CrossMark

Serum 25-hydroxyvitamin D levels and risk of lung cancer and histologic types: a Mendelian randomisation analysis of the HUNT study

Yi-Qian Sun[1], Ben M. Brumpton[2,3,4], Carolina Bonilla[4], Sarah J. Lewis[4], Stephen Burgess[5,6], Frank Skorpen[1], Yue Chen[7], Tom I.L. Nilsen[8,9], Pål Richard Romundstad [8] and Xiao-Mei Mai[8]

Affiliations: [1]Dept of Clinical and Molecular Medicine (IKOM), Norwegian University of Science and Technology (NTNU), Trondheim, Norway. [2]K.G. Jebsen Center for Genetic Epidemiology, Dept of Public Health and Nursing, Norwegian University of Science and Technology (NTNU), Trondheim, Norway. [3]Dept of Thoracic Medicine, St Olavs Hospital, Trondheim University Hospital, Trondheim, Norway. [4]MRC Integrative Epidemiology Unit, School of Social and Community Medicine, University of Bristol, Bristol, UK. [5]MRC Biostatistics Unit, University of Cambridge, Cambridge, UK. [6]Cardiovascular Epidemiology Unit, Dept of Public Health and Primary Care, University of Cambridge, Cambridge, UK. [7]School of Epidemiology and Public Health, Faculty of Medicine, University of Ottawa, Ottawa, Canada. [8]Dept of Public Health and Nursing, Norwegian University of Science and Technology (NTNU), Trondheim, Norway. [9]Clinic of Anesthesia and Intensive Care, St Olavs Hospital, Trondheim University Hospital, Trondheim, Norway.

Correspondence: Yi-Qian Sun, Dept of Clinical and Molecular Medicine (IKOM), Norwegian University of Science and Technology (NTNU), Erling Skjalgssons Gate 1, Laboratoriesenteret, 7491 Trondheim, Norway. E-mail: yi-qian.sun@ntnu.no

ABSTRACT We aimed to investigate potential causal associations between serum 25-hydroxyvitamin D (25(OH)D) levels and incidence of lung cancer overall and histologic types.

We performed a Mendelian randomisation analysis using a prospective cohort study in Norway, including 54 580 individuals and 676 incident lung cancer cases. A 25(OH)D allele score was generated based on the vitamin D-increasing alleles rs2282679, rs12785878 and rs10741657. Hazard ratios with 95% confidence intervals for incidence of lung cancer and histologic types were estimated in relation to the allele score. The inverse-variance weighted method using summarised data of individual single nucleotide polymorphisms was applied to calculate the Mendelian randomisation estimates.

The allele score accounted for 3.4% of the variation in serum 25(OH)D levels. There was no association between the allele score and lung cancer incidence overall, with HR 0.99 (95% CI 0.93–1.06) per allele score. A 25 nmol·L^{-1} increase in genetically determined 25(OH)D level was not associated with the incidence of lung cancer overall (Mendelian randomisation estimate HR 0.96, 95% CI 0.54–1.69) or any histologic type.

Mendelian randomisation analysis did not suggest a causal association between 25(OH)D levels and risk of lung cancer overall or histologic types in this population-based cohort study.

This article has supplementary material available from erj.ersjournals.com

Introduction

Vitamin D has been suggested to have a number of anticarcinogenic potentials, such as stimulating differentiation, inducing apoptosis, and inhibiting invasion and metastasis [1, 2]. Epidemiological studies of the associations between circulating vitamin D and various cancers have shown inconsistent results [3], [4]. Lung cancer has been the most common cancer type for several decades worldwide and it is also the most deadly cancer [5]. The main histologic types of lung cancer are small cell lung cancer, adenocarcinoma and squamous cell carcinoma [6]. Adenocarcinoma is the most common histologic type of lung cancer in many countries [7]. Unlike small cell lung cancer and squamous cell carcinoma, the association between smoking and adenocarcinoma is much weaker [8]. Thus, identifying risk factors other than tobacco smoking is necessary for further prevention of lung cancer overall and certain histologic types.

Two meta-analyses of observational cohort studies suggested an inverse association between serum vitamin D and risk of lung cancer overall [9, 10]. However, conventional observational epidemiological studies have limited capability to identify causal associations due to potential bias from confounding and reverse causation [11]. Although well-designed prospective cohort studies can reduce the possibility of reverse causation, residual and unmeasured confounding is inevitable in observational studies. However, Mendelian randomisation studies have been suggested to be able to overcome these limitations and to help make causal inferences of modifiable risk factors on health-related outcomes, provided that the crucial assumptions are satisfied [11, 12].

Therefore, we performed a Mendelian randomisation study using three single nucleotide polymorphisms (SNPs) as instrumental variables for serum 25-hydroxyvitamin D (25(OH)D), the primary circulating form of vitamin D, to explore potential causal associations of serum 25(OH)D levels with incidence of lung cancer overall and different histologic types in a population-based prospective cohort.

Material and methods

Study population and data linkage

The Nord-Trøndelag Health Study (HUNT) is a large population-based health study in Norway consisting of three separate surveys: HUNT1 (1984–1986), HUNT2 (1995–1997) and HUNT3 (2006–2008). The current study was based on data from HUNT2, in which 65 227 subjects aged ⩾20 years living in the county of Nord-Trøndelag participated (response rate 70%). All participants completed a general questionnaire including questions on health, lifestyle and socioeconomic status. Blood samples were drawn and body weight and height were measured at a clinical examination. The HUNT Research Center received updated information about deaths from all causes and emigration of HUNT participants from the Norwegian National Registry in which the dates of such events were recorded for all people living in Norway.

Using the unique 11-digit personal identification number of all residents in Norway, the data on HUNT2 participants were linked with data from the Cancer Registry of Norway (www.kreftregisteret.no). The Tenth Revision of the International Statistical Classification of Diseases and Related Health Problems topography codes C33–C34 were used to identify incident lung cancer cases among the HUNT2 study participants. Histologic types of lung cancer were classified according to the International Classification of Diseases of Oncology [13]. The participants in HUNT2 were followed from the date of participation to the date of lung cancer diagnosis, death, emigration or end of follow-up (December 31, 2014), whichever occurred first.

We excluded subjects who reported ever-cancer (n=2400) in the HUNT2 questionnaire at baseline, lung cancer cases diagnosed before the participation date (n=13) in the HUNT2 study and subjects who did not have information on genotype (n=8234), leaving 54 580 subjects in the analysis cohort. Moreover, a 10% random sample (n=6613) of the HUNT2 participants was selected as a subcohort for serum 25(OH)D measurement. After further excluding individuals without serum and genotype information, 5546 individuals remained in the analysis subcohort.

Measurement and standardisation of serum 25(OH)D levels

Serum 25(OH)D level is widely recognised as the best available proxy measure for body vitamin D status [14, 15]. Serum 25(OH)D levels were measured at the HUNT Biobank using LIAISON 25 OH Vitamin D TOTAL (DiaSorin, Saluggia, Italy), a fully automated, antibody-based, chemiluminescence assay. The detection range of the assay is 10–375 nmol·L^{-1}. As seasonal fluctuations in 25(OH)D levels were expected due to the high-latitude geographical position of Norway, a cosinor model based on month of blood draw was used to calculate a season-standardised 25(OH)D level (nmol·L^{-1}) that represents the annual average value of 25(OH)D for each subject [16]. The season-standardised 25(OH)D was calculated using the package cosinor version 1.1 in R version 3.4.2 (www.r-project.org).

Genotyping and imputation of SNPs and allele score as instrumental variables

DNA was isolated from blood samples collected in HUNT2 and stored at the HUNT Biobank. Genotyping was performed using HumanCoreExome (Illumina, San Diego, CA, USA) arrays as described elsewhere [17]. Imputation was performed on samples of recent European ancestry using Minimac3 version 2.0.1 (http://genome.sph.umich.edu/wiki/Minimac3) [18] from a merged reference panel constructed from the Haplotype Reference Consortium panel release version 1.1 [19] and a local reference panel based on 2201 whole-genome sequenced HUNT participants [20]. In total, three SNPs located in or near genes for vitamin D synthesis and metabolism were selected as instrumental variables for serum 25(OH)D based on two widely cited genomewide association studies [21, 22]: rs2282679 (*GC*), rs12785878 (*NADSYN1/DHCR7*) and rs10741657 (*CYP2R1*). Information on rs6013897 that was included in Wang *et al.* [22] and its proxy SNPs was not available in the HUNT study; this SNP, however, showed the weakest effect on 25(OH)D level [22, 23]. The effect allele (25(OH)D-increasing allele) was coded as 1 and the other allele was coded as 0 (rs2282679: T=1; rs12785878: T=1; rs10741657: A=1). A 25(OH)D allele score, which was the sum of the number of effect alleles of rs2282679, rs12785878 and rs10741657, was generated to increase the statistical power of the analyses [24]. The R^2-values for linkage disequilibrium between these three SNPs were calculated [25].

Statistical analyses

Linear regression was applied to calculate the F-statistic and R^2-value between SNPs or the allele score and season-standardised 25(OH)D levels. Values of the F-statistic >10 suggest that the SNPs or allele score are valid instrumental variables [11]. Linear regression was used to estimate the associations between the allele score and continuous covariates in order to test the assumption that the instrumental variables were not associated with potential confounders for the association between serum 25(OH)D and lung cancer; logistic regression was used in corresponding analyses of binary covariates. To test if there was a causal association between serum 25(OH)D and risk of lung cancer, we used Cox proportional hazards regression to calculate hazard ratios with 95% confidence intervals for the incidence of lung cancer overall or histologic types in relation to the allele score. Age was used as the time scale in the models. The proportional hazards assumption was satisfied for all SNPs and the allele score. In analyses estimating the risk of a specific histologic type, all other subtypes were censored at the date of diagnosis.

To calculate Mendelian randomisation estimates of serum 25(OH)D on lung cancer risk, we generated summarised data of coefficients and standard errors from linear regression of individual SNPs on season-standardised 25(OH)D levels in the subcohort (n=5546), as well as coefficients (ln(hazard ratio)) and standard errors from Cox regression of individual SNPs on risk of lung cancer overall or a histologic type in the cohort (n=54 580). Inverse-variance weighted (IVW) and median-based methods were used for the summarised data to calculate Mendelian randomisation estimates of serum 25(OH)D for lung cancer overall and histologic types [26]. An IVW estimate of the causal effect combines the ratio estimates using each genetic variant in a fixed-effect meta-analysis model [27]. To test for pleiotropy we used MR-Egger to calculate the intercept and 95% confidence intervals [28]. Additionally, we tested for heterogeneity between SNPs using I^2 and Cochran's Q-statistic. To test the robustness of our findings, we performed a two-sample Mendelian randomisation as sensitivity analysis using summarised data of SNP–25(OH)D associations derived from a previous consortium study (n~35 000) [23].

Analyses with summarised data of individual SNPs were carried out using the package MendelianRandomization version 0.2.2 in R version 3.4.2. All other statistical analyses were performed with Stata/SE version 14.2 (StataCorp, College Station, TX, USA).

Ethics

The study was approved by the Norwegian Regional Committees for Medical and Health Research Ethics. All participants gave their informed consent on participation in HUNT, linkage to previous HUNT surveys and specific registries.

Results

During a median follow-up of 18 years, a total of 676 incident lung cancer cases were diagnosed among the 54 580 cohort participants. Table 1 shows the distribution of baseline characteristics in the cohort (n=54 580) and subcohort (n=5546) of the HUNT2 study. In general, the distribution of baseline characteristics was similar between the cohort and subcohort. Supplementary table S1 presents the characteristics of SNPs included in the 25(OH)D allele score in the HUNT2 study. There was no evidence of departure from the Hardy–Weinberg equilibrium for the three SNPs. The allele frequency was in line with that of the 1000 Genomes Phase 3 data (www.internationalgenome.org). The R^2-values for linkage disequilibrium between the three SNPs were <0.1.

TABLE 1 Distribution of baseline characteristics in the cohort and subcohort of the HUNT2 study (1995–1997)

	Cohort	Subcohort[#]
Subjects n	54 580	5546
Age years	49.2±16.6	49.1±16.6
Season-standardised 25(OH)D level nmol·L^{-1}		48.3±17.1
Lung cancer cases n	676	77
Sex		
Female	52.7	52.8
Male	47.3	47.2
Smoking		
Never	42.3	42.5
Ever	55.8	55.4
Unknown	1.9	2.1
Family history of cancer		
No	74.5	74.1
Yes	25.5	25.9
Education years		
<10	33.6	32.9
⩾10	61.8	62.5
Unknown	4.6	4.6
Economic difficulties		
No	49.8	51.0
Yes	21.8	21.1
Unknown	28.3	27.9
Body mass index kg·m^{-2}		
<25	39.8	40.6
⩾25	59.4	58.8
Unknown	0.7	0.6
Physical activity		
Inactive	21.4	21.1
Active	48.8	49.0
Unknown	29.9	29.8
Alcohol consumption		
Never	33.6	32.9
⩾1 times per month	58.0	59.0
Unknown	8.4	8.1
Chronic bronchitis		
No	94.6	95.0
Yes	3.4	3.4
Unknown	2.0	1.7

Data are presented as mean±SD or %, unless otherwise stated. 25(OH)D: 25-hydroxyvitamin D. [#]: those with measured serum 25(OH)D levels and genotype information.

F-statistics and R^2-values between SNP/allele score and season-standardised 25(OH)D levels are presented in table 2. The SNP rs2282679 had the highest F-statistic and R^2-value among the three SNPs, showing 4.0 nmol·L^{-1} increase in 25(OH)D per effect allele. The 25(OH)D allele score had a F-statistic of 197 and

TABLE 2 Coefficient, F-statistic and R^2-value of linear regression between single nucleotide polymorphism/allele score and season-standardised 25-hydroxyvitamin D (25(OH)D) level (nmol·L^{-1}) in the subcohort of the HUNT2 study[#]

	Coefficient[¶] (95% CI)	p-value	F-statistic	R^2-value
rs2282679	4.00 (3.30–4.71)	1.63×10^{-28}	124	0.022
rs12785878	1.92 (1.26–2.59)	1.36×10^{-08}	32	0.006
rs10741657	2.48 (1.84–3.13)	6.98×10^{-14}	56	0.010
Allele score[+]	2.74 (2.36–3.12)	5.15×10^{-44}	197	0.034

[#]: n=5546; [¶]: change of season-standardised 25(OH)D (nmol·L^{-1}) per effect allele or per allele score; [+]: sum of the number of effect alleles of rs2282679, rs12785878 and rs10741657.

TABLE 3 Associations between the 25-hydroxyvitamin D allele score and risk of lung cancer overall and histologic types in the HUNT2 study[#]

	Cases n	Rate per 100 000 person-years[¶]	HR[+] (95% CI)	p-value
Lung cancer overall	676	73.8	0.99 (0.93–1.06)	0.83
Small cell lung cancer	90	9.8	0.92 (0.77–1.10)	0.35
Adenocarcinoma	189	20.6	0.92 (0.82–1.04)	0.20
Squamous cell carcinoma	141	15.4	0.98 (0.85–1.14)	0.82
Other/unknown subtypes	256[§]	27.9	1.08 (0.97–1.21)	0.14

[#]: n=54 580; [¶]: time at risk for outcomes: 916 480 person-years; [+]: per vitamin D-increasing allele score; [§]: among which 87 cases were other subtypes.

accounted for 3.4% of the variation of serum 25(OH)D levels. The associations between the allele score and the potential confounders are presented in supplementary table S2. In general, taking account of multiple testing, there were no clear associations observed.

Table 3 shows that the 25(OH)D allele score was not associated with the incidence of lung cancer overall, with HR 0.99 (95% CI 0.93–1.06) per allele score. There was no clear association between the allele score and risk of any histologic type of lung cancer. Based on Mendelian randomisation estimates using either the IVW method or weighted median method, there was little evidence that genetically determined season-standardised 25(OH)D was associated with risk of lung cancer overall or any histologic type (table 4 and figure 1). Using the IVW method, the Mendelian randomisation estimate HR for lung cancer overall was 0.96 (95% CI 0.54–1.69) per 25 nmol·L^{-1} increase in the genetically determined 25(OH)D level.

As shown in table 4, I^2 and Cochran's Q-statistic showed no evidence for heterogeneity between the SNPs (I^2 0.00, 95% CI 0.00–0.24; p=0.87 for lung cancer overall). The p-value of the intercept by the MR-Egger method was 0.79 (intercept −0.03, 95% CI −0.25–0.19) for lung cancer overall, suggesting no substantial pleiotropic effect of these SNPs (table 5).

Mendelian randomisation estimates of a 10% increase in genetically determined 25(OH)D level with risks of lung cancer and histologic types in a two-sample Mendelian randomisation are presented in supplementary tables S3 and S4 and supplementary figure S1 as sensitivity analyses. All estimates in the two-sample Mendelian randomisation were similar to those derived from the primary analyses.

Discussion
Main findings
In this Mendelian randomisation analysis of a population-based prospective cohort study including 54 580 subjects, we found no substantial evidence of a causal association of serum 25(OH)D level with the incidence of lung cancer overall, small cell lung cancer, adenocarcinoma or squamous cell carcinoma.

Comparison with other studies
The finding of the current Mendelian randomisation study is inconsistent with the conclusion from two meta-analyses of observational studies [9, 10]. The results of the meta-analyses may be largely driven by the inclusion of a large cohort study showing an inverse association between 25(OH)D levels and incidence of

TABLE 4 Mendelian randomisation estimates of the associations between a 25 nmol·L^{-1} increase in genetically determined season-standardised 25-hydroxyvitamin D and risk of lung cancer overall and histologic types in the HUNT2 study[#]

|	Inverse-variance weighted method				Weighted median method	
	HR (95% CI)	p-value	I^2 (95% CI)	p-value of Q-statistic	HR (95% CI)	p-value
Lung cancer overall	0.96 (0.54–1.69)	0.88	0.00 (0.00–0.24)	0.87	1.00 (0.54–1.85)	0.99
Small cell lung cancer	0.58 (0.12–2.69)	0.48	0.45 (0.00–0.84)	0.16	0.82 (0.14–4.76)	0.82
Adenocarcinoma	0.54 (0.18–1.57)	0.26	0.00 (0.00–0.62)	0.76	0.61 (0.19–1.95)	0.41
Squamous cell carcinoma	0.64 (0.19–2.17)	0.47	0.08 (0.00–0.90)	0.34	0.64 (0.17–2.44)	0.52
Other/unknown subtypes	2.22 (0.86–5.75)	0.10	0.00 (0.00–0.72)	0.69	2.69 (0.94–7.71)	0.07

[#]: n=54 580.
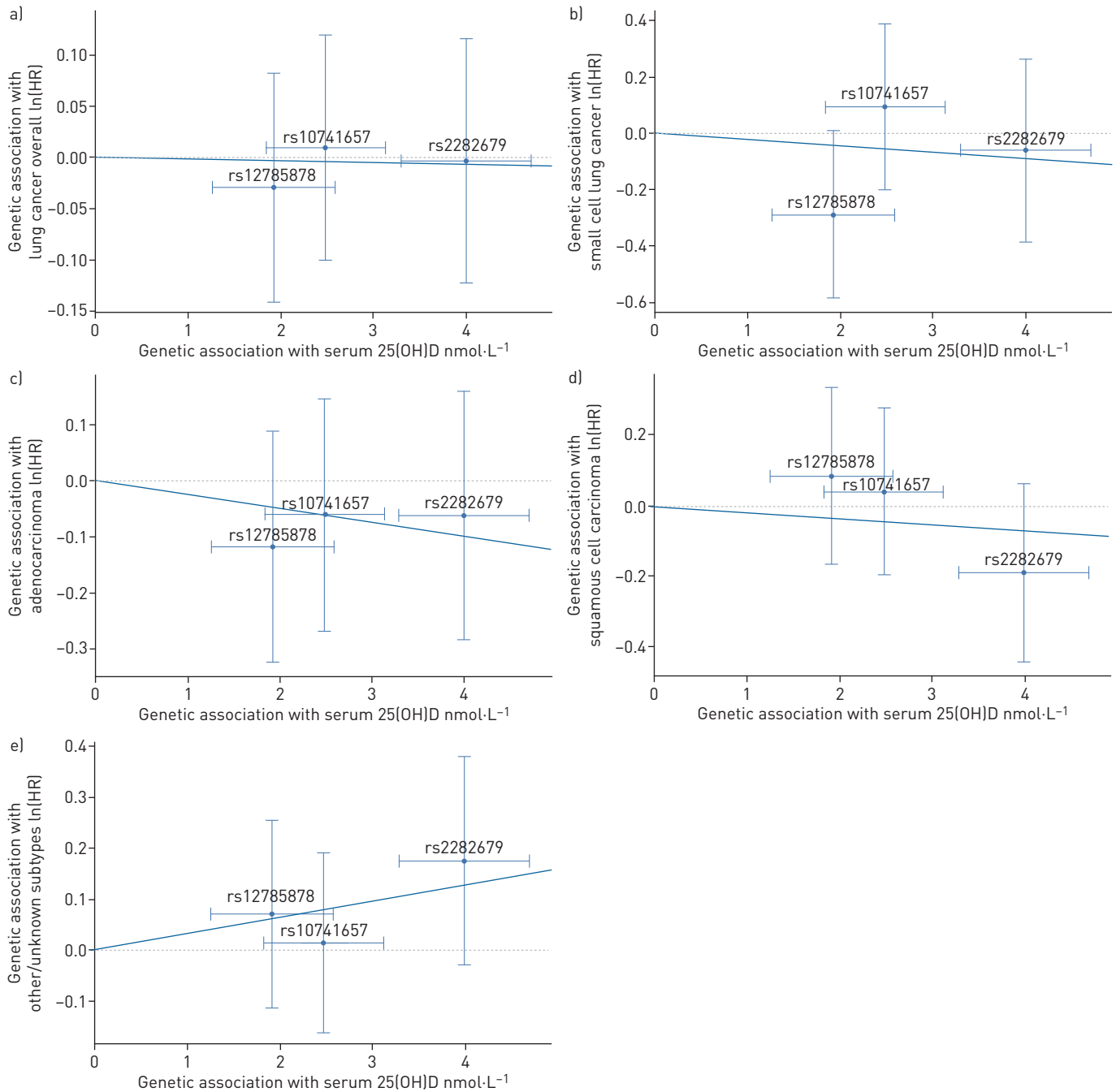
FIGURE 1 Effect of single nucleotide polymorphism (SNP)-determined serum 25-hydroxyvitamin D (25(OH)D) on the risk of lung cancer overall and histologic types in the HUNT2 study (n=54 580): a) lung cancer overall, b) small cell lung cancer, c) adenocarcinoma, d) squamous cell carcinoma and e) other/unknown subtypes. The ln(hazard ratio) for risk of outcome of each SNP was plotted against each SNP's effect on increase of serum 25(OH)D (nmol·L^{-1}). Each trend line, derived from the inverse-variance weighted method and set through the origin of the axes arbitrarily, represents ln(hazard ratio) for the risk of lung cancer or histologic type per nmol·L^{-1} increase in genetically determined serum 25(OH)D. Vertical and horizontal lines around points show 95% confidence intervals for each SNP.

lung cancer [29], whereas others showed no association [9, 10]. The current study is also inconsistent with results from our own observational study from the same cohort showing that lower 25(OH)D levels were associated with a lower risk of adenocarcinoma, particularly in obese individuals [30]. The present Mendelian randomisation analysis conformed to our speculation that residual confounding by adiposity or adiposity-related factors could have biased the observational results [30].

Few Mendelian randomisation studies have explored the potential causal association between circulating vitamin D levels and lung cancer risk. Our findings are consistent with those of the study by

TABLE 5 MR-Egger pleiotropy test of associations between a 25 nmol·L^{-1} increase in genetically determined season-standardised 25-hydroxyvitamin D and risk of lung cancer overall and histologic types in the HUNT2 study[#]

	Intercept (95% CI)	p-value
Lung cancer overall	−0.03 [−0.25–0.19]	0.79
Small cell lung cancer	−0.27 [−1.27–0.72]	0.59
Adenocarcinoma	−0.14 [−0.55–0.27]	0.50
Squamous cell carcinoma	0.36 [−0.12–0.84]	0.14
Other/unknown subtypes	−0.09 [−0.45–0.28]	0.64

[#]: n=54 580.

DIMITRAKOPOULOU *et al.* [31] who used summarised data of a consortium (Transdisciplinary Research in Cancer of the Lung–International Lung Cancer Consortium (TRICL-ILCCO)) including a large number of lung cancer cases (n=12 537). Having assumed sufficient statistical power to detect moderate effects, they found no causal association between circulating vitamin D concentration and risk of lung cancer and certain histologic types (adenocarcinoma and squamous cell carcinoma) [31]. Although the conclusions are the same, our study differs from the study by DIMITRAKOPOULOU *et al.* [31] in study design. Our study investigated the association in a homogeneous population-based prospective cohort study with a long follow-up duration, while TRICL-ILCCO mainly consisted of case–control studies from different geographical areas and ethnicities (http://ilcco.iarc.fr). Selection bias is more likely in case–control studies than in prospective cohort studies and survivor bias in Mendelian randomisation studies has recently been discussed as a methodological issue [32]. Nevertheless, we need to note that the causal effect estimates in Mendelian randomisation studies generally reflect a lifetime risk, regardless of the follow-up time.

Strengths and limitations

The current study is one of the first Mendelian randomisation analyses using a long-term prospective population-based study to investigate the associations of serum 25(OH)D levels with the risk of lung cancer and histologic types. Information about diagnosis of lung cancer at the Cancer Registry of Norway is nearly complete and reasonably accurate [33]. However, there were many cases with unknown subtypes, resulting in limited statistical power in the analyses of histologic types.

Compared with observational studies, Mendelian randomisation studies are not vulnerable to reverse causation and unmeasured confounding when the assumptions of the studies are satisfied. We performed both one-sample Mendelian randomisation in which we measured serum 25(OH)D levels in a reasonably large subcohort and two-sample Mendelian randomisation as a sensitivity analysis. Even though two-sample Mendelian randomisation is becoming common with the access to MR-Base (www.mrbase.org), one-sample Mendelian randomisation still has its advantages, such as testing the important assumptions of Mendelian randomisation directly. F-statistics and R^2-values from the regression of the SNP/allele score on 25(OH)D levels indicated sufficient strength of the instrumental variables of the exposure in the current study. The variation of 25(OH)D explained by the three SNPs used in the present study was larger (3.4% *versus* 1.9%) than that explained by four SNPs in the study by VIMALESWARAN *et al.* [23]. We could also investigate the associations between SNP/allele score with a broad range of measured and reported characteristics at baseline. Even though the instruments may still be associated with unmeasured confounding, they were not associated with important confounders such as smoking and socioeconomic status in HUNT2. The last important assumption in Mendelian randomisation is that the instrument (SNP or allele score) should be associated with the outcome of interest (lung cancer) only *via* the exposure (circulating vitamin D levels). We found no violation of this essential assumption according to MR-Egger tests, but fewer genetic instruments may have a relatively low power to detect horizontal pleiotropy [28].

This study had several potential limitations. Nonparticipation in HUNT2 was ∼30% of the population. As participants in the HUNT studies were shown to be healthier than nonparticipants, our findings might differ to some degree from the true situation in the general Norwegian population [34]. The applied Mendelian randomisation analysis was based on the assumption that the exposure–outcome relationship is a linear, dose–response relationship [11]. We were not able to investigate the nonlinear association in Mendelian randomisation due to the lack of methods for binary outcomes [35], whereas many of the reported associations between circulating 25(OH)D levels and health outcomes were nonlinear [30, 36]. The sample size of this study was likely insufficient to reveal a weak-to-moderate effect of vitamin D on lung cancer risk based on the wide confidence intervals of the Mendelian randomisation estimates, but the

sample size of our cohort seemed adequate to detect risk factors with large effects on lung cancer, such as smoking (supplementary table S5). In addition, our results were consistent with the null findings of the aforementioned Mendelian randomisation study that reported sufficient study power of a case–control design [31]. Nevertheless, consortia consisting of data from European population-based prospective studies with long follow-up duration are warranted to further investigate the causality of vitamin D on lung cancer in Mendelian randomisation analysis. Mendelian randomisation studies are also called for in Asia and the Middle East where populations are reported to have lower vitamin D levels than populations from Europe [37]. In addition, results from ongoing large clinical trials are expected to clarify the causal association of vitamin D with cancer and other adverse outcomes, especially in individuals with low vitamin D status prior to intervention [38, 39].

Conclusions

In summary, Mendelian randomisation analysis indicated that serum 25(OH)D levels were not causally associated with the risk of lung cancer overall or histologic types in a population-based prospective cohort study.

References

1   Deeb KK, Trump DL, Johnson CS. Vitamin D signalling pathways in cancer: potential for anticancer therapeutics. *Nat Rev Cancer* 2007; 7: 684–700.

2   Feldman D, Krishnan AV, Swami S, *et al.* The role of vitamin D in reducing cancer risk and progression. *Nat Rev Cancer* 2014; 14: 342–357.

3   Giovannucci E, Liu Y, Rimm EB, *et al.* Prospective study of predictors of vitamin D status and cancer incidence and mortality in men. *J Natl Cancer Inst* 2006; 98: 451–459.

4   Jacobs ET, Kohler LN, Kunihiro AG, *et al.* Vitamin D and colorectal, breast, and prostate cancers: a review of the epidemiological evidence. *J Cancer* 2016; 7: 232–240.

5   International Agency for Research on Cancer. GLOBOCAN 2012: Estimated Cancer Incidence, Mortality and Prevalence Worldwide in 2012. 2012. http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx Date last accessed: April 27, 2018.

6   International Agency for Research on Cancer. Pathology and Genetics of Tumours of the Lung, Pleura, Thymus and Heart. Lyon, IARC Press, 2004.

7   Cheng TY, Cramb SM, Baade PD, *et al.* The international epidemiology of lung cancer: latest trends, disparities, and tumor characteristics. *J Thorac Oncol* 2016; 11: 1653–1671.

8   Lee PN, Forey BA, Coombs KJ. Systematic review with meta-analysis of the epidemiological evidence in the 1900s relating smoking to lung cancer. *BMC Cancer* 2012; 12: 385.

9   Liu J, Dong Y, Lu C, *et al.* Meta-analysis of the correlation between vitamin D and lung cancer risk and outcomes. *Oncotarget* 2017; 8: 81040–81051.

10 Zhang L, Wang S, Che X, *et al.* Vitamin D and lung cancer risk: a comprehensive review and meta-analysis. *Cell Physiol Biochem* 2015; 36: 299–305.

11 Lawlor DA, Harbord RM, Sterne JA, *et al.* Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat Med* 2008; 27: 1133–1163.

12 Lawlor DA. Commentary: Two-sample Mendelian randomization: opportunities and challenges. *Int J Epidemiol* 2016; 45: 908–915.

13 International Agency for Research on Cancer. International Classification of Diseases for Oncology (ICD-O). 3rd Edn. Lyon, IARC Press, 2013.

14   Holick MF. Vitamin D deficiency. *N Engl J Med* 2007; 357: 266–281.

15   Zerwekh JE. Blood biomarkers of vitamin D status. *Am J Clin Nutr* 2008; 87: 1087s–1091s.

16   Degerud E, Hoff R, Nygard O, *et al.* Cosinor modelling of seasonal variation in 25-hydroxyvitamin D concentrations in cardiovascular patients in Norway. *Eur J Clin Nutr* 2016; 70: 517–522.

17   Ferreira MA, Vonk JM, Baurecht H, *et al.* Shared genetic origin of asthma, hay fever and eczema elucidates allergic disease biology. *Nat Genet* 2017; 49: 1752–1757.

18   Das S, Forer L, Schonherr S, *et al.* Next-generation genotype imputation service and methods. *Nat Genet* 2016; 48: 1284–1287.

19   McCarthy S, Das S, Kretzschmar W, *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* 2016; 48: 1279–1283.

20   Zhou W, Fritsche LG, Das S, *et al.* Improving power of association tests using multiple sets of imputed genotypes from distributed reference panels. *Genet Epidemiol* 2017; 41: 744–755.

21   Ahn J, Yu K, Stolzenberg-Solomon R, *et al.* Genome-wide association study of circulating vitamin D levels. *Hum Mol Genet* 2010; 19: 2739–2745.

22   Wang TJ, Zhang F, Richards JB, *et al.* Common genetic determinants of vitamin D insufficiency: a genome-wide association study. *Lancet* 2010; 376: 180–188.

23   Vimaleswaran KS, Berry DJ, Lu C, *et al.* Causal relationship between obesity and vitamin D status: bi-directional Mendelian randomization analysis of multiple cohorts. *PLoS Med* 2013; 10: e1001383.

24   Burgess S, Thompson SG. Use of allele scores as instrumental variables for Mendelian randomization. *Int J Epidemiol* 2013; 42: 1134–1144.

25   Johnson AD, Handsaker RE, Pulit SL, *et al.* SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics* 2008; 24: 2938–2939.

26   Yavorska OO, Burgess S. MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. *Int J Epidemiol* 2017; 46: 1734–1739.

27   Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol* 2013; 37: 658–665.

28   Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol* 2015; 44: 512–525.

29   Afzal S, Bojesen SE, Nordestgaard BG. Low plasma 25-hydroxyvitamin D and risk of tobacco-related cancer. *Clin Chem* 2013; 59: 771–780.

30   Sun YQ, Langhammer A, Wu C, *et al.* Associations of serum 25-hydroxyvitamin D level with incidence of lung cancer and histologic types in Norwegian adults: a case-cohort analysis of the HUNT study. *Eur J Epidemiol* 2017; 59: 771–780.

31   Dimitrakopoulou VI, Tsilidis KK, Haycock PC, *et al.* Circulating vitamin D concentration and risk of seven cancers: Mendelian randomisation study. *BMJ* 2017; 359: j4761.

32   Vansteelandt S, Dukes O, Martinussen T. Survivor bias in Mendelian randomization analysis. *Biostatistics* 2017; in press [https://doi.org/10.1093/biostatistics/kxx050].

33   Larsen IK, Smastuen M, Johannesen TB, *et al.* Data quality at the Cancer Registry of Norway: an overview of comparability, completeness, validity and timeliness. *Eur J Cancer* 2009; 45: 1218–1231.

34   Langhammer A, Krokstad S, Romundstad P, *et al.* The HUNT study: participation is associated with survival and depends on socioeconomic status, diseases and symptoms. *BMC Med Res Methodol* 2012; 12: 143.

35   Staley JR, Burgess S. Semiparametric methods for estimation of a nonlinear exposure-outcome relationship using instrumental variables with application to Mendelian randomization. *Genet Epidemiol* 2017; 41: 341–352.

36   Sun YQ, Langhammer A, Skorpen F, *et al.* Serum 25-hydroxyvitamin D level, chronic diseases and all-cause mortality in a population-based prospective cohort: the HUNT Study, Norway. *BMJ Open* 2017; 7: e017256.

37   van Schoor NM, Lips P. Worldwide vitamin D status. *Best Pract Res Clin Endocrinol Metab* 2011; 25: 671–680.

38   Kupferschmidt K. Uncertain verdict as vitamin D goes on trial. *Science* 2012; 337: 1476–1478.

39   Pilz S, Rutters F, Dekker JM. Disease prevention: vitamin D trials. *Science* 2012; 338: 883.