



OPEN

## Understanding the effect of smoking and drinking behavior on Parkinson's disease risk: a Mendelian randomization study

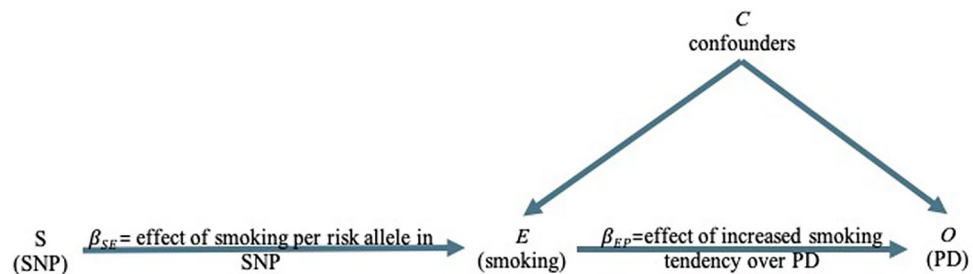
Carmen Domínguez-Baleón<sup>1</sup>, Jue-Sheng Ong<sup>2</sup>, Clemens R. Scherzer<sup>1,3,4</sup>, Miguel E. Rentería<sup>1,2,6</sup> & Xianjun Dong<sup>1,4,5,6</sup>

Previous observational studies have identified correlations between Parkinson's disease (PD) risk and lifestyle factors. However, whether or not those associations are causal remains unclear. To infer causality between PD risk and smoking or alcohol intake, we conducted a two-sample Mendelian randomization study using genome-wide association study summary statistics from the GWAS & Sequencing Consortium of Alcohol and Nicotine use study (1.2 million participants) and the latest meta-analysis from the International Parkinson's Disease Genomics Consortium (37,688 PD cases and 18,618 proxy-cases). We performed sensitivity analyses, including testing for pleiotropy with MR-Egger and MR-PRESSO, and multivariable MR modeling to account for the genetic effects of competing substance use traits on PD risk. Our results revealed causal associations of *alcohol intake* (OR 0.79; 95% CI 0.65–0.96;  $p = 0.021$ ) and *smoking continuation* (which compares *current vs. former smokers*) (OR 0.64; 95% CI 0.46–0.89;  $p = 0.008$ ) with lower PD risk. Multivariable MR analyses showed that the causal association between *drinks per week* and PD is unlikely due to confounding by smoking behavior. Finally, frailty analyses suggested that the causal effects of both *alcohol intake* and *smoking continuation* on PD risk estimated from MR analysis are not explained by the presence of survival bias alone. Our findings support the role of smoking as a protective factor against PD, but only when comparing *current vs. former smokers*. Similarly, increased alcohol intake had a protective effect over PD risk, with the *alcohol dehydrogenase 1B (ADH1B)* locus as a potential candidate for further investigation of the mechanisms underlying this association.

Parkinson's disease (PD) is a complex genetic disease, and growing evidence exists that lifestyle and environmental factors<sup>1</sup> and numerous genetic variants modulate both its susceptibility<sup>2,3</sup> and progression<sup>4,5</sup>. Environmental factors such as exposure to pesticides<sup>6,7</sup> or history of melanoma<sup>8</sup> have been associated with increased PD risk, whereas alcohol intake<sup>9–11</sup>, smoking<sup>12–14</sup>, caffeine consumption<sup>15,16</sup>, and educational attainment<sup>17</sup> are associated with reduced risk<sup>1,18</sup>. However, these relationships could be the product of confounding variables or reverse causation<sup>19,20</sup>. The traditional approach to investigating causality is the *randomized controlled trial* (RCT), where participants are randomly allocated to either a control or a treatment group to reduce bias<sup>21</sup>. However, running an RCT may not be possible in some situations due to the ethical and health concerns associated with subjecting participants to exposures that might be detrimental to their health, such as smoking, drinking, or pesticides<sup>22–24</sup>.

Mendelian randomization (MR) is an exciting alternative methodology that overcomes these difficulties using genetic variants (e.g., single nucleotide polymorphisms, or SNPs) as instruments to infer causality between exposure variables and health outcomes in a way that is independent of confounding factors. The method takes advantage of the fact that genetic variants are randomly distributed in the population just as participants would randomly be allocated in an RCT<sup>20,22,24,25</sup>. A *two-sample MR* experimental design requires summary statistics

<sup>1</sup>Center for Advanced Parkinson Research and Precision Neurology Program, Brigham & Women's Hospital, Harvard Medical School, Boston, MA, USA. <sup>2</sup>Department of Genetics & Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia. <sup>3</sup>Department of Neurology, Massachusetts General Hospital, Boston, MA, USA. <sup>4</sup>Department of Neurology, Brigham and Women's Hospital, Boston, MA, USA. <sup>5</sup>Genomics and Bioinformatics Hub, Brigham and Women's Hospital, Boston, MA, USA. <sup>6</sup>These authors jointly supervised this work: Miguel E. Rentería and Xianjun Dong. ✉email: miguel.renteria@qimrberghofer.edu.au; xdong@rics.bwh.harvard.edu



**Figure 1.** Directed acyclic graph of the effect of smoking on PD using genetic variants (SNPs) as proxies. (Adapted from Lawlor et al.<sup>22</sup>). An SNP (S) is associated with the exposure, for example, smoking (E), and therefore the SNP (S) can be used as an instrument to determine the effect of smoking (E) on PD, which is the outcome of interest (O). An association between the smoking-risk-increasing SNP (S) and elevated risk of PD is consistent with a causal relationship between E and O if three assumptions are met: (1) The SNP (S) must be strongly associated with smoking (E); (2) The SNP (S) must be independent of any confounding factors (C) that may be influencing PD (O) and (3) The SNP (S)'s association with PD (O) is explained only via smoking (E). SNP, single nucleotide polymorphism.

from two independent genome-wide association studies (GWAS) for the exposure and outcome variables. It examines whether the SNPs (S) that influence the exposure (E) also affect the outcome (O), but not the other way around. This way, MR facilitates the inference of a direct causal relationship between the exposure and the outcome (see Fig. 1)<sup>22,23</sup>.

The inverse association of PD risk with smoking and alcohol intake has been consistently reported in the literature<sup>9–14,18</sup>. A recent meta-analysis across 61 case–control and eight cohort studies by Li et al.<sup>13</sup> estimated pooled relative risk (RR) of a 0.59 (95% CI 0.56–0.62) for PD in smokers. Furthermore, another meta-analysis of 32 studies by Zhang et al.<sup>10</sup> estimated a RR of 0.78 (95% CI 0.67–0.92) when comparing participants with the highest versus the lowest levels of alcohol intake while also controlling for the effects of smoking and caffeine intake. Notably, between-study heterogeneity has been noted, which implies that studies in much larger and homogeneous samples might still be needed to confirm the causal nature of this association.

Mendelian randomization has recently been applied to infer causality between PD risk and environmental and lifestyle variables. Notably, Nalls et al.<sup>26</sup> reported little evidence for a causal relationship between current tobacco use and PD risk, despite finding a significant inverse genetic correlation between the two traits, estimated using LD score regression. Also, Noyce et al.<sup>27</sup> conducted a comprehensive investigation of 401 exposure traits, including smoking and drinking, and reported negative causal associations. However, a more detailed examination of the possible horizontal pleiotropy (i.e., a situation where candidate SNP instruments might be associated with the outcome through pathways independent of the exposure of interest) is required to disentangle the interplay between smoking behavior, alcohol intake, and PD risk.

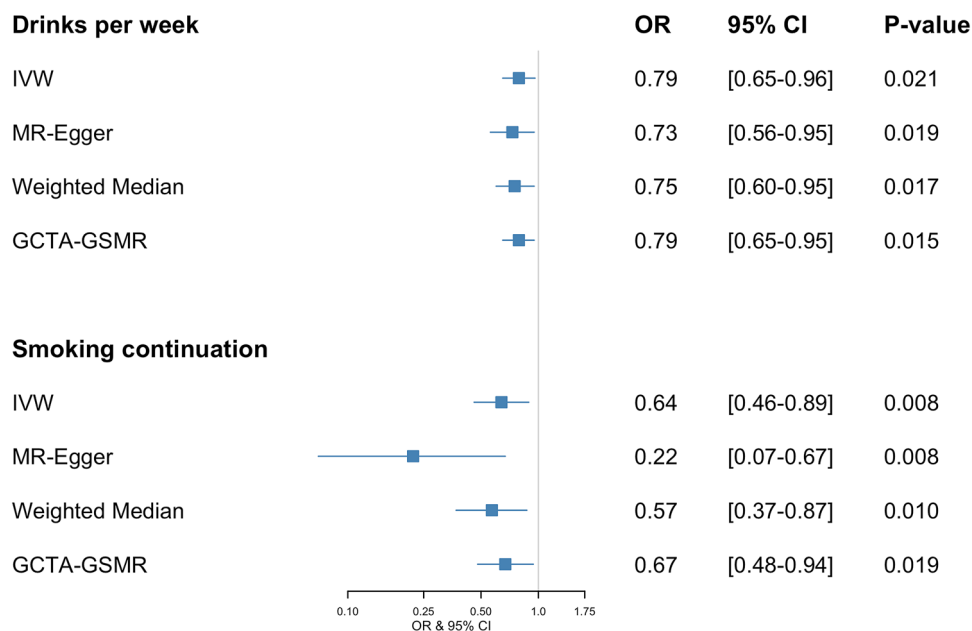
Here we leverage the availability of updated GWAS summary statistics from two recent studies by the GWAS & Sequencing Consortium of Alcohol and Nicotine use (GSCAN)<sup>28</sup> and the International Parkinson's Disease Genomics Consortium<sup>26</sup> (IPDGC) to expand on previous MR investigations. We specifically examine the causal effects of five lifestyle variables (number of alcoholic drinks per week, smoking initiation, heaviness, continuation, and age at initiation) on PD risk. Notably, given that smoking and alcohol intake are both phenotypically and genetically correlated, we performed multivariable MR analysis to control for horizontal pleiotropy on competing substance use factors, thus providing a better adjusted and more robust inference model. Lastly, we performed frailty analyses to test if survival bias was responsible for the MR estimates obtained in this study.

## Results

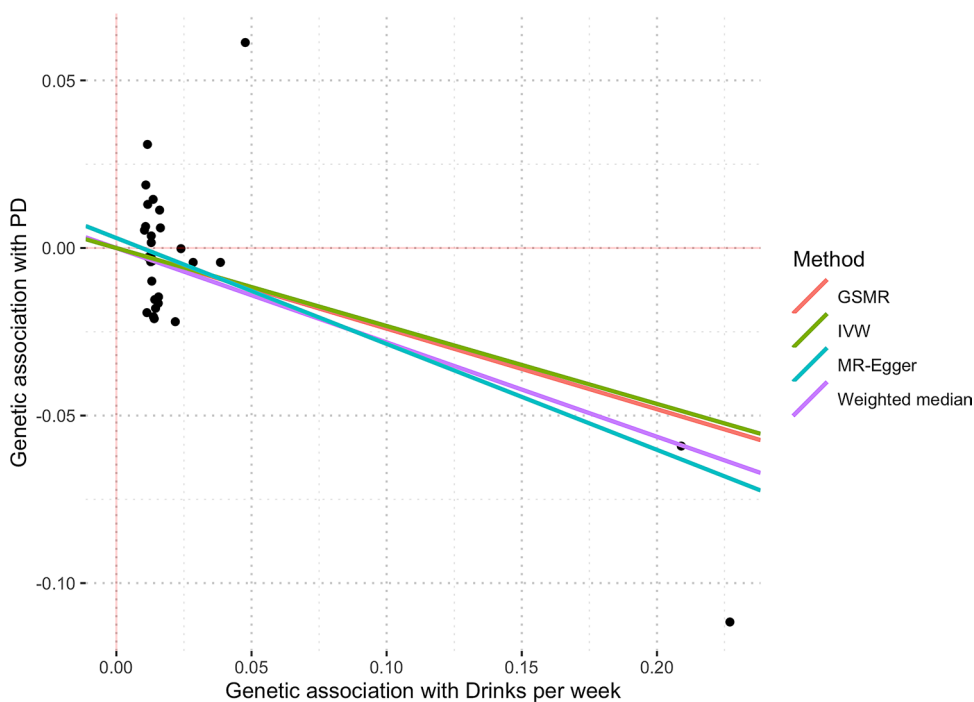
**Overview.** We found strong evidence supporting an inverse (i.e., protective) effect for the *number of drinks per week* and *smoking continuation* (which compares *current vs. former smokers*) traits on PD risk. Nonetheless, there was no clear association between PD risk and *smoking initiation* (*ever vs. never being a regular smoker*), *smoking heaviness* (*number of cigarettes per day*), or the *age at smoking initiation*. The results are summarized in Fig. 2 and Supplementary Fig. S3. Variant data used in this study can be found in Supplementary Tables S1–S5.

**Alcohol intake.** Horizontal pleiotropy analysis with MR-PRESSO identified two potentially pleiotropic outliers (rs1260326 and rs34121753), which have previously been associated with other traits, such as triglyceride<sup>29</sup> and serum acid levels<sup>30</sup> for the former and hematocrit levels for the latter<sup>31</sup>. After excluding outliers, our MR estimates showed that the genetic tendency to have more *drinks per week* is suggested to be causally associated with a reduced PD risk. The OR from the IVW estimate was 0.79 (95% CI 0.65–0.96;  $p = 0.021$ ) (Fig. 2), with no evidence for horizontal pleiotropy (MR-Egger estimate of 0.73 (95% CI 0.56–0.95;  $p = 0.019$ ) intercept of 0.003;  $p = 0.348$ ). Estimates derived from GCTA-GSMR yielded very similar findings: OR was 0.79 (95% CI 0.65–0.95;  $p = 0.015$ ). There was moderate evidence of heterogeneity of effect estimates between variants (Q statistic = 37.43,  $I^2 = 19.85\%$ ,  $p = 0.16$ ). Figure 3 shows all the derived MR estimates from alternative models for the effect of *drinks per week* on PD risk. We did not find evidence for unbalanced pleiotropy via manual inspection in the MR funnel plots (see Supplementary Fig. S1). Notably, the rs1229984 variant in the *alcohol dehydrogenase 1B* (*ADH1B*)

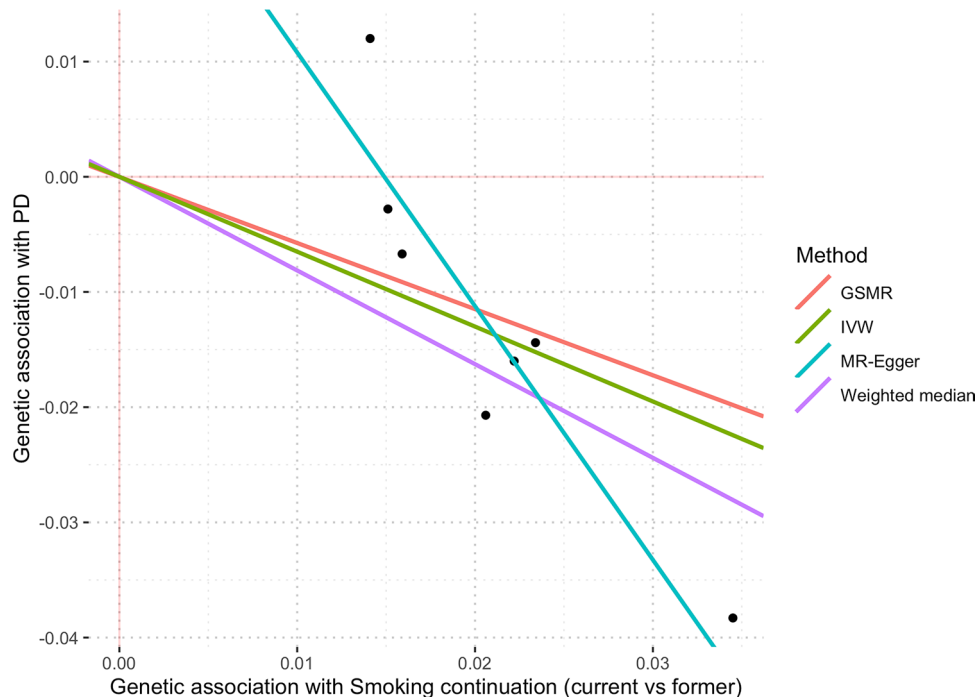
### Mendelian Randomization results



**Figure 2.** Forest plot showing OR and 95% confidence intervals of MR effect estimates. OR values for *smoking continuation* are expressed in *per doubling of odds* units.



**Figure 3.** Scatter plot showing MR effect estimates of *drinks per week* over PD. Each SNP-PD association is plotted against SNP-*drinks per week* association, and the corresponding MR estimates for IVW, MR-Egger, Weighted-Median, and GSMR are plotted. The two rightmost variants shown on the plot are rs29001570 and rs1229984, which correspond to the Alcohol Dehydrogenase 5 and Alcohol Dehydrogenase 1B loci.



**Figure 4.** Scatter plot showing MR effect estimates of *smoking continuation* over PD. Each SNP-PD association is plotted against SNP-*smoking continuation* association, and the corresponding MR estimates for IVW, MR-Egger, Weighted-Median and GSMR are plotted.

locus had the greatest association with *drinks per week*. It is known that individuals with one or two A alleles (i.e., AG or AA genotypes) at rs1229984 are more likely to find drinking unpleasant and thus have a reduced risk for alcoholism<sup>32,33</sup>. After removing rs1229984 and re-running all analyses, the resulting estimate was only slightly attenuated towards the null, and the point estimate remained in the same direction [0.84 (95% CI 0.63–1.11;  $p=0.216$ )], albeit with lower precision. On the other hand, the effect estimate of the rs1229984 single variant on PD risk was OR 0.75 (95% CI 0.59–0.96;  $p=0.025$ ).

There was very weak evidence for an association through reverse causality as evaluated in our bidirectional MR analysis. Greater genetic predisposition towards PD did not influence *drinks per week*, with an IVW estimate of 1.006 (95% CI 0.999–1.013;  $p=0.083$ ) and the equivalent GSMR estimate of 1.037 (95% CI 1.009–1.066;  $p=0.009$ ) per doubling of odds on PD. These findings infer that the association between alcohol drinks/week and PD is likely explained by unidirectional causality where genetically increased alcohol consumption reduces PD risk.

Finally, multivariable MR analysis accounting for the genetic effect of alcohol SNPs on smoking phenotypes found no attenuation of the effect of alcohol on PD risk [i.e., conditional association OR 0.72 (95% CI 0.53–0.98;  $p=0.034$ )], per standard deviation increase in the number of drinks per week when the genetic effect of alcohol SNPs on smoking behavior, obesity and education attainment have been taken into consideration (see “Methods” section).

**Smoking continuation.** MR-PRESSO identified a single SNP outlier (rs9607805, known to be associated with neuroticism<sup>34</sup>), which was excluded from the analysis. IVW results showed that current smokers have a lower risk of PD in contrast to former smokers. Based on the derived OR on PD per doubling of odds for *smoking continuation* was 0.64 (95% CI 0.46–0.89;  $p=0.008$ ) (see Fig. 2). There was weak evidence for unbalanced horizontal pleiotropy biasing the result (MR-Egger estimate was 0.22 (95% CI 0.07–0.67;  $p=0.008$ ) per doubling of odds with an intercept (in log(OR)) of 0.033,  $p=0.051$ ). No evidence of heterogeneity among variants (Q statistics = 5.48,  $I^2=0\%$ ,  $p=0.48$ ) was observed. GSMR analysis results yielded very similar results to IVW, with OR = 0.67 (95% CI 0.48–0.94;  $p=0.019$ ) per doubling of odds. Figure 4 shows MR estimates for the effect of *smoking continuation* on PD. Given the existing discrepancies in results from IVW/GSMR and MR-Egger, it is impossible to rule out the possibility that pleiotropic effects may influence the association between *being a current smoker* and PD. Manual inspection of the funnel plot showed asymmetry around the point estimate, which might indicate the presence of horizontal pleiotropy (see Supplementary Fig. S2).

Bidirectionality analysis showed that the effect of predisposition to Parkinson’s disease over *smoking continuation* is very close to the null, with a non-significant  $p$ -value and an IVW estimate of 0.997 (95% CI 0.992–1.001;  $p=0.14$ ) per doubling of odds. GSMR analysis yielded similar results: 0.96 (95% CI 0.93–0.992;  $p=0.015$ ), per doubling of odds of PD on *being a current smoker*. Notably, the multivariable MR OR estimate for *smoking continuation* controlling for the genetic effect on alcohol consumption and other smoking behavior phenotypes

could not be meaningfully interpreted due to the low number of SNP instruments, which yields an extensive confidence interval OR 0.66 (95% CI 0.04–9.96;  $p=0.76$ ).

**Smoking initiation, heaviness, and age at initiation.** The predisposition to *ever being a regular smoker* (as opposed to never initiating smoking) OR per doubling of odds for IVW was 0.90 (95% CI 0.74–1.09;  $p=0.308$ ), GSMR 0.89 (95% CI 0.78–0.97,  $p=0.112$ ) and MR-Egger 0.39 (95% CI 0.15–1.07;  $p=0.068$ ) per doubling of odds with moderate evidence of heterogeneity between variants (Q statistic = 152.78,  $I^2=44.36\%$ ,  $p=9.18e-6$ ). *Smoking heaviness*, characterized by the number of cigarettes smoked per day, showed a IVW OR of 1.06 (95% CI 0.88–1.29;  $p=0.526$ ), a GSMR estimate of 1.07 (95% CI 0.92–1.23,  $p=0.373$ ) and MR-Egger estimate was OR = 1.07 (95% CI 0.75–1.51,  $p=0.72$ ) with moderate evidence of heterogeneity (Q statistic = 36.78,  $I^2=42.9\%$ ,  $p=0.017$ ). Lastly, *age at smoking initiation* IVW estimate was 0.83 (95% CI 0.53–1.30;  $p=0.416$ ), GSMR OR = 0.84 (95% CI 0.52–1.38,  $p=0.492$ ) and a MR-Egger estimate of 1.002 (95% CI 0.16–6.29,  $p=0.998$ ), showing moderate evidence of heterogeneity (Q statistic = 5.43  $I^2=8$ ,  $p=0.36$ ) (see Supplementary Fig. S3). Derived MR estimates and corresponding funnel plots for these traits can be found in Supplementary Fig. S5–S10.

**Frailty analysis.** Neurodegenerative disorders that are strongly associated with age, such as PD, may be subject to bias if selective mortality occurs. For example, if individuals who consume alcohol or continue smoking die prematurely before having a PD diagnosis, then those individuals who don't consume alcohol or quit smoking might be more represented among PD patients. Thus, survival bias might explain the association between either alcohol consumption or being a current smoker and PD risk. In order to test if survival bias was sufficient to explain the MR results estimated from this study, we followed the methodology used by Noyce<sup>35</sup>, which consisted of a series of simulations that were performed to obtain the effect sizes that are expected from the frailty effects alone. If the effect sizes obtained from the frailty analysis were to be similar to the effect sizes obtained from the Mendelian Randomization analysis, then it would be suggested that the survival bias is large enough to explain the results. In the case of *drinks per week*, the simulations showed that there is a negligible influence of frailty effects biasing the MR results as the effect induced solely by frailty in the IVW analysis was an OR of 0.994 (95% CI 0.909–1.087) compared to the MR result OR of 0.78 (95% CI 0.69–0.88) (Fig. 5). On the other hand, for *smoking continuation* frailty effects do not exert a bias over the MR results as the effect of frailty alone in the IVW approach was 1.02 OR (95% CI 0.77–1.29) in contrast to the result from the MR analysis, which had a 0.64 OR (95% CI 0.46–0.89) (Fig. 6). Finally, we performed a one-sample test to determine if the difference between the IVW frailty analysis results and the estimated IVW MR values was significant. A  $p$ -value  $<2.2e-16$  was obtained for both *drinks per week* and *smoking continuation* traits, indicating that the difference between the frailty analysis results and our MR analyses are significantly different.

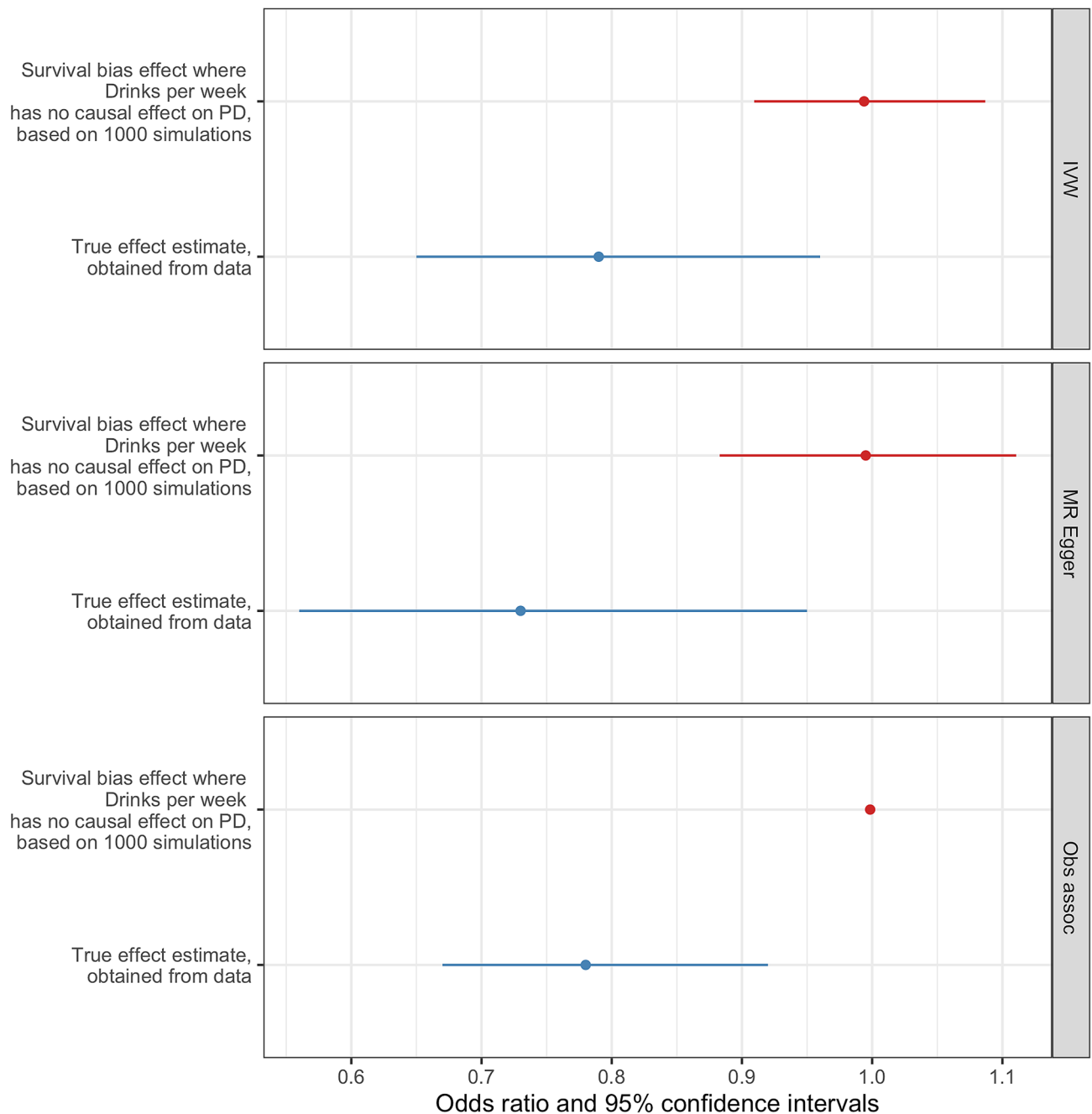
Both the exposure and PD datasets include participant data from the UK Biobank (UKBB) and 23andMe cohorts. To test whether the two-sample MR findings might be biased due to the sample overlapping<sup>36</sup>, we derived genetic associations with the exposures of interest using a non-overlapping data subset<sup>37</sup>. We used a publicly available dataset by Liu et al.<sup>28</sup>, which excludes data from both UKBB and 23andMe. SNP-exposure estimates (for instruments used in our main analysis) from the GWAS conducted on these subsets were used to re-evaluate our MR associations. For *drinks per week*, the OR from the IVW estimate was 0.82 (95% CI 0.62–1.08;  $p=0.155$ ), and no evidence for horizontal pleiotropy (MR-Egger estimate of 0.67 (95% CI 0.44–1.02;  $p=0.064$ ) and an intercept value of 0.005 ( $p=0.224$ )), and a GCTA-GSMR estimate of 0.65 (95% CI 0.48–0.88;  $p=0.007$ ) (see Fig. 7). For *smoking continuation*, the IVW estimate was 0.83 (95% CI 0.73–0.95;  $p=0.006$ ), with a MR-Egger estimate of 0.61 (95% CI 0.43–0.89;  $p=0.009$ ) with an intercept of 0.024 ( $p=0.084$ ), indicating no presence of horizontal pleiotropy, and the GCTA-GSMR estimate of 0.74 (95% CI 0.54–1.02;  $p=0.067$ ) (see Fig. 7). *Smoking initiation*, *smoking heaviness*, and *age at smoking initiation* showed similar associations as those reported in the main analysis but with wider confidence intervals. The complete findings from the analysis using the dataset excluding the 23andMe and UKBB samples can be found in Supplementary Fig. S4.

## Discussion

In the present study, we provide genetic evidence to support a protective role of moderate alcohol intake against PD risk (OR 0.79, 95% CI 0.65–0.96;  $p=0.021$ ). Moreover, we found that being a *current smoker* vs. a *former smoker* was causally associated with reduced PD risk (OR 0.64 95% CI 0.46–0.89;  $p=0.008$ ), whereas no causal association could be established for *smoking initiation*, *smoking heaviness*, and the *age at smoking initiation*.

Interestingly, in the case of alcohol intake, both the effect estimate and direction of causal association were consistent (OR 0.84, 95% CI 0.63–1.11;  $p=0.216$ ) even after removing the rs1229984 variant, which had the strongest association with *drinks per week* and corresponded to the *alcohol dehydrogenase 1B* locus, in which the reduction in precision that is shown by wider confidence intervals and a non-significant  $p$ -value is expected as the variant rs1229984 has the strongest association with the drinks per week phenotype, but the value of the causal estimate (0.84) remained in the same direction and is in fact very close to the original estimate (0.79) and therefore we can conclude that the effect that alcohol drinking has over PD risk is not solely driven by the variant rs1229984 and thus horizontal pleiotropy is not underlying the association. Nonetheless, the individual estimate for the effect of the rs1229984 variant over PD remained similar to the original result but with lower precision, indicated by broader confidence intervals (OR 0.75 95% CI 0.59–0.96;  $p=0.025$ ), which positions the rs1229984 SNP as a suitable candidate for investigating the mechanism behind the protective effects of alcohol intake on PD risk. The ADH1B protein metabolizes ethanol into acetaldehyde during alcohol metabolism, and the risk allele rs1229984 (A) produces a substitution of Arginine for a Histidine that changes affinity for the NAD<sup>+</sup> coenzyme and increases ethanol oxidation by a 70–80-fold<sup>38</sup>. Individuals that carry the rs1229984 (A)

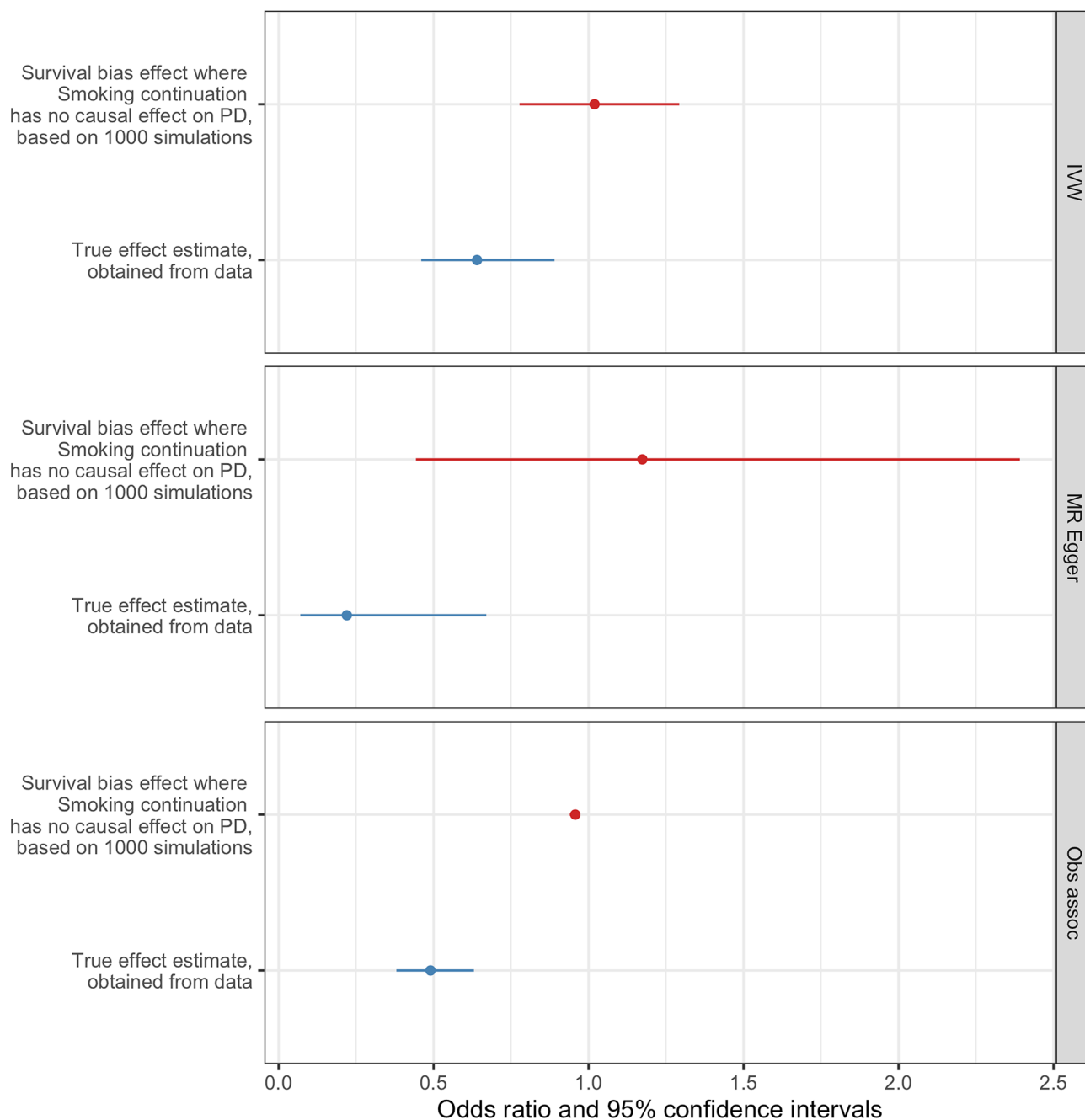




**Figure 5.** Frailty analysis results for *drinks per week*. The effect estimates obtained from simulations in which the effect *drinks per week* has over PD risk are solely explained by the presence of survival bias (red) are compared to the causal effect estimates between *drinks per week* and PD risk obtained from our MR analysis (blue), for the IVW, MR-Egger and observational approaches.

allele, therefore, present a condition known as “alcohol flush reaction”, characterized by a visible flush on their face and body, and in some cases, symptoms such as nausea, headache, and general discomfort. Hence, carriers of the A allele tend to limit their alcohol intake. In line with our findings, a recent study by García-Martín et al.<sup>39</sup> reported an inverse association between the reduced alcohol intake phenotype of the rs1229984 (A) allele and increased PD risk. Considering that this variant is both implicated in alcohol metabolism and strongly associated with alcohol intake, it is a promising candidate for investigating the mechanism by which alcohol drinking influences PD risk. Another strongly associated SNP with *drinks per week* was rs29001570 (see Supplementary Fig. S1), located in the *Alcohol Dehydrogenase 5 locus (ADH5)*, which has also been previously associated with alcohol intake<sup>40</sup>. *ADH5* is known to be involved in formaldehyde metabolism but has a low affinity for alcohol, so the mechanism by which it associates with alcohol intake is not well understood.

The average level for *drinks per week* phenotype across Liu et al.<sup>28</sup> meta-analysis is 7.6 drinks per week, which would broadly refer to a moderate drinking pattern according to the Dietary Guidelines for Americans

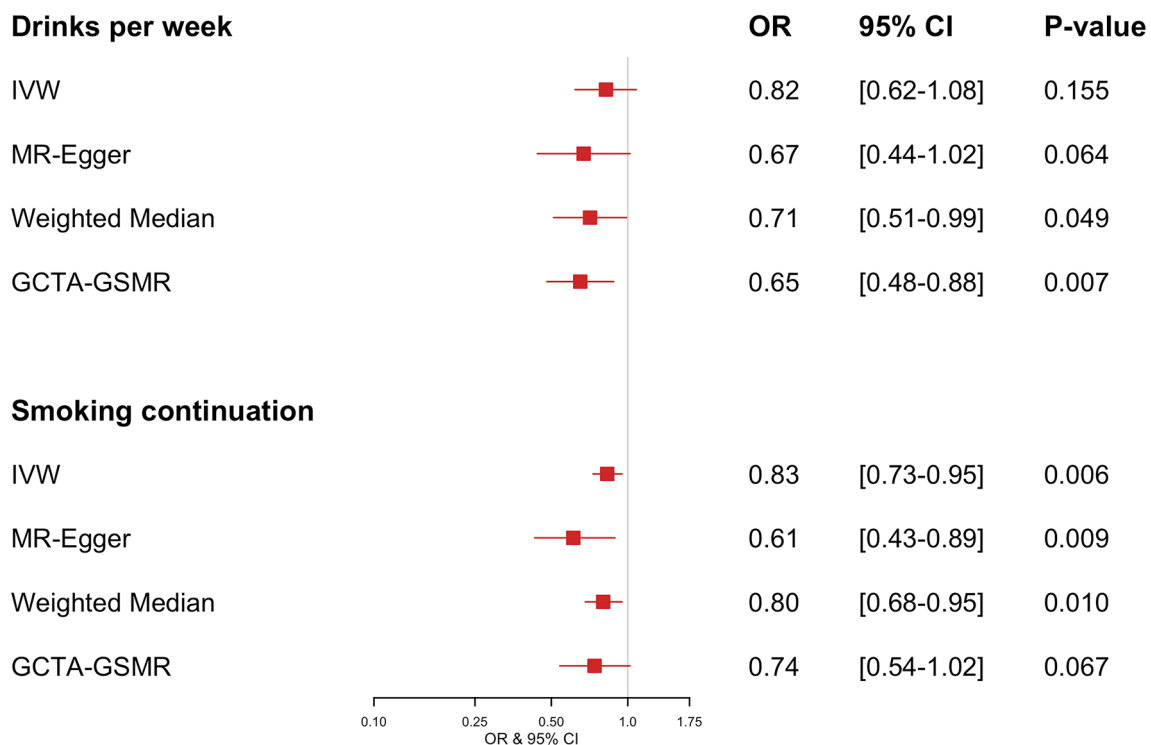


**Figure 6.** Frailty analysis results for *smoking continuation*. Comparison between the effect estimates obtained from simulations in which the effect *smoking continuation* has over PD risk are explained by the presence of survival bias (red) and the resulting effect estimates between *smoking continuation* and PD risk obtained from our MR analysis (blue), estimated for the IVW, MR-Egger and observational approaches.

2015–2020<sup>41</sup>, which defines moderate alcohol consumption as up to 1 drink per day (7 drinks per week). It has also been shown that moderate alcohol consumers are at reduced risk of all-cause mortality [HR 0.78, 95% CI 0.74 to 0.82]<sup>42</sup>. This indicates that there is a low chance that survival bias may be affecting our *drinks per week* results, an idea that is further supported by our frailty analysis results.

We found that variants associated with being a *current smoker* in the *smoking continuation* GWAS were suggested to be causally associated with reduced PD risk. Although the IVW result showed evidence for an association with OR 0.64 (95% CI 0.46–0.89;  $p = 0.008$ ) per doubling of odds on PD, the MR-Egger estimate suggested the presence of unbalanced horizontal pleiotropy, suggesting multiple potential (independent) biological modes of action linking the smoking variants onto PD risk, although the point estimate remain widely consistent with the IVW findings. Pleiotropic effects might be exerted in part by the lower genetic correlation between *drinks per week* and *smoking continuation* phenotypes reported by Liu et al.<sup>28</sup> ( $r_g = 0.11$ ,  $p = 6.45 \times 10^{-5}$ ). However, it is unlikely that this correlation is solely responsible for the effect *smoking continuation* has on PD risk. Importantly,

## MR Results from Liu et.al. excluding UKBB &amp; 23andMe dataset



**Figure 7.** Forest plot showing OR and 95% confidence intervals of MR effect estimates calculated from the Liu et al. dataset that excluded UKBB and 23andMe. OR values for *smoking continuation* are expressed in *per doubling of odds* units.

Exposure	Total SNPs	Sample size	R <sup>2</sup>	F-statistic	Power (%)
Age at smoking initiation	7	341,427	0.00102	49.74375	33
Smoking initiation	87	1,232,091	0.00441	62.74564	89
Drinks per week	33	941,280	0.00513	147.00378	93
Smoking continuation	8	547,219	0.00105	72.13881	34
Smoking heaviness	25	337,334	0.00983	113.90956	100

**Table 1.** The number of variants, R<sup>2</sup>, F-statistic and statistical power for each of the five phenotypes under investigation. Power was calculated to detect a difference in PD risk with an OR of 0.8 in the IPDGC cases and controls with an alpha of 5%.

given that the discovery GWAS ignored time to cessation in the *former smokers* group, the phenotype can be conceivably more heterogeneous, making extrapolation of causality from these genetic-associations seemingly complicated. As the present study might lack sufficient power to characterize predictive instruments for *smoking continuation* (see Table 1), future studies leveraging even larger sample sizes for both *smoking continuation* and PD will be needed to validate these findings.

Notably, most smokers have attempted to quit smoking at least once in their lifetime<sup>43</sup>, and tobacco dependence is associated with longer smoking duration and failure of quitting attempts<sup>44</sup>. Therefore, *being a current smoker* could simply be interpreted as individuals being less likely to quit because they have a stronger genetic predisposition to smoking, whereas *former smokers* have a comparatively weaker predisposition and can quit successfully. Thus, our results support the hypothesis of smoking being a protective factor against PD.

It is important to note that the exclusion of never smokers in the *smoking continuation* phenotype (current vs. former smokers) influences the results' interpretation. Our *smoking continuation* MR result suggests that prolonged smoking confers the protective effect against PD risk compared to shorter-term smoking. In contrast, the *smoking initiation* (ever vs. never smoker) result did not reach statistical significance. This might be because the "ever smoker" group contains both "current" and "former" smokers. The "former" smokers might be masking



the protective effect seen in “current” smokers, therefore yielding a non-significant effect when the “ever smoker” group is compared against the “never smoker” group. Unfortunately, the Liu et al.<sup>28</sup> study did not compare ‘never and former smokers’ vs. ‘current smokers’. We hypothesize that, under that scenario, smoking’s protective effect would be similar to that seen for *smoking continuation*. Future studies could investigate this question.

Weak instrument bias can arise when there is overlap between the two datasets used in the two-sample MR approach. The degree of overlap and its effects cannot be calculated, but it can potentially bias the resulting estimate towards the observational association<sup>37</sup>. To overcome the potential *weak instrument bias* due to overlapping samples in the exposure and outcome datasets (i.e., both include participants from the UKBB and 23andMe), we repeated the MR analysis using a Liu et al. data subset that excluded both UKBB and 23andMe. The results for both *drinks per week* and *smoking continuation* remained significant for most of the methods evaluated. For *drinks per week*, the OR from the IVW estimate was 0.82 (95% CI 0.62–1.08;  $p=0.155$ ), and for *smoking continuation*, IVW estimate was 0.83 (95% CI 0.73–0.95;  $p=0.006$ ). As we can conclude by comparing Figs. 2 and 7, the estimates across the four methods evaluated (IVW, MR-Egger, Weighted Median, GTCA-GSMR) for both *drinks per week* and *smoking continuation* were slightly shifted towards the null, and the 95% confidence intervals were widened. This observation was expected because removing the UKBB and 23andMe cohorts decreased the sample size from 1.2 million to ~250,000 participants. Such a reduction in sample size results in lower precision of the estimate shown by the widening of confidence intervals and the reduction of  $p$ -value significance. However, despite the considerable reduction in sample size, the resultant point estimates were similar to those reported in our main analysis. These findings provide additional assurance that our initial MR results were unlikely driven to be by inadequate sample overlap management.

Interpreting univariate MR results can be challenging because SNP instruments associated with the exposure of interest can exert pleiotropic effects on correlated risk factors. However, discarding SNPs based on an observed association with correlated traits can greatly diminish statistical power to detect an association. By conditioning out the association of each instrument on pleiotropic risk factors alongside the exposure of interest in our adjusted model, we evaluated the direct effect that our exposure of interest has on PD risk. While this approach is conceptually useful to tackle these issues, there are several potential concerns. Firstly, while results from the multivariable MR analysis conceptually help obtain a marginal estimate of the effect size between an exposure of interest and the outcome by conditioning competing risk factors, the model can potentially over-correct true associations as the underlying model does not distinguish between horizontal (independent of causality) and vertical (in line with causality) pleiotropy. For instance, the genetic associations with alcohol and smoking can be heavily confounded by (genetic effects on) socio-economic status<sup>28</sup>. In practice, the single SNP result for alcohol is likely a more reliable predictor for habitual alcohol intake given its well-understood role in regulating alcohol metabolism. However, unlike most variants, it is not associated with proxies of obesity, smoking, or educational attainment. Our results suggest that the protective relationship of alcohol intake (*drinks per week*) on PD remains consistent even after controlling for the (potential) pleiotropic effects of these SNPs on smoking behavior, obesity, and educational attainment. Thus, the causal association between alcohol intake and PD is unlikely to be the product of confounding via smoking behavior.

The frailty analysis we performed aimed to test if the results obtained from the MR analysis in this study could be explained by the effects of frailty alone, in which a survival bias due to selective mortality was responsible for the causal estimate between either alcohol consumption or *smoking continuation* and PD risk. We found that, in the case of *drinks per week*, survival bias introduced by the frailty effect is not sufficient enough to explain the magnitude of the association between *drinks per week* and PD risk shown by the MR results. Similarly, for smoking continuation, there was no evidence that the frailty effects were solely responsible for the estimate obtained from the MR analysis. Therefore, our frailty analysis results rule out that survival bias can explain the causal association between that either *alcohol intake* or *smoking continuation* and PD risk we found through the Mendelian randomization analysis.

A handful of recent studies have attempted to examine the causal relationship between smoking and alcohol intake on PD risk using Mendelian randomization. Nalls et al.<sup>26</sup> reported little evidence for a causal relationship between PD and current smoking status (OR 0.545; 95% CI 0.230–1.291;  $p$ -value = 0.1681). Noyce et al.<sup>27</sup> reported an inverse causal association between both current smoking status and increased alcohol intake and PD risk, but those associations were not explored in detail, and sensitivity analyses results were not explored or discussed in depth. Even though the Noyce et al. results and ours are similar, an important difference is the source of both the exposure (UKBB vs. Liu et al. 2019) and outcome (Chang et al. 2017 vs. Nalls et al. 2019) GWAS summary statistics datasets. The analyses presented here used datasets that comprised substantially larger and better-powered samples than those previously used. Also, we performed a series of sensitivity analyses to minimize the possibility of bias due to pleiotropy.

One limitation of the current study is that we could not perform subgroup analyses, for instance, to determine whether the protective effect of alcohol intake on PD risk is the same in both men and women. That is an important question as high-volume, and high-frequency drinking are more prevalent in men than women<sup>45</sup>. At the same time, we could not consider the survival effects of alcohol intake and smoking behavior. Heavy smoking increases mortality, so heavy smokers may be underrepresented among PD patients<sup>46</sup>. Finally, given that the PD GWAS consisted of a case–control design, the effects of smoking and drinking behavior can only be interpreted in the context of risk of developing PD, but are not informative of the effect that these behaviors might exert on disease progression of individuals with a current PD diagnosis.

Finally, it is crucial to be aware that smoking and drinking pose serious health risks to individuals. Evaluation and translation of these epidemiological clues for therapeutics have been difficult and thus far unsuccessful. However, a better understanding of the underlying biology and their relationship with PD risk could ultimately help delineate novel targets for prevention or treatment without the adverse health effects of smoking and drinking.

## Methods

**Data sources.** We obtained GWAS summary statistics from the GWAS & Sequencing Consortium of Alcohol and Nicotine use (GSCAN) consortium (Liu et al.<sup>28</sup> [www.conservancy.umn.edu/handle/11299/201564](http://www.conservancy.umn.edu/handle/11299/201564)), in which 556 genetic variants in 406 loci were associated with the following phenotypes: *smoking initiation* (ever being a regular smoker), *age at smoking initiation*, *smoking heaviness* (cigarettes per day), *smoking continuation* (current vs. former smoker), and the number of *drinks per week*, in up to 1.2 million participants (Phenotype definitions can be found in the Supplementary Note). The beta coefficients, standard errors, and p-values were extracted from all five datasets. To address the potential instrumental bias issue, we obtained from the GSCAN consortium an additional summary statistics dataset, which excluded both UKBB and 23anMe and consisted of ~250,000 participants. The sample sizes for each GWAS study are listed in Table 1.

We also obtained summary statistics from the most recent PD GWAS from the International Parkinson's Disease Genomics Consortium<sup>26</sup> (IPDGC; [www.pdgenetics.org/](http://www.pdgenetics.org/)) and 23andMe Inc., through direct request. This dataset consisted of 7.8 million analyzed genetic variants in 37,688 cases, 18,618 UKBB proxy cases (first-degree relatives of PD patients), and 1.4 million controls.

**Statistical analyses.** All datasets were “clumped” using PLINK (<http://zzz.bwh.harvard.edu/plink/clump.shtml>). Variants selected as index SNPs were those with the smallest p-values < 5e-8, while clumped SNPs were those in high linkage disequilibrium with the index SNP ( $R^2$  threshold of 0.001) or within a distance of 10,000 kb based on the 1000 Genomes project reference dataset. From the 406 variants identified by Liu et al.<sup>28</sup> (10 associated with the *age at smoking initiation*, 55 with *smoking heaviness*, 378 with *smoking initiation*, 24 with *smoking continuation*, and 99 with *drinks per week*), we identified 7 associated variants with the *age at smoking initiation*, 29 with *smoking heaviness*, 96 with *smoking initiation*, 8 with *smoking continuation* and 42 with *drinks per week*, after clumping.

To evaluate whether the SNP instruments selected for each exposure were strongly associated with the exposure, the proportion of phenotypic variance for each trait explained by the SNPs ( $R^2$ ) and  $F$  statistics were calculated for the five traits (see Table 1).  $R^2$  for each trait was calculated based on the formula

$$\sum_i \frac{2m_i(1 - m_i)\beta_i^2}{\text{Var}(Y)},$$

where  $m_i$  is the minor allele frequency and  $\beta_i$  the magnitude of the association between the  $i$ -th SNP on trait  $Y$ <sup>47</sup>. The  $F$  statistic was calculated as previously described by Noyce et al.<sup>35</sup>:

$$F = \frac{R^2(n - 1 - k)}{(1 - R^2) \times k},$$

where  $n$  is the sample size and  $k$  the number of SNPs.

The beta coefficients, standard errors, and p-values were extracted from the PD dataset for those index SNPs previously identified in each study. We then inspected how many index SNPs were shared between PD and each of the five traits. In the case of the *age at smoking initiation* and *smoking continuation*, all index SNPs were also found in the PD dataset ( $n=7$  and 8, respectively). Moreover, 25 SNPs were found for the *smoking heaviness*, 87 SNPs for *smoking initiation*, and 33 SNPs for *drinks per week*. No suitable LD-proxies were found in the PD dataset for any of the missing variants. For the bidirectional-MR, 88 index SNPs were identified from the PD GWAS dataset and further extracted from the *smoking continuation*, *smoking initiation*, and *drinks per week* GWAS datasets.

Power calculations were conducted using the method by Brion et al.<sup>48</sup> ([shiny.cns.genomics.com/mRnd/](http://shiny.cns.genomics.com/mRnd/)) and the  $R^2$  previously calculated, in which the power to detect an effect of OR 0.8 on PD risk in the iPDGC dataset with an alpha of 5% was evaluated for each trait (see Table 1). *Smoking heaviness* showed the greatest percentage of power with 100%, followed by *drinks per week* with 93% and *smoking initiation* with 89%. The traits with the lowest power values were *smoking continuation* and *age at smoking initiation* with 34% and 33%, respectively.

To evaluate potential instrumental bias, we extracted the beta coefficients, standard errors, and p-values from the excluding UKBB and 23andMe datasets for those index SNPs for each trait previously identified in the original Liu et al.<sup>28</sup> dataset. Finally, we repeated the MR analyses with this new smoking and drinking dataset and the same PD dataset as the previous analysis.

**Effect allele harmonization.** Harmonization across both the GSCAN and PD datasets was carried out as described by Hartwig et al.<sup>36</sup> before the MR analyses. First, all beta coefficients of the smoking phenotypes datasets were ensured to be positive (to facilitate interpretation of plots and as a requirement for further MR analysis). Negative beta coefficients were ‘flipped’, e.g., they were multiplied by  $-1$ , the effect allele became the alternative allele, and the effect allele frequency was subtracted from 1. Then, we compared the effect allele coding in the smoking dataset and the Parkinson's disease dataset and ensured that both had the same allele coding. Variants in the PD dataset with different effect alleles in the smoking dataset were ‘flipped’ as previously described.

**Mendelian randomization analyses.** Before the MR analyses, we first harmonized both the GSCAN and PD datasets to correctly align the effect alleles. Four different MR methods were then selected to conduct the analysis: Inverse Variance Weighted (IVW)<sup>49</sup>, MR-Egger<sup>50,51</sup>, MR-PRESSO<sup>52,53</sup> and GTCA-GSMR<sup>25,53</sup>, each with unique characteristics. A fundamental assumption required for MR inferences to be valid is that the instrument (SNP) and the outcome must be associated only through the exposure. If this assumption is violated, the presence of horizontal pleiotropy (an SNP associated with some other trait besides our exposure of interest) may bias

MR estimates, potentially leading to erroneous conclusions<sup>20,22,54</sup>. In the case of the IVW model, it calculates the causal estimate by using a weighted linear regression that combines each instrumental variable's wald ratios and weighting them by the inverse variance of the gene-outcome association while constraining the intercept to zero; therefore, it can only be interpreted under strong assumptions on the absence of pleiotropy<sup>49</sup>. The other three models exploit the statistical properties of MR estimates that are less vulnerable to these biases. For instance, MR-Egger provides a conservative estimate of the MR effect size in the presence of directional pleiotropy by fitting an intercept term to model the bias. If the intercept term is significantly different from zero, this is taken as evidence for horizontal pleiotropy<sup>50,51</sup>. MR-PRESSO identifies pleiotropic outliers that potentially contribute towards genetic heterogeneity under an IVW framework through non-parametric approaches<sup>52</sup>. In contrast, the GTCA-GSMR framework is a parametric-based method that derives the MR estimate using a generalized two-step least squares (2SLS) approach while accounting for pleiotropic outliers with the HEIDI-outlier technique<sup>25</sup>. Applying these models in parallel can help triangulate MR inferences through parametric and non-parametric approaches, making results more robust and less likely to be affected by weak violations of MR assumptions. We also generated funnel plots to inspect the presence of pleiotropic outliers manually, as asymmetry in the distribution of effect sizes indicates unbalanced horizontal pleiotropy. Cochran's Q and I<sup>2</sup> statistics were calculated to test for heterogeneity across estimates<sup>50,51</sup>.

To facilitate the interpretation of the effect estimates from binary variables, the odds ratios were multiplied by 0.693 to obtain the OR 'per doubling of odds', a concept introduced by Burgess et al.<sup>55</sup> that refers to a doubling in the prevalence of the binary exposure.

**Multivariable MR analysis adjusting for competing substance use phenotypes.** While our univariate MR analyses (i.e., fitting our exposure of interests one at a time) indicate whether genetic predisposition to these risk factors is associated with PD risk, it remains unclear whether other highly correlated environmental risk factors potentially mediate those relationships. To evaluate the direct effect between our exposures of interest and PD risk, we fitted another IVW regression model by regressing out the genetic association between our SNP instruments (for each trait respectively) and the following putative risk factors in our MR analyses: BMI, years of education, smoking status and cigarettes per day (for the DPW MR only); alcohol intake (for the smoking-related trait MR only). For BMI (kg/m<sup>2</sup>) and education attainment (in SD units), we derived the genetic effect size estimates from the UKBB cohort. The UKBB data were QC-ed as per previous work<sup>47</sup>. The GWAS for BMI (UKB data-field ID: 21001) included 437,458 individuals, while normalized educational attainment (UKB data-field ID: 6138) included 214,999 individuals. We used BOLT-LMM<sup>56</sup> mixed model software adjusting for recruitment age, sex and the first ten ancestral principal components. For alcohol and smoking-related traits, we adopted the effect size estimates from the GSCAN summary statistics.

**Frailty analysis.** We evaluated if survival bias could explain the MR results obtained from this study by following the methodology described by Noyce et al.<sup>35</sup> in which the model assumptions are that either the *drinks per week* or *smoking continuation* are associated with mortality and PD is associated with age. In summary, the approach consisted in the following steps:

1. Simulating a large population (n = 6 million) where individuals had alleles at each evaluated SNPs (either 31 for *drinks per week* or 7 for *smoking continuation*).
2. Values for *drinks per week* or *smoking continuation* were simulated based on mean and sd values reported by Liu et al.<sup>28</sup>.
3. Age values were generated to match the age distributions reported in Nalls et al.<sup>26</sup>.
4. Influence of our traits of interest on survival was stimulated following an all-cause mortality hazard ratio of 1.12 for *heavy alcohol consumption*<sup>42</sup> and 1.62 for *current smokers*<sup>57</sup> used to generate a Gompertz–Makeham mortality curve. Alive-death status was then sampled from the probability of death obtained from the survival curve.
5. Simulation of PD status was based upon the age-related diagnosis reported from Driver et al.<sup>58</sup>.
6. A sample of 56,306 simulated cases with PD and 1.4 million controls following the age distributions reported by Nalls et al.<sup>26</sup> was extracted. Finally, observational associations and MR analysis were performed over the simulated individuals. The known observational effect estimates were extracted from those reported by Zhang et al.<sup>10</sup> for *alcohol consumption* and Gallo et al.<sup>14</sup> for *smoking continuation*.
7. This process was repeated 1000 times to obtain the effect size distribution that can be attributable to frailty effects alone.

Code for the frailty analysis is accessible in the github repository.

### Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request. To apply and access 23andMe summary statistics, please visit <http://research.23andme.com/dataset-access/> for more information.

### Code availability

All code for this study is available at our GitHub repository: <http://github.com/cdomingu/MRandPD>.

Received: 6 November 2020; Accepted: 8 June 2021

Published online: 07 July 2021

## References

- Ascherio, A. & Schwarzschild, M. A. The epidemiology of Parkinson's disease: Risk factors and prevention. *Lancet Neurol.* **15**, 1257–1272 (2016).
- Mittal, S. *et al.*  $\beta$ -Adrenoreceptor is a regulator of the  $\alpha$ -synuclein gene driving risk of Parkinson's disease. *Science* **357**, 891–898 (2017).
- Dong, X. *et al.* Enhancers active in dopamine neurons are a primary link between genetic variation and neuropsychiatric disease. *Nat. Neurosci.* **21**, 1482–1492 (2018).
- Liu, G. *et al.* Specifically neuropathic Gaucher's mutations accelerate cognitive decline in Parkinson's. *Ann. Neurol.* **80**, 674–685 (2016).
- Liu, G. *et al.* Prediction of cognition in Parkinson's disease with a clinical–genetic score: A longitudinal analysis of nine cohorts. *Lancet Neurol.* **16**, 620–629 (2017).
- Ascherio, A. *et al.* Pesticide exposure and risk for Parkinson's disease. *Ann. Neurol.* **60**, 197–203 (2006).
- Wang, A. *et al.* Parkinson's disease risk from ambient exposure to pesticides. *Eur. J. Epidemiol.* **26**, 547–555 (2011).
- Pan, T., Li, X. & Jankovic, J. The association between Parkinson's disease and melanoma. *Int. J. Cancer* **128**, 2251–2260 (2011).
- Eriksson, A. K., Löfvig, S., Callaghan, R. C. & Allebeck, P. Alcohol use disorders and risk of Parkinson's disease: Findings from a Swedish national cohort study 1972–2008. *BMC Neurol.* **13**, 190 (2013).
- Zhang, D., Jiang, H. & Xie, J. Alcohol intake and risk of Parkinson's disease: A meta-analysis of observational studies. *Mov. Disord.* **29**, 819–822 (2014).
- Bettiol, S. S., Rose, T. C., Hughes, C. J. & Smith, L. A. Alcohol consumption and Parkinson's disease risk: A review of recent findings. *J. Parkinson's Dis.* **5**, 425–442 (2015).
- Thacker, E. L. *et al.* Temporal relationship between cigarette smoking and risk of Parkinson disease. *Neurology* **68**, 764–768 (2007).
- Li, X., Li, W., Liu, G., Shen, X. & Tang, Y. Association between cigarette smoking and Parkinson's disease: A meta-analysis. *Arch. Gerontol. Geriatr.* **61**, 510–516 (2015).
- Gallo, V. *et al.* Exploring causality of the association between smoking and Parkinson's disease. *Int. J. Epidemiol.* **48**, 912–925 (2019).
- Ross, G. W. *et al.* Association of coffee and caffeine intake with the risk of Parkinson disease. *J. Am. Med. Assoc.* **283**, 2674–2679 (2000).
- Costa, J., Lunet, N., Santos, C., Santos, J. & Vaz-Carneiro, A. Caffeine exposure and the risk of Parkinson's disease: A systematic review and meta-analysis of observational studies. *J. Alzheimer's Dis.* **20**, S221–S238 (2010).
- Glatt, S. L. *et al.* Risk factors for dementia in Parkinson's disease: Effect of education. *Neuroepidemiology* **15**, 20–25 (1996).
- Campdelacru, J. Parkinson's disease and Alzheimer disease: Environmental risk factors. *Neurologia* **29**, 541–549 (2014).
- Boef, A. G. C., Dekkers, O. M. & Le Cessie, S. Mendelian randomization studies: A review of the approaches used and the quality of reporting. *Int. J. Epidemiol.* **44**, 496–511 (2015).
- Smith, G. D. & Hemani, G. Mendelian randomization: Genetic anchors for causal inference in epidemiological studies. *Hum. Mol. Genet.* **23**, 89–98 (2014).
- Kendall, J. M. Designing a research project: Randomised controlled trials and their principles. *Emerg. Med. J.* **20**, 164–168 (2003).
- Lawlor, D. A., Harbord, R. M., Sterne, J. A. C., Timpson, N. & Davey Smith, G. Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. *Stat. Med.* **27**, 1133–1163 (2008).
- Noyce, A. J. & Nalls, M. A. Mendelian randomization—the key to understanding aspects of Parkinson's disease causation?. *Mov. Disord.* **31**, 478–483 (2016).
- Burgess, S., Small, D. S. & Thompson, S. G. A review of instrumental variable estimators for Mendelian randomization. *Stat. Methods Med. Res.* **26**, 2333–2355 (2017).
- Zhu, Z. *et al.* Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat. Commun.* <https://doi.org/10.1038/s41467-017-02317-2> (2018).
- Nalls, M. A. *et al.* Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: A meta-analysis of genome-wide association studies. *Lancet Neurol.* **18**, 1091–1102 (2019).
- Noyce, A. J. *et al.* The Parkinson's disease Mendelian randomization research portal. *Mov. Disord.* **34**, 1864–1872 (2019).
- Liu, M. *et al.* Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat. Genet.* **51**, 237–244 (2019).
- Willer, C. J. *et al.* Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat. Genet.* **40**, 161–169 (2008).
- Köttgen, A. *et al.* Genome-wide association analyses identify 18 new loci associated with serum urate concentrations. *Nat. Genet.* **45**, 145–154 (2013).
- Astle, W. J. *et al.* The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell* **167**, 1415–1429.e19 (2016).
- Muramatsu, T. *et al.* Alcohol and aldehyde dehydrogenase genotypes and drinking behavior of Chinese living in Shanghai. *Hum. Genet.* **96**, 151–154 (1995).
- Bierut, L. J. *et al.* ADH1B is associated with alcohol dependence and alcohol consumption in populations of European and African ancestry. *Mol. Psychiatry* **17**, 445–450 (2012).
- Luciano, M. *et al.* Association analysis in over 329,000 individuals identifies 116 independent variants influencing neuroticism. *Nat. Genet.* **50**, 6–11 (2018).
- Noyce, A. J. *et al.* Estimating the causal influence of body mass index on risk of Parkinson disease: A Mendelian randomisation study. *PLoS Med.* **14**, 1–19 (2017).
- Hartwig, F. P., Davies, N. M., Hemani, G. & Smith, G. D. Counterfactual causation: Avoiding the downsides of a powerful, widely applicable but potentially fallible technique. *Int. J. Epidemiol.* **45**, 1717–1726 (2016).
- Burgess, S., Davies, N. M. & Thompson, S. G. Bias due to participant overlap in two-sample Mendelian randomization. *Genet. Epidemiol.* **40**, 597–608 (2016).
- Edenberg, H. J. The genetics of alcohol metabolism: Role of alcohol dehydrogenase and aldehyde dehydrogenase variants. *Alcohol Res. Health* **30**, 5–13 (2007).
- García-Martin, E. *et al.* Association between the missense alcohol dehydrogenase rs1229984T variant with the risk for Parkinson's disease in women. *J. Neurol.* **266**, 346–352 (2019).
- Clarke, T. K. *et al.* Genome-wide association study of alcohol consumption and genetic overlap with other health-related traits in UK biobank (N=112117). *Mol. Psychiatry* **22**, 1376–1384 (2017).
- U.S. Department of Health and Human Services and U.S. Department of Agriculture. *2015–2020 Dietary Guidelines for Americans* (USDA, 2015).
- Xi, B. *et al.* Relationship of alcohol consumption to all-cause, cardiovascular, and cancer-related mortality in U.S. adults. *J. Am. Coll. Cardiol.* **70**, 913–922 (2017).



43. Borland, R., Partos, T. R., Yong, H. H., Cummings, K. M. & Hyland, A. How much unsuccessful quitting activity is going on among adult smokers? Data from the International Tobacco Control Four Country cohort survey. *Addiction* **107**, 673–682 (2012).
44. Javitz, H. S., Brigham, J., Lessov-Schlaggar, C. N., Krasnow, R. E. & Swan, G. E. Association of tobacco dependence and quit attempt duration with Rasch-modeled withdrawal sensitivity using retrospective measures. *Addiction* **104**, 1027–1035 (2009).
45. Wilsnack, R. W., Wilsnack, S. C., Kristjanson, A. F., Vogeltanz-Holm, N. D. & Gmel, G. Gender and alcohol consumption: Patterns from the multinational GENACIS project. *Addiction* **104**, 1487–1500 (2009).
46. Ezzati, M. & Lopez, A. D. Estimates of global mortality attributable to smoking in 2000. *Lancet* **362**, 847–852 (2003).
47. Gharahkhani, P. *et al.* Effect of increased body mass index on risk of diagnosis or death from cancer. *Br. J. Cancer* **120**, 565–570 (2019).
48. Brion, M. J. A., Shakhbazov, K. & Visscher, P. M. Calculating statistical power in Mendelian randomization studies. *Int. J. Epidemiol.* **42**, 1497–1501 (2013).
49. Burgess, S., Butterworth, A. & Thompson, S. G. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet. Epidemiol.* **37**, 658–665 (2013).
50. Hemani, G., Bowden, J. & Davey Smith, G. Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Hum. Mol. Genet.* **27**, R195–R208 (2018).
51. Bowden, J. *et al.* Assessing the suitability of summary data for two-sample mendelian randomization analyses using MR-Egger regression: The role of the I<sup>2</sup> statistic. *Int. J. Epidemiol.* **45**, 1961–1974 (2016).
52. Verbank, M., Chen, C. Y., Neale, B. & Do, R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat. Genet.* **50**, 693–698 (2018).
53. Ong, J. & MacGregor, S. Implementing MR-PRESSO and GCTA-GSMR for pleiotropy assessment in Mendelian randomization studies from a practitioner’s perspective. *Genet. Epidemiol.* **43**, 609–616 (2019).
54. Taylor, A. E. *et al.* Mendelian randomization in health research: Using appropriate genetic variants and avoiding biased estimates. *Econ. Hum. Biol.* **13**, 99–106 (2014).
55. Burgess, S. & Labrecque, J. A. Mendelian randomization with a binary exposure variable: Interpretation and presentation of causal estimates. *Eur. J. Epidemiol.* **33**, 947–952 (2018).
56. Loh, P. R., Kichaev, G., Gazal, S., Schoech, A. P. & Price, A. L. Mixed-model association for biobank-scale datasets. *Nat. Genet.* **50**, 906–908 (2018).
57. Qiao, Q., Tervahauta, M., Nissinen, A. & Tuomilehto, J. Mortality from all causes and from coronary heart related to smoking and changes in smoking during a 35-year follow-up of middle-aged finnish men. *Eur. Heart J.* **21**, 1621–1626 (2000).
58. Driver, J. A., Logroscino, G., Gaziano, J. M. & Kurth, T. Incidence and remaining lifetime risk of Parkinson disease in advanced age. *Neurology* **72**, 432–438 (2009).

## Acknowledgements

We thank the research participants and employees of 23andMe for their contribution to this study. Also, we thank Dr. Alastair Noyce and Dr. Gibran Hemani for sharing the Frailty Analysis code. CDB was supported by a Parkinson’s Foundation-APDA Summer Student Fellowship (PF-APDA-SFW-1913) and the UNAM “Programa 2019 para Actividades Especiales de Cooperación Interinstitucional (PAECI)”. XD was supported by Grants from the American Parkinson’s Disease Association. MER was supported by the Australian National Health & Medical Research Council and the Australian Research Council through an NHMRC-ARC Dementia Research Development Fellowship (GNT1002821). CRS was supported in part by NIH Grants U01NS082157, U01NS095736, U01NS100603, and the American Parkinson Disease Association.

## Author contributions

M.E.R. and X.D. designed, conceived, and jointly supervised the study together with C.R.S. C.D.B. and J.S.O. performed and interpreted the statistical analyses. C.D.B. drafted the first version of the manuscript with the support of M.E.R., X.D. and J.S.O. All authors assisted in the preparation and editing of the manuscript and approved its final version.

## Competing interests

CRS has collaborated with Pfizer, Opko, Proteome Sciences, Genzyme Inc; has consulted for Genzyme; has served as Advisor to the Michael J. Fox Foundation, NIH, Department of Defense; is on the Scientific Advisory Board of the American Parkinson Disease Association; has received funding from the NIH, the U.S. Department of Defense, the Harvard NeuroDiscovery Center, the Michael J. Fox Foundation, and American Parkinson Disease Association. XD has received funding from the NIH, the Michael J. Fox Foundation, and American Parkinson Disease Association. The other authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-93105-y>.

**Correspondence** and requests for materials should be addressed to M.E.R. or X.D.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021