International Journal of Epidemiology, 2019, 1438–1446
doi: 10.1093/ije/dyz081
Advance Access Publication Date: 10 May 2019
Original article

Mendelian Randomization

The effect of smoking intensity on all-cause and cause-specific mortality—a Mendelian randomization analysis

Gunnhild Åberge Vie [1]* Robyn E Wootton,[2,3,4,5] Johan Håkon Bjørngaard,[1] Bjørn Olav Åsvold,[6,7] Amy E Taylor,[2,8] Maiken Elvestad Gabrielsen,[7] George Davey Smith [2,5,8] Pål Richard Romundstad[1] and Marcus R Munafò[2,3,4,5]

[1]Department of Public Health and Nursing, NTNU, Norwegian University of Science and Technology, Trondheim, Norway, [2]MRC Integrative Epidemiology Unit (IEU), University of Bristol, Bristol, UK, [3]UK Centre for Tobacco and Alcohol Studies, University of Bristol, Bristol, UK, [4]Tobacco and Alcohol Research Group, School of Psychological Science, University of Bristol, Bristol, UK, [5]NIHR Bristol Biomedical Research Centre, University Hospitals Bristol NHS Foundation Trust and University of Bristol, Bristol, UK, [6]Department of Endocrinology, St Olavs Hospital, Trondheim University Hospital, Trondheim, Norway, [7]KG Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, NTNU, Norwegian University of Science and Technology, Trondheim, Norway and [8]Department of Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK

*Corresponding author. Faculty of Medicine and Health Sciences, Department of Public Health and Nursing, NTNU, Pb 8905, 7491 Trondheim, Norway. E-mail: gunnhild.vie@ntnu.no

Editorial decision 25 March 2019; Accepted 5 April 2019

Abstract

Background: Smoking is an important cause of mortality and recent studies have suggested that even low-intensity smoking might be associated with increased mortality. Still, smoking is associated with lower socio-economic status as well as other potential risk factors, and disease onset might motivate smoking cessation, thus residual confounding and reverse causality might bias results. We aimed to assess the evidence of a causal relationship between smoking intensity and cause-specific as well as all-cause-mortality using Mendelian randomization analyses.

Methods: We included 56 019 participants from the Norwegian HUNT2 Study and 337 103 participants from UK Biobank, linked to national registry data on causes of death. We estimated associations of self-reported smoking as well as the genetic variant rs1051730 as an instrument for smoking intensity with all-cause and cause-specific mortality. We subsequently meta-analysed the results from the two cohorts.

Results: Each effect allele of the rs1051730 was associated with a 9% increased hazard of all-cause mortality [95% confidence interval (CI) 6–11] among ever smokers. Effect alleles were also associated with death by neoplasms [hazard ratio (HR) 1.11, 95% CI 1.06–1.15], circulatory diseases (HR 1.06, 95% CI 1.01–1.11) and respiratory diseases (HR 1.15, 95%

CI 1.05–1.26) among ever smokers. The association was stronger among ever than never smokers for all-cause mortality ($p < 0.001$), neoplasms ($p = 0.001$) and respiratory diseases ($p = 0.038$).

Conclusions: Our results indicate a causal effect of smoking intensity on all-cause mortality and death by neoplasms and respiratory diseases. There was weaker evidence of a causal effect of smoking intensity on death by circulatory diseases.

Key words: Genetic variation, cigarette smoking, cause of death, epidemiology

Key Messages

- Observational studies indicate that even low-intensity smoking increases mortality.
- We used Mendelian randomization to investigate the causal associations between smoking intensity and all-cause and cause-specific mortality.
- Our study suggests that increased smoking intensity increases all-cause mortality and mortality by respiratory disease and neoplasms.
- Our study finds weaker evidence of a causal effect of smoking intensity on death by circulatory disease. This is consistent with the effect of being a smoker compared with not being a smoker being stronger than the effect of higher compared with lower smoking intensity on cardiovascular health.

Introduction

The association between tobacco smoking and increased mortality has been recognized for several decades.[1] The association with cardiovascular diseases, cancer and respiratory diseases is well established[2,3] and suggestive evidence exists for associations with several other diseases.[4] There is a clear dose–response relationship between smoking and mortality[5,6] and risk attenuates after smoking cessation.[7] Although there may be health benefits from reducing tobacco consumption from high to low intensity, the effect seems to be small[8,9] and even long-term very-low-intensity smoking has recently been associated with mortality[6] and cardiovascular diseases.[10]

The prevalence of heavy smoking has decreased over time in high-income countries,[11] whereas the prevalence of low-intensity smoking has been more stable, thereby constituting a larger share of smokers.[12] However, there is a clear social gradient in smoking[13] and smoking intensity is also higher among individuals with lower educational attainment.[14] Moreover, smoking is associated with other health behaviours and disease onset could motivate smoking cessation. Residual confounding and reverse causality are therefore critical issues in studies of the consequences of smoking.

The presumed causality between smoking and all-cause mortality has been supported by a study using Mendelian randomization.[15,16] Two single-nucleotide polymorphisms (SNPs) rs1051730 and rs16969968 are strongly associated with tobacco consumption among smokers. This association is believed to be causal for rs16969968 and the two SNPs are in perfect linkage disequilibrium with each other and considered interchangeable.[17] They are located in the CHRNA5-A3-B4 nicotinic receptor gene cluster and each additional copy of the effect allele corresponds to about one additional cigarette per day among smokers, with even stronger associations observed for blood cotinine levels, which is a more objective measure of tobacco consumption.[18] The rs1051730 can thus, under the assumptions that the risk alleles are evenly distributed within the population and have no effects on outcome other than through smoking behaviour, serve as an unconfounded measure of smoking intensity.[16] As smoking intensity can only affect mortality among smokers, we would expect associations between the SNP and mortality among ever smokers only. Never smokers can therefore be used as a negative control population, and any association between the SNP and outcomes among never smokers would suggest that the SNP is not a valid instrument for smoking intensity.[19] Stratification on a consequence of the exposure could cause a collider bias; however, a simulation study has indicated that a moderate selection bias will not severely affect the estimates of a Mendelian randomization study.[20] Smoking intensity ought not to cause smoking initiation but could cause smoking cessation; conditioning on ever

smoking is therefore less likely to introduce bias than conditioning on former vs current smoking.[21]

We aimed to expand the existing literature by assessing the evidence of a causal relationship between smoking intensity and cause-specific as well as all-cause-mortality, using Mendelian randomization.[16]

Methods

Study design

We performed Mendelian randomization analyses of the associations between the single-nucleotide polymorphism rs1051730 and all-cause and cause-specific mortality in two different study cohorts: the second wave of the Nord-Trøndelag Health Study (the HUNT2 Study 1995–97) and UK Biobank (2006–2010). We subsequently meta-analysed the effect estimates. For comparison, we also estimated the association between self-reported smoking and mortality. Both cohorts included participants of both sexes. Age at baseline ranged from 19 to 101 years in the HUNT2 Study and from 40 to 69 years in UK Biobank. Further details about the study populations and genotyping can be found in the Supplementary Methods, available as Supplementary data at *IJE* online.

Smoking behaviour

For analyses of self-reported smoking status, we categorized participants as never smokers, former smokers, current low-intensity smokers (<10 cigarettes per day), medium-intensity smokers (10–19 cigarettes per day) or high-intensity smokers (20 or more cigarettes per day). For the main Mendelian randomization analyses, we combined former and current smokers into ever smokers, in order to avoid potential collider bias from conditioning on a consequence of smoking intensity. However, assuming that the effects of smoking intensity on mortality are reversible, we would expect stronger association among current compared with former smokers. We therefore additionally categorized smokers into five subgroups based on time since smoking cessation: current smokers, former smokers who quit less than 5 years, 5–9 years, 10–19 years or ≥20 years before baseline, respectively. The potential bias from separate analyses by smoking status is further discussed in the Supplementary Methods, available as Supplementary data at *IJE* online.

Outcome

Date and cause of death until 31 December 2012 were collected from the Cause of Death Registry in Norway and until June 2017 in the UK. We defined all-cause mortality as death from any cause. We recoded all ICD-9-codes to ICD-10-codes ($n = 13$)[22] before we categorized causes of death according to the European short list of causes of death (see Supplementary Table 1, available as Supplementary data at *IJE* online).[23]

We excluded diseases of blood, blood-forming organs and immune system, skin and subcutaneous tissue, complications of pregnancies, conditions originating in the perinatal period, and congenital malformations and chromosomal abnormalities (ICD-10 codes D50–D89, O00–Q99) from the analyses of cause-specific mortality, due to few deaths from these causes.

Statistical analyses

We assessed the statistical evidence of associations between the number of effect alleles and smoking intensity as well as with possible confounders (age, sex, education and alcohol intake) within strata of smoking status, using linear regression for continuous variables and chi-square test for categorical variables. If effect alleles are not associated with measured confounders within strata of smoking status, bias from stratification is less likely.

All associations were analysed separately within UK Biobank and the HUNT2 cohorts, and estimates from the two cohorts were subsequently meta-analysed. A random-effects model was used for associations with self-reported smoking, whereas associations with the genetic instrument were performed assuming a fixed effect. We additionally performed random-effects meta-analyses because I^2 indicated substantial heterogeneity for several causes of death. We performed analyses using complete cases in Stata 15.0 and R (R Core Team, 2014).

Associations with self-reported smoking

We analysed the association between self-reported smoking status and mortality using Cox proportional hazard models with age as the time scale, adjusted for alcohol intake and in strata of education and sex. The association between smoking and body mass index is likely bidirectional; as smoking is associated with lower body mass index, a high body mass index can motivate smoking initiation.[24,25] We considered the mediating effect to be more important than the confounding effect in a setting with mostly middle-aged study participants, and hence did not adjust for body mass index. Different causes of mortality were assessed in separate models and participants were censored at time of death from any other cause than the one used as an outcome in any given model, as well as at time of emigration or end of follow-up. We assessed the proportional-hazards assumption using Schoenfeld residuals and visual inspection of log-log plots.

Due to indication of non-proportionality, we performed additional analyses in strata of age.

Mendelian randomization

We similarly analysed the association between the rs1051730 effect allele and death using Cox proportional-hazards models with age as the time scale and stratified by sex. As the association between the SNP and number of cigarettes smoked was consistent with an additive effect, we treated the number of effect alleles as a continuous variable in the analyses.

As smoking intensity can only affect mortality among smokers, we would expect associations between the SNP and mortality among ever smokers only. We therefore performed analyses separately for ever smokers and never smokers, and we report the statistical evidence of heterogeneity in the association over these subgroups as the p for difference. Similar associations in never smokers as in ever smokers would suggest a pathway from the effect alleles to the outcome that is not mediated by smoking intensity; hence, it would suggest that the association is not evidence of a causal relationship.

Shoenfeld's residuals did not provide statistical evidence of non-proportional hazards for the association between the SNP and mortality in strata of smokers/never smokers. We still examined the association separately before and after age 80, as visual inspection of log-log plots indicated possible violation of the proportional-hazards assumptions. These analyses were performed in the HUNT Study only, as there were no deaths after age 80 in the UK Biobank sample.

To assess the potential benefit of smoking cessation, we also performed Mendelian randomization analyses separately for five groups based on current smoking status and time since smoking cessation for former smokers. We do not show results for causes of death where the number of deaths in the strata of previous smokers was less than 20 in the UK Biobank and the HUNT Study combined.

Results

Our study sample included 56 019 participants from the HUNT2 Study and 337 103 participants from UK Biobank. Of these, 11 303 and 9634 participants, respectively, died within follow-up. The most common causes of death were neoplasms and diseases of the circulatory system (Supplementary Table 1, available as Supplementary data at *IJE* online). Participants in the HUNT2 Study were followed for up to 17.4 years and median age at death was 82 years (interquartile range 75–88) compared with a maximum of 11 years of follow-up and a median age at death of 67 years (interquartile range 62–70) in the UK Biobank sample.

We confirmed the association between rs1051730 and intensity of smoking; for each effect allele, current smokers smoked on average 0.7 cigarettes [95% confidence interval (CI) 0.5–0.8] more per day in the HUNT2 Study and 1.0 (95% CI 0.9–1.1) cigarettes more per day in the UK Biobank sample, with a near linear increase in the number of cigarettes smoked per effect allele. The SNP was also associated with being a current rather than former smoker ($p < 0.001$) and with lower age at participation among ever smokers ($p = 0.005$ in the HUNT Study, $p = 0.002$ in UK Biobank). Possible confounders were evenly distributed according to number of effect alleles (Supplementary Tables 2–5, available as Supplementary data at *IJE* online). For each risk allele, the odds of ever having smoked daily increased by 2% (95% CI 1–3, $p < 0.001$) and 4% (95% CI 1–6, $p = 0.005$) in UK Biobank and the HUNT Study, respectively.

Self-reported smoking

We found a clear dose–response relationship between self-reported smoking and all-cause mortality, mortality by neoplasm, circulatory and respiratory diseases, diseases of the digestive system and external causes of death (Figure 1), supported by strong statistical evidence of associations with current smoking intensity (all p-values <0.001). Former smokers had an intermediate risk between never and low-intensity smokers for all of these causes of death, except that there was no increased risk of death by external causes for former compared with never smokers. We found indication of non-proportionality of hazards for all-cause mortality and death by mental disorders, neurological, circulatory and respiratory diseases. Additional analyses indicated weaker associations between smoking and mortality at older ages (Supplementary Tables 8 and 9, available as Supplementary data at *IJE* online).

Mendelian randomization analyses

Mendelian randomization analyses indicated a causal mechanism behind the observed association between smoking intensity and all-cause mortality among ever smokers [hazard ratio (HR) 1.09, 95% CI 1.06–1.11 per risk allele—see Figure 2]. Furthermore, Mendelian randomization results supported a causal effect of smoking intensity on death by neoplasms (HR 1.11, 95% CI 1.06–1.15 per risk allele) and respiratory diseases (HR 1.15, 95% CI 1.05–1.26 per risk allele). The number of effect alleles was not associated with higher mortality from either of these causes among never smokers (HRs 1.00, 1.00 and 0.95, and p for difference <0.001, 0.001 and 0.032, respectively).

Each effect allele was associated with an increased hazard rate of death by circulatory diseases among ever

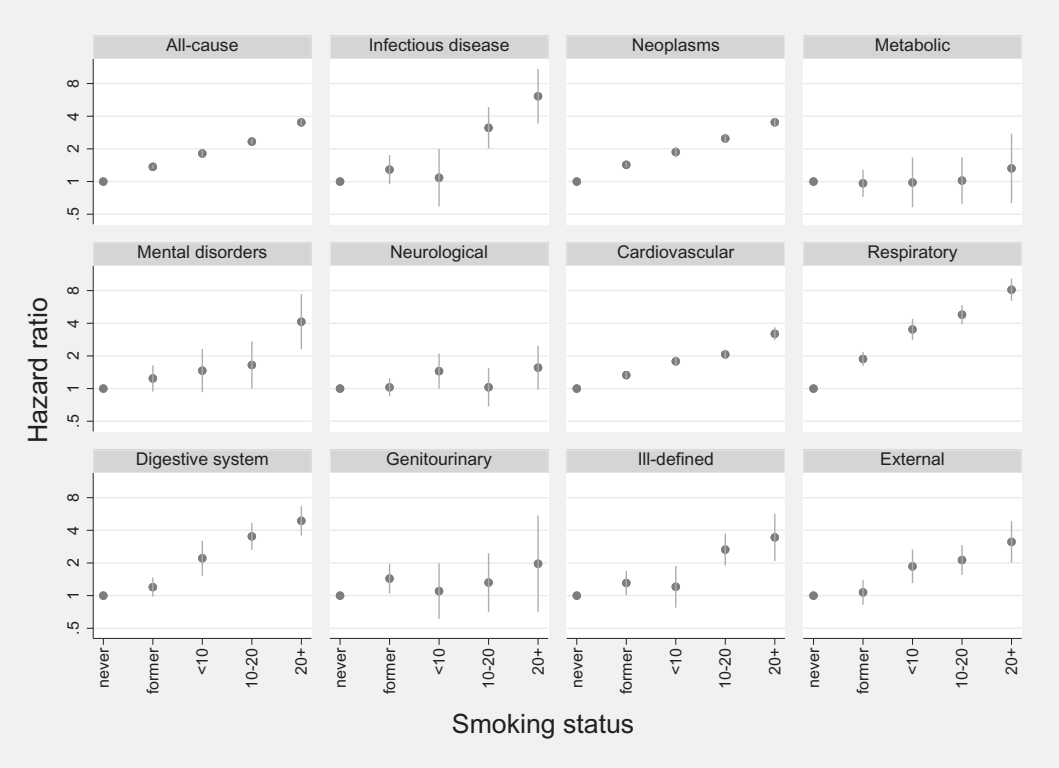

Figure 1. Associations between smoking status (given as never smoker, former smoker or by number of cigarettes smoked per day among current smokers) and mortality, adjusted for age, sex, education and alcohol intake. Results from a random-effects meta-analysis of estimates from the HUNT2 Study (1995–97) ($n = 52\,561$) and UK Biobank (2006–10). Graphs are truncated at hazard ratio = 0.5.

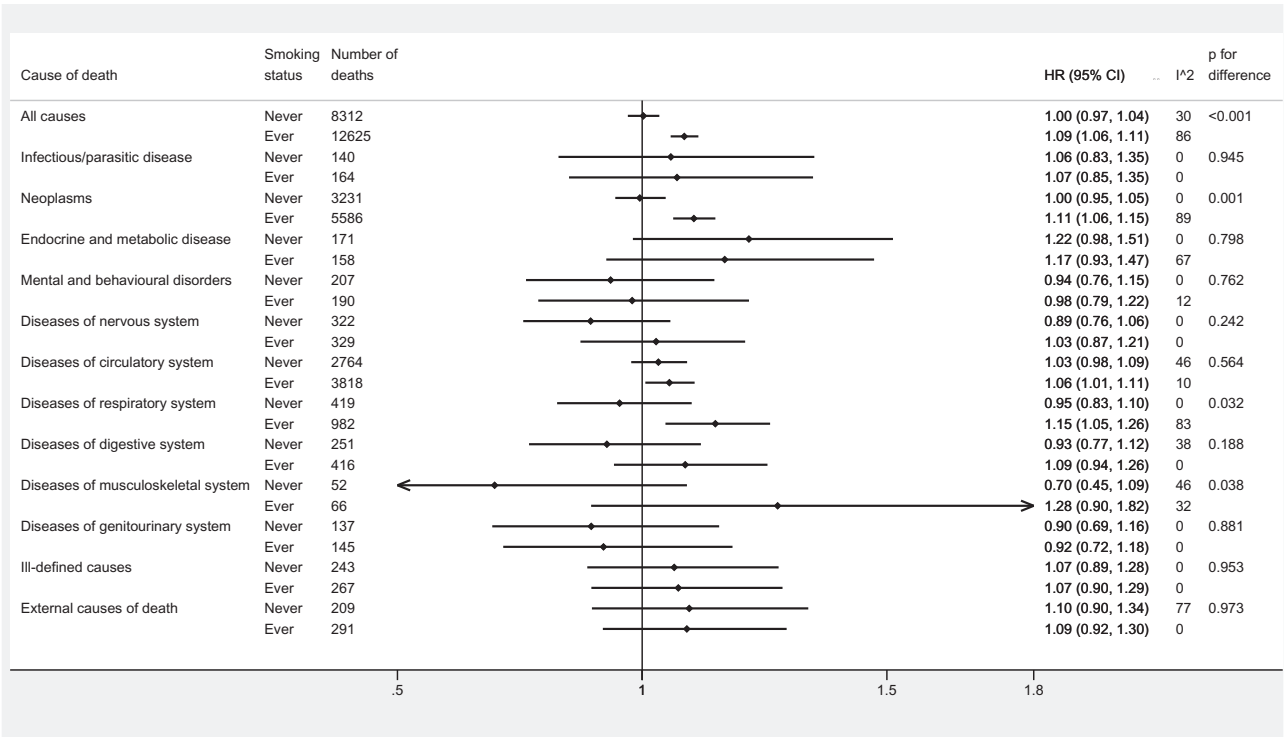

Figure 2. Associations between number of smoking increasing alleles of rs1071530 and mortality. Results from a fixed-effect meta-analysis of estimates from the HUNT2 Study (1995–97) and UK Biobank (2006–10).

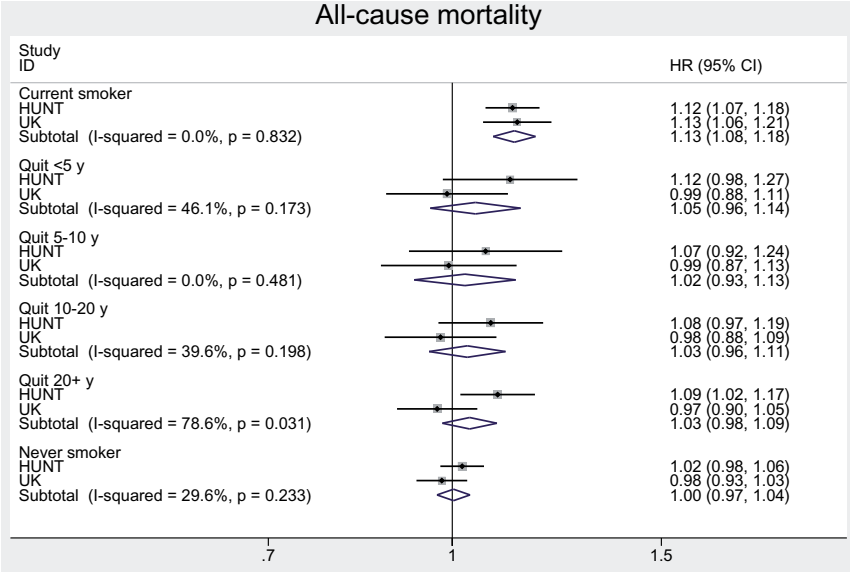

Figure 3. Associations between number of smoking increasing alleles of rs1071530 and all-cause mortality. Results from a fixed-effect meta-analysis of estimates from the HUNT2 Study (1995–97) ($n = 55\,593$) and UK Biobank (2006–10). Results are presented separately by smoking status.

smokers (HR 1.06, 95% CI 1.01–1.11). The effect estimate among never smokers was only marginally smaller (HR 1.03, 95% CI 0.98–1.09), with weak statistical evidence of a difference (p for difference 0.568). In additional analyses, this association among never smokers was only found among those 80 years or older in the HUNT2 Study (Supplementary Table 12 and Supplementary Figure 6, available as Supplementary data at *IJE* online). Furthermore, we found some statistical evidence that the association between effect alleles and mortality by musculoskeletal diseases differed between smokers and non-smokers (p for difference 0.038), but there was only weak evidence of an association among ever smokers (HR 1.28, 95% CI 0.90–1.82) and we did not find a corresponding association with self-reported smoking intensity.

For all other causes of death, we did not find strong statistical evidence of either associations between effect alleles and mortality among ever smokers or a stronger association among ever compared with never smokers.

There was substantial heterogeneity between the estimates from the HUNT2 Study and UK Biobank for several outcomes (I^2 up to 89%). Random-effects meta-analyses gave similar effect estimates to fixed-effect meta-analyses, but with lower precision (Supplementary Figure 3, available as Supplementary data at *IJE* online).

Furthermore, we found low heterogeneity between the HUNT2 Study and UK Biobank in the association among current smokers, whereas there was substantial heterogeneity among former smokers (Figure 3 and Supplementary Figures 4–8, available as Supplementary data at *IJE* online) and analyses of current smokers supported a causal effect of smoking intensity on all-cause mortality and death by neoplasm and respiratory diseases. We found only weak statistical evidence of an association between effect alleles and death by circulatory disease among current smokers (HR 1.05, 95% CI 0.97–1.13, $p = 0.213$) but with low heterogeneity between the HUNT2 Study and UK Biobank. For all-cause mortality and death by neoplasms, meta-analyses indicated that the excess risk associated with each effect allele attenuated with time (p for difference between subgroups 0.001) (Figure 3 and Supplementary Figure 4, available as Supplementary data at *IJE* online). The association between effect alleles and death by circulatory diseases did not seem to attenuate with time since smoking cessation (p for difference between subgroups 0.742) (Supplementary Figure 6, available as Supplementary data at *IJE* online); for other causes of death, the number of events was too low to draw clear conclusions.

Discussion

Using data from two large-scale cohort studies, we observed a clear dose–response relationship between self-reported smoking and all-cause mortality and mortality by neoplasm, circulatory diseases, respiratory diseases, diseases of the digestive system and external causes of death. Mendelian randomization analyses indicated a causal mechanism behind the observed association for all-cause mortality as well as for death by neoplasms and respiratory diseases. We found weaker evidence for a causal effect of smoking intensity on circulatory diseases as well. Results from Mendelian randomization analyses did not support a causal association between smoking intensity and mortality by diseases of the digestive system, mental diseases or external causes of death.

Strengths and limitations

The validity of Mendelian randomization analyses relies on three main assumptions: (i) the genetic instrument must be associated with the exposure of interest, (ii) it must only be associated with the outcome through the exposure of interest and (iii) it must be independent of confounders. The association between rs157030 and smoking intensity has been established previously and was confirmed in our study samples. The second and third assumptions cannot be directly verified, but can be supported by lack of associations between effect alleles and known confounders and by the use of negative controls.[19] An association between effect alleles and age at participation among ever smokers is to be expected, as smoking intensity increases mortality. The lack of associations between effect alleles and mortality among never smokers support the interpretation of a causal effect of smoking intensity on all-cause mortality, as well as death by neoplasms and respiratory diseases. A full instrumental variable estimation of the causal effect per cigarette smoked was not possible, as the number of cigarettes smoked does not fully describe the effect of the SNP on tobacco consumption.[18]

In the Mendelian randomization analyses, we adjusted for age and sex only, both of which had complete data, and missing on smoking status was a negligible problem. However, there is the potential for misclassification of the self-reported smoking variables. Misclassification of ever smokers as never smokers would bias the use of never smokers as a negative control; however, we have no indication of this from the analyses of all-cause mortality. Dates and causes of deaths were collected from national registries, so we thus had the best information available about causes of death; however, as causes of death are rarely determined based on post-mortem examinations, we cannot exclude misclassifications in causes of death either. This study gives a broad overview of the associations between a genetic variant associated with smoking intensity and mortality, but combining causes into broad categories could also mask the association with specific sub-causes.

Furthermore, whereas genetic variants are determined at conception, study participants did not enter the studies until adulthood. As we lack information about the association between effect alleles and mortality before study inclusion, there is a possibility that lower survival among exposed individuals could bias the results.[26]

Participation in the HUNT2 Study was about 70%[27] and the HUNT2 Study is fairly representative of the general population but, despite this, non-participants in the HUNT2 Study tend to have somewhat higher mortality as compared with participants.[28] In contrast, participation in UK Biobank was only about 5%, suggesting a higher selection to participation.[29] Representative samples are not a prerequisite for valid associations in longitudinal studies.[30] Still, selective participation could induce bias.[31] Whereas the low mortality associated with effect alleles among former smokers in the UK Biobank sample could be due to selection bias, the consistency of estimated associations between the two samples among current smokers suggest that selection bias is a lesser problem for this group of participants.

We found a weak association of the SNP with smoking initiation and stronger evidence of an association with smoking cessation. However, we found only weak evidence of associations between the SNP and important confounders within strata of smoking status. This weighs against any important collider bias being introduced by stratification, but we still cannot exclude it.[17]

Comparison with other studies

Evidence of a causal effect of smoking intensity on all-cause mortality, deaths by neoplasms and respiratory diseases is consistent with existing knowledge about the consequences of smoking.[3] A possible effect of smoking intensity on death by musculoskeletal and connective tissue diseases should not be given much emphasis, as the estimates were imprecise and inconsistent between methods. On the other hand, smoking has previously been associated with rheumatoid arthritis and systemic lupus erythematosus.[3]

Although smoking is also a well-established risk factor for circulatory diseases, our findings regarding a causal link were not conclusive. As described earlier, never smokers are used as a negative control group and associations between effect alleles and outcomes among never smokers indicate a pathway circumventing smoking intensity; hence, the association among smokers might not represent a causal effect of smoking intensity. As effect estimates for death by circulatory diseases were similar among never compared with ever smokers, we cannot exclude pleiotropic effects of the SNP as an explanation. However, the finding of an association between the effect allele and death from circulatory diseases among never smokers was inconsistent and the modest association found in ever smokers might therefore still represent a causal effect of smoking intensity on death by circulatory diseases. The gradient in risk associated with self-reported smoking intensity is also gentler for death by circulatory diseases compared with neoplasms and respiratory diseases. This is in accordance with a recent review in which smoking just one cigarette per day was found to confer up to half of the excess relative risk of cardiovascular diseases associated with smoking 20 cigarettes per day.[10] Finding evidence of a

weak causal effect only is thus consistent with the effects of smoking on circulatory diseases depending more on smoking status than on smoking intensity.

Our results did not indicate causality in the associations of smoking with infectious diseases, diseases of the digestive system and renal failure previously identified by Carter *et al.*,[4] although, again, it is important to bear in mind that the SNP we used is mainly an instrument of smoking intensity, not smoking status.

Our results strengthen the large body of evidence suggesting that smoking increases the risk of death from cancer and pulmonary disease. As this is a Mendelian randomization study, we can be even more confident that these are causal effects and not spurious associations due to reverse causation or residual confounding. This study looked at smoking intensity beyond smoking initiation and showed that higher smoking intensity was associated with greater risk of death. Although reducing smoking intensity might be insufficient to prevent cardiovascular deaths, our results imply that it is important to help individuals reduce smoking intensity or quit smoking as soon as possible.

Supplementary Data

Supplementary data are available at *IJE* online.

Funding

This work was supported by the Norwegian Research Council (grant number 250335) to G.Å.V.; NIHR Biomedical Research Centre at the University Hospitals Bristol NHS Foundation Trust and the University of Bristol; and the MRC Integrative Epidemiology Unit (grant number MC_UU_12013/6) to R.E.W. This study was supported by the NIHR Biomedical Research Centre at the University Hospitals Bristol NHS Foundation Trust and the University of Bristol. The views expressed in this publication are those of the authors and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health and Social Care. B.O.Å. and M.E.G. work in a research unit funded by Stiftelsen Kristian Gerhard Jebsen; Faculty of Medicine and Health Sciences, NTNU; The Liaison Committee for education, research and innovation in Central Norway; and the Joint Research Committee between St. Olavs Hospital and the Faculty of Medicine and Health Sciences, NTNU. The funders had no influence on the study design, data collection and analysis, decision to publish or preparation of the manuscript.

Acknowledgements

This research has been conducted using the UK Biobank Resource (application 9142). The Nord-Trøndelag Health Study (The HUNT Study) is a collaboration between HUNT Research Centre (Faculty of Medicine and Health Sciences, NTNU, Norwegian University of Science and Technology), Nord-Trøndelag County Council, Central Norway Regional Health Authority and the Norwegian Institute of Public Health. All participants in the HUNT2 Study and UK Biobank gave their informed consent to participation. The present study was approved by the Regional Ethical Committee in Mid-Norway (REK-2011/975) and UK Biobank has received ethics approval from the National Health Service National Research Ethics Service (ref. 11/NW/0382).

Conflict of interest: None declared.

References

1. Mucha L, Stephenson J, Morandi N, Dirani R. Meta-analysis of disease risk associated with smoking, by gender and intensity of smoking. *Gend Med* 2006;**3**:279–91.
2. Taghizadeh N, Vonk JM, Boezen HM. Lifetime smoking history and cause-specific mortality in a cohort study with 43 years of follow-up. *PLoS One* 2016;**11**:e0153310.
3. US Department of Health and Human Services. *The Health Consequences of Smoking—50 Years of Progress: A Report of the Surgeon General*. Atlanta, GA: US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, 2014.
4. Carter BD, Abnet CC, Feskanich D *et al*. Smoking and mortality—beyond established causes. *N Engl J Med* 2015;**372**:631–40.
5. Thun MJ, Carter BD, Feskanich D *et al*. 50-Year trends in smoking-related mortality in the United States. *N Engl J Med* 2013;**368**:351–64.
6. Inoue-Choi M, Liao LM, Reyes-Guzman C, Hartge P, Caporaso N, Freedman ND. Association of long-term, low-intensity smoking with all-cause and cause-specific mortality in the National Institutes of Health-AARP diet and health study. *JAMA Intern Med* 2017;**177**:87–95.
7. Banks E, Joshy G, Weber MF *et al*. Tobacco smoking and all-cause mortality in a large Australian cohort study: findings from a mature epidemic with current low smoking prevalence. *BMC Med* 2015;**13**:38.
8. Pisinger C, Godtfredsen NS. Is there a health benefit of reduced tobacco consumption? A systematic review. *Nicotine Tob Res* 2007;**9**:631–46.
9. Gerber Y, Myers V, Goldbourt U. Smoking reduction at midlife and lifetime mortality risk in men: a prospective cohort study. *Am J Epidemiol* 2012;**175**:1006–12.
10. Hackshaw A, Morris JK, Boniface S, Tang J-L, Milenković D. Low cigarette consumption and risk of coronary heart disease and stroke: meta-analysis of 141 cohort studies in 55 study reports. *BMJ* 2018;**360**:j5855.
11. Pierce JP, Messer K, White MM, Cowling DW, Thomas DP. Prevalence of heavy smoking in California and the United States, 1965–2007. *JAMA* 2011;**305**:1106–12.
12. Jamal A, Phillips E, Gentzke AS *et al*. Current cigarette smoking among adults—United States, 2016. *MMWR Morb Mortal Wkly Rep* 2018;**67**:53–59.
13. Hiscock R, Bauld L, Amos A, Fidler JA, Munafò M. Socioeconomic status and smoking: a review. *Ann N Y Acad Sci* 2012;**1248**:107–23.
14. Nagelhout GE, de Korte-de Boer D, Kunst AE *et al*. Trends in socioeconomic inequalities in smoking prevalence, consumption, initiation, and cessation between 2001 and 2008 in the

Netherlands. Findings from a national population survey. *BMC Public Health* 2012;**12**:303.
15. Rode L, Bojesen SE, Weischer M, Nordestgaard BG. High tobacco consumption is causally associated with increased all-cause mortality in a general population sample of 55,568 individuals, but not with short telomeres: a Mendelian randomization study. *Int J Epidemiol* 2014;**43**:1473–83.
16. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet* 2014;**23**:R89–98.
17. Furberg H, Kim Y, Dackor J *et al*. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet* 2010;**42**:441.
18. Munafo MR, Timofeeva MN, Morris RW *et al*. Association between genetic variants on chromosome 15q25 locus and objective measures of tobacco exposure. *J Natl Cancer Inst* 2012;**104**: 740–48.
19. Davies NM, Thomas KH, Taylor AE *et al*. How to compare instrumental variable and conventional regression analyses using negative controls and bias plots. *Int J Epidemiol* 2017;**46**:2067–77.
20. Gkatzionis A, Burgess S. Contextualizing selection bias in Mendelian randomization: how bad is it likely to be? *Int J Epidemiol* 2018. doi: 10.1093/ije/dyy202.
21. Taylor M, Rode L, Bjorngaard J *et al*. Is smoking heaviness causally associated with alcohol use? A Mendelian randomization study in four European cohorts. *Int J Epidemiol* 2018;**47**: 1098–105.
22. National Center for Health Statistics. *ICD-9*. 1998. ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Publications/ICD9-CM/1996/DISEASES/TABULAR/ (1 June 2018, date last accessed).
23. European Commission. Metadata. European Shortlist for Causes of Death, *2012*. 2012. http://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=LST_NOM_DTL&StrNom=COD_2012&StrLanguageCode=EN&IntPcKey=&StrLayoutCode=HIERARCHIC (1 June 2018, date last accessed).
24. Taylor A, Richmond R, Palviainen T *et al*. The effect of body mass index on smoking behaviour and nicotine metabolism: a Mendelian randomization study. *Hum Mol Gen* 2019;**28**:1322–30.
25. Carreras-Torres R, Johansson M, Haycock PC *et al*. Role of obesity in smoking behaviour: Mendelian randomisation study in UK Biobank. *BMJ* 2018;**361**:k1767.
26. VanderWeele TJ, Tchetgen EJT, Cornelis M, Kraft P. Methodological challenges in Mendelian randomization. *Epidemiology (Cambridge, Mass)* 2014;**25**:427.
27. Krokstad S, Langhammer A, Hveem K *et al*. Cohort profile: the HUNT study, Norway. *Int J Epidemiol* 2013;**42**:968–77.
28. Langhammer A, Krokstad S, Romundstad P, Heggland J, Holmen J. The HUNT study: participation is associated with survival and depends on socioeconomic status, diseases and symptoms. *BMC Med Res Methodol* 2012;**12**:143.
29. Fry A, Littlejohns TJ, Sudlow C *et al*. Comparison of sociodemographic and health-related characteristics of UK biobank participants with those of the general population. *Am J Epidemiol* 2017;**186**:1026–34.
30. Rothman KJ. Six persistent research misconceptions. *J Gen Intern Med* 2014;**29**:1060–64.
31. Ebrahim S, Davey Smith G. Commentary: Should we always deliberately be non-representative? *Int J Epidemiol* 2013;**42**: 1022–26.