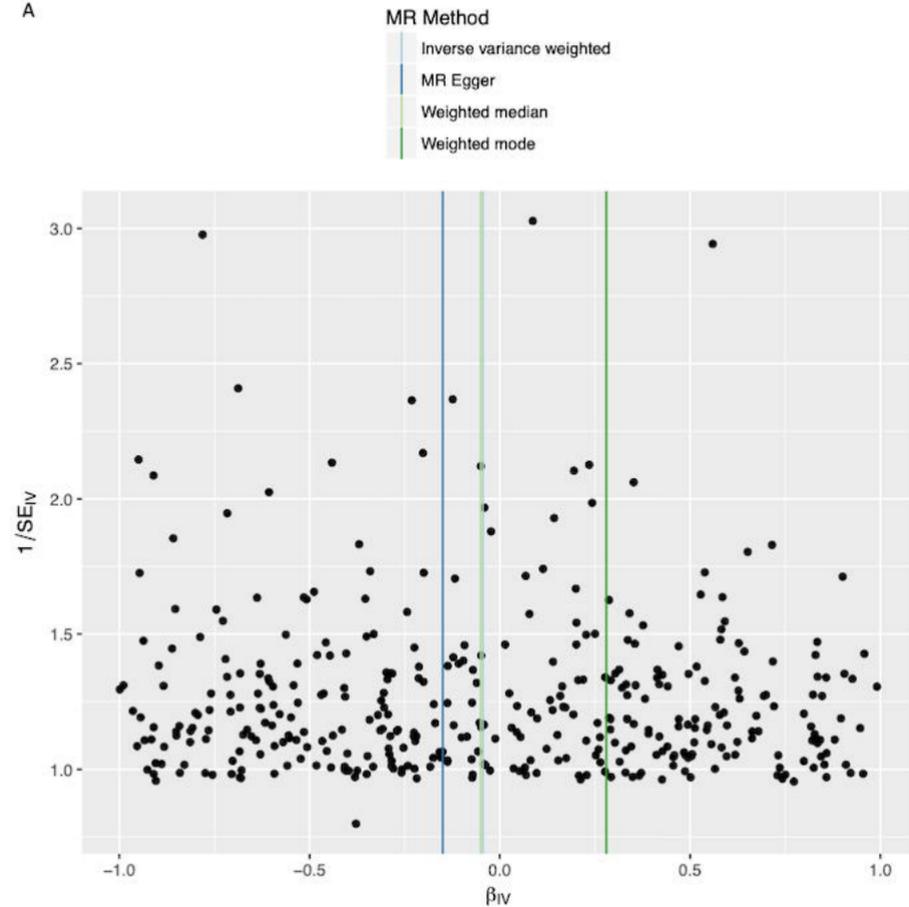


Figure 2

A



B

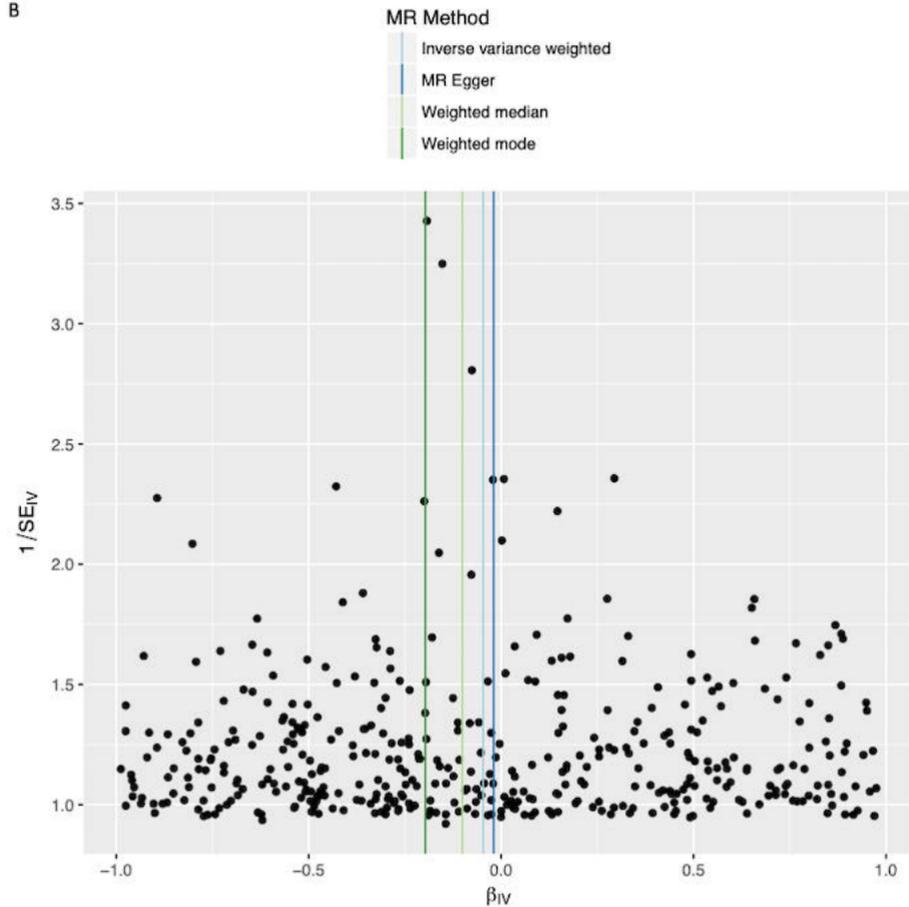
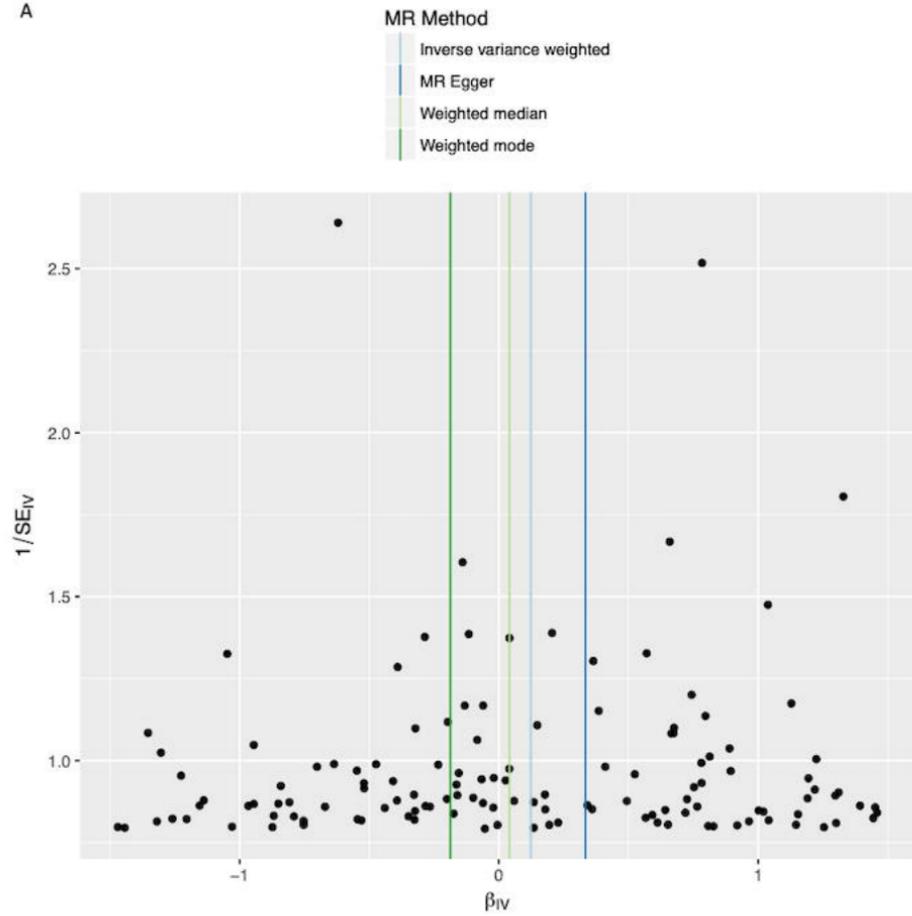


Figure 3

A



B

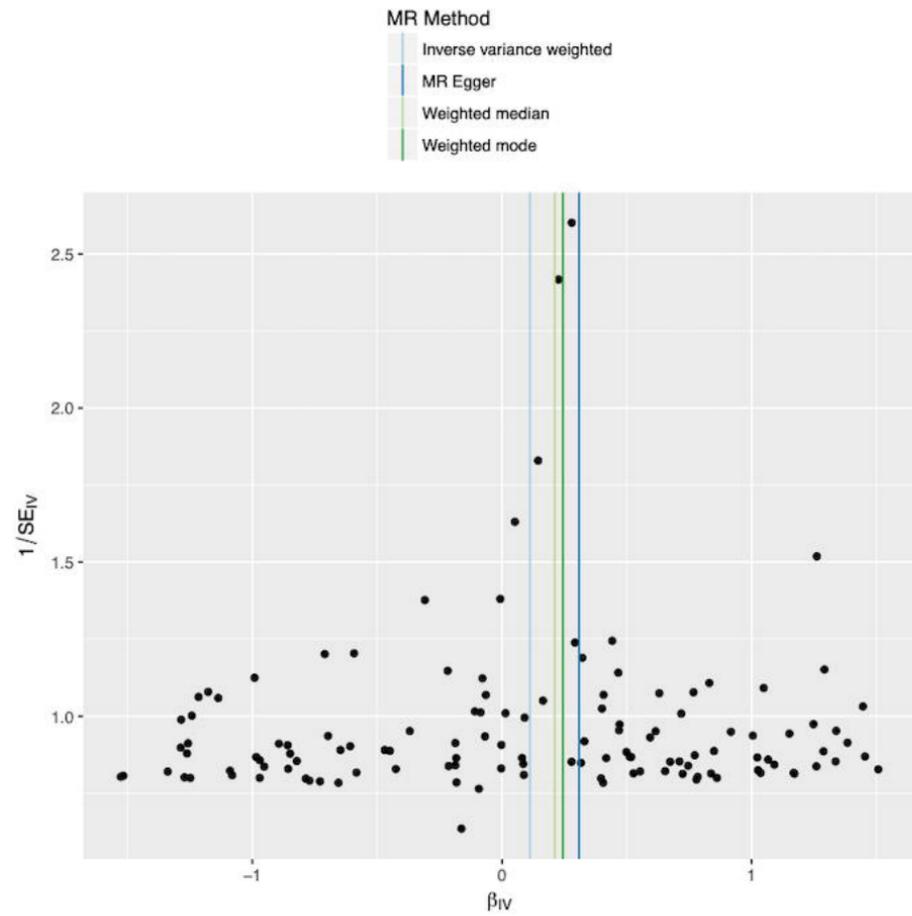
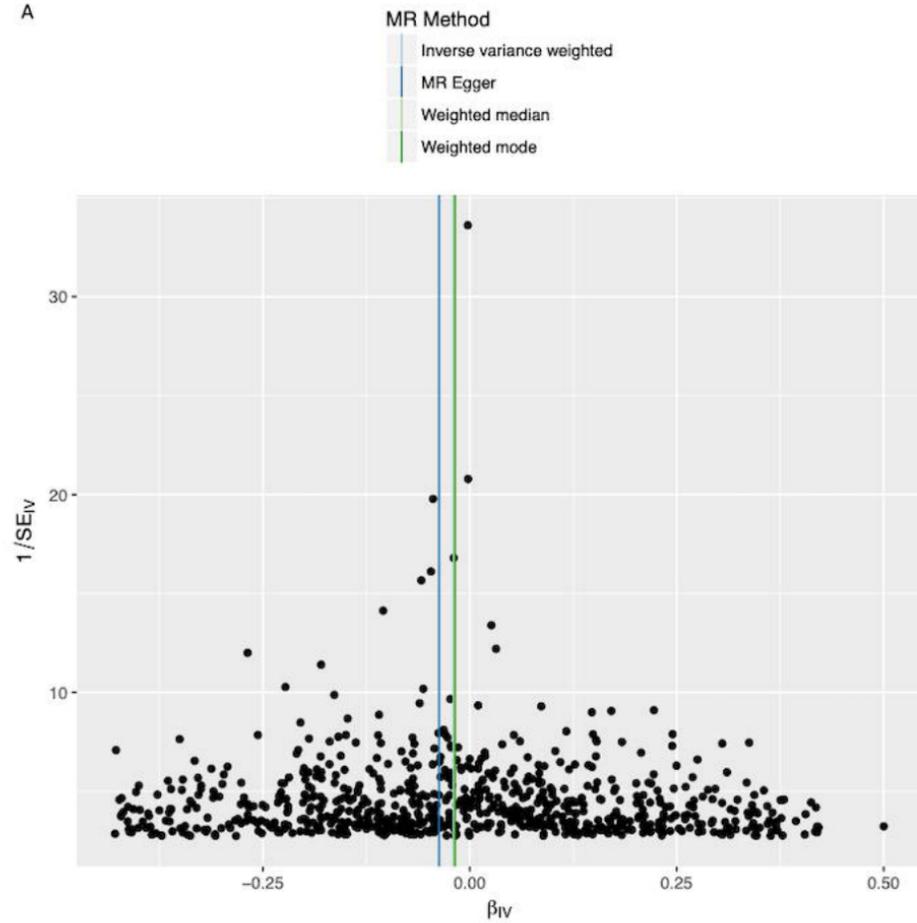


Figure 4

A



B

