Article type: Original Article

Title: No association between circulating concentrations of vitamin D and risk of lung cancer: An analysis in 20 prospective studies in the Lung Cancer Cohort Consortium (LC3)

Author List: D.C. Muller[1,2*], A.M. Hodge[3,4*], A. Fanidi[1,5], D. Albanes[6], XM. Mai[7], XO. Shu[8], S.J. Weinstein[6], T.L. Larose[1,9], X. Zhang[10], J. Han[11,12], M.J. Stampfer[10,13,14], S.A. Smith-Warner[13,14], J. Ma[10], J.M. Gaziano[15,16,17], H.D. Sesso[13,15,16], V.L. Stevens[18], M.L. McCullough[18], T.M. Layne[6], R. Prentice[19], M. Pettinger[19], C.A. Thomson[20], W. Zheng[8], YT. Gao[21], N. Rothman[6], YB. Xiang[22], H. Cai[23], R. Wang[24], JM. Yuan[24,25], WP. Koh[26], L.M. Butler[25,27], Q. Cai[23], W.J. Blot[23], J. Wu[23], PM. Ueland[28,29], Ø. Midttun[30], A. Langhammer[31], K. Hveem[9,31], M. Johansson[32], J. Hultdin[33], K. Grankvist[33], A. A. Arslan[34,35], L. Le Marchand[36], G. Severi[3,37,38]#, M. Johansson[1]#, P. Brennan[1]#

* denotes co-first authors

# denotes co-last authors

1. Genetic Epidemiology Group, International Agency for Research on Cancer, Lyon, France
2. Department of Epidemiology and Biostatistics, Imperial College London, London, United Kingdom
3. Cancer Epidemiology Center, Cancer Council Victoria, Melbourne, Australia
4. Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, University of Melbourne, Parkville, Australia
5. MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Cambridge, United Kingdom

6. Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Bathesda, USA
7. Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, NTNU, Norwegian University of Science and Technology, Trondheim, Norway
8. Division of Epidemiology, Department of Medicine, Vanderbilt University School of Medicine, Nashville, USA
9. K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health & Nursing, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Trondheim, Norway
10. Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, USA
11. Department of Epidemiology, Richard M. Fairbanks School of Public Health, Indiana University, Indianapolis, USA
12. Melvin and Bren Simon Cancer Center, Indiana University, Indianapolis, USA
13. Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, USA
14. Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, USA
15. Division of Aging, Department of Medicine, Brigham and Women's Hospital, Boston, USA
16. Division of Preventive Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, USA
17. Boston VA Medical Center, Boston, USA
18. Epidemiology Research Program, American Cancer Society, Atlanta, USA
19. Division of Public Health Sciences Fred Hutchinson Cancer Research Center, Seattle, USA
20. Health Promotion Sciences, Mel and Enid Zuckerman College of Public Health, University of Arizona, Tucson, USA
21. Department of Epidemiology, Shanghai Cancer Institute, Shanghai Jiaotong University, Shanghai, China

22. State Key Laboratory of Oncogene and Related Genes & Department of Epidemiology, Shanghai Cancer Institute, Renji Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China
23. Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, USA
24. UPMC Hillman Cancer Center, University of Pittsburgh, Pittsburgh, USA
25. Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, USA
26. Duke-NUS Graduate Medical School Singapore, Singapore, Singapore
27. Division of Cancer Control and Population Sciences, University of Pittsburgh Cancer Institute, Pittsburgh, USA
28. Laboratory of Clinical Biochemistry, Department of Clinical Science, University of Bergen, Bergen, Norway
29. Haukeland University Hospital, Bergen, Norway
30. Bevital AS, Bergen, Norway
31. HUNT Research Centre, Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, NTNU, Norwegian University of Science and Technology, Levanger, Norway
32. Department of Radiation Sciences, Oncology, Umeå University, Umeå, Sweden
33. Department of Medical Biosciences, Clinical Chemistry, Umeå University, Umeå, Sweden
34. Departments of Obstetrics and Gynecology, Population Health and Environmental Medicine, New York University School of Medicine, New York, USA
35. Department of Population Health and Environmental Medicine, New York University School of Medicine, New York, USA
36. Epidemiology Program, Cancer Research Center of Hawaii, University of Hawaii
37. Italian Institute for Genomic Medicine (IIGM), Torino, Italy.

38. Centre de Recherche en Epidemiologie et Santé des Populations (CESP) UMR1018 Inserm, Facultés de Médicine Université Paris-Saclay, Villejuif, France.

Corresponding Authors:

Dr. Paul Brennan, Genetic Epidemiology Group, International Agency for Research on Cancer (IARC-WHO), 150 cours Albert Thomas, Lyon, 69008, France, Tel: +33 4 72 73 85 33, Email: gep@iarc.fr

AND

Dr. David C. Muller, Department of Epidemiology and Biostatistics, Imperial College London, Norfolk Place, London W21PG, United Kingdom, Tel: +44 (0)20 7594 0856, Email: david.muller@imperial.ac.uk

Abstract

Background: There is observational evidence suggesting that high vitamin D concentrations may protect against lung cancer. To investigate this hypothesis in detail, we measured circulating vitamin D concentrations in pre-diagnostic blood from 20 cohorts participating in the Lung Cancer Cohort Consortium (LC3).

Patients and methods: The study included 5,313 lung cancer cases and 5,313 controls selected from. Blood samples for the cases were collected, on average, 5 years prior to lung cancer diagnosis. Controls were individually matched to the cases by cohort, sex, age, race/ethnicity, date of blood collection, and smoking status in 5 categories. Liquid chromatography coupled with tandem mass spectrometry was used to separately analyze 25-hydroxyvitamin D_2 ($25(OH)D_2$) and 25-hydroxyvitamin D_3 ($25(OH)D_3$) and their concentrations were combined to give an overall measure of 25(OH)D. We used conditional logistic regression to calculate odds ratios (OR) and 95% confidence intervals (CI) for 25(OH)D as both a continuous and categorical variable.

Results: Overall, no apparent association between 25(OH)D and risk of lung cancer was observed (multivariable adjusted OR for a doubling in concentration: 0.98, 95% confidence interval: 0.91, 1.06). Similarly, we found no clear evidence of interaction by cohort, sex, age, smoking status, or histology.

Conclusion: This study did not support an association between vitamin D concentrations and lung cancer risk.

Keywords: serum 25-hydroxyvitamin D, vitamin D, lung cancer, case-control, prospective, consortium

Key Message: Results from this prospective study of 20 international cohorts show no association between circulating concentrations of vitamin D and risk of lung cancer.

Introduction

Lung cancer is the most common cause of cancer death worldwide, accounting for over 20% of all cancer deaths (1). Although avoidance of tobacco consumption remains the most important strategy for lung cancer prevention, a substantial proportion of lung cancer cases occur among former smokers, especially in countries where tobacco cessation campaigns have been successful. Identifying additional strategies to reduce lung cancer risk would be of particular relevance for high risk groups such as former smokers, and could have a substantial public health impact.

There is evidence that higher circulating vitamin D concentrations may reduce the risk of developing lung cancer (2,3). Vitamin D is produced primarily in the skin after exposure to ultraviolet B (UVB) radiation from sunlight. A secondary source of vitamin D is through diet or via dietary supplements (4). Vitamin D is subsequently hydroxylated in the liver to form 25-hydroxyvitamin D (25(OH)D), the major circulating metabolite of vitamin D, which is then converted into its active form (1,25-dihydroxyvitamin D) in the kidneys and other tissues (5). The necessity of adequate concentrations of vitamin D for bone health is well known (6), and vitamin D has also been implicated in the

development of several cancers, particularly colorectal cancer (7,8). As reported in two meta-analyses, a number of prospective cohort and nested case-control studies have examined vitamin D concentrations, measured as 25-hydroxyvitamin D (25(OH)D) in blood samples, in terms of risk of lung cancer. In the first meta-analysis, pooled results from 9 prospective studies reported a 17% decrease in lung cancer risk between participants with 'high' and 'low' concentrations of 25(OH)D (2). The second meta-analysis reported similar results, although ten of the studies were included in the first meta-analysis (2). Pooled results from the second meta-analysis showed a nonlinear inverse association between 25(OH)D and lung cancer risk (3).

While these meta-analyses are informative, they are clearly constrained by the manner in which results are reported in the literature, and cannot account for variation that may occur between studies, including use of different laboratories for measurement of vitamin D. In order to provide a more definitive answer to the role of circulating vitamin D and risk of lung cancer, we conducted a pooled analysis of 25(OH)D concentrations in 5,313 case-control pairs from 20 cohorts within the Lung Cancer Cohort Consortium (LC3).

Methods

The LC3 was established in 2009 as a project within the US National Cancer Institute (NCI) Cohort Consortium as a coordinated large-scale effort to study the role of B-vitamins in lung cancer (9). Inclusion criteria for participating cohorts included the occurrence of at least 200 incident lung cancer cases

with baseline questionnaire data and either plasma or serum samples – typically cryopreserved at <80°C. Twenty cohorts fulfilled those criteria and agreed to participate in the LC3, resulting in a combined cohort population of over 2,000,000 participants from Asia, Australia, Europe, and North America. All cohorts are listed in Supplementary Table 1, along with the numbers of cases and controls contributed. Brief details on design of the cohorts and their follow-up procedures are also provided in the supplementary materials.

Selection of cases and controls

Lung cancer cases were defined on the basis of the International Classification of Diseases for Oncology, Second Edition (ICD-O-2), and included all invasive cancers coded as C34.0-C34-9. Altogether, 11,399 incident lung cancer cases with pre-diagnostic blood samples were identified from the participating cohorts. We selected a total of 5,545 lung cancer cases for subsequent blood based analysis, and to optimize the statistical power in smoking stratified risk analysis, never and former smoking cases were oversampled. For each case, one control was randomly chosen from risk-sets consisting of all cohort members alive and free of cancer (except non-melanoma skin cancer) at the time of diagnosis of the index case. Matching criteria were cohort, sex, date of blood collection (± 1 month, relaxed to ± 3 months for sets without available controls), and date of birth (± 1 year, relaxed to ± 3 years for sets without available controls), as well as smoking status in 5 categories: never smokers, short and long-term quitters among former smokers (<10 years, ≥10 years since quitting), and light and heavy smokers among current smokers (< 15, ≥15 cigarettes per day). After excluding cases

who were not correctly matched on smoking status (*n*=124 cases), who had insufficient blood samples (*n*=42), or had a revised date of diagnosis prior to blood draw (*n*=13), a total of 5,364 lung cancer case-control pairs remained eligible for analysis. Further exclusion of 51 participants whose matched pair was missing a 25(OH)D value resulted in a final set of 5,313 matched case-control pairs.

Biochemical analyses

Vitamin D analyses were conducted as part of a coordinated laboratory analysis along with a panel of 40 other biomarkers, focused primarily on B-vitamins and other biomarkers that are involved in the one carbon metabolism pathway (9). Blood samples from all cases and controls were sent on dry ice to the Bevital AS laboratory (Bergen, Norway, www.bevital.no). Liquid chromatography coupled with tandem mass spectrometry was used to separately analyze 25-hydroxyvitamin D_2 (25(OH)D_2) and 25-hydroxyvitamin D_3 (25(OH)D_3) (11). The limit of detection was 3.3 nmol/L, and within-day and between-day coefficients of variation ranged from 4.4%–8.2%. Circulating cotinine was also assessed with liquid chromatography coupled with tandem mass spectrometry. The limit of detection was 1 nmol/L, and the within-day and between-day coefficients of variation ranged from 2%–6%. The laboratory is Vitamin D External Quality Assessment Scheme–certified (DEQAS, London, United Kingdom, www.deqas.org).

Statistical Analysis

We used conditional logistic regression to calculate odds ratios (ORs) and 95% confidence intervals (CIs) for 25(OH)D as both a continuous and categorical variable. For categorical analyses, 25(OH)D was grouped according to the quartiles of its distribution among control participants, as well as an alternative grouping with extreme and moderate deficiency (<25 nmol/L) as the reference category. For continuous analyses, 25(OH)D was $\log_2$ transformed, so ORs correspond to the expected fold-change in odds of lung cancer for a doubling in 25(OH)D concentration. All models were adjusted for circulating cotinine in four groups, defined by quartiles of cotinine among current smoking participants. We investigated potential interactions between 25(OH)D and sex, smoking status, cohort region, age at baseline, body mass index (BMI, kg/m^2), time between blood draw and lung cancer diagnosis, as well as histological subtype. These analyses were conducted using R version 3.3.2 (12).

Because $25(OH)D_3$ is strongly affected by exposure to UVB radiation, we used season-adjusted concentrations. We modeled seasonal variation of log transformed $25(OH)D_3$ using two pairs of sine and cosine functions of day of blood collection in a hierarchical linear regression model, allowing model parameters to vary by cohort. The use of trigonometric functions allows estimation of the periodic variation in $25(OH)D_3$ and produces smooth predictions with no artificial discontinuities from season to season, or year to year. We included 2 pairs of sine and cosine functions in the models because the inclusion of additional terms did not improve model fit, nor did it substantially affect parameter estimates in the final models. For each participant, season-adjusted 25(OH)D was calculated by adding or subtracting

the expected seasonal deviation from the mean 25(OH)D_3 concentration based on the day of blood draw, and adding the 25(OH)D_2 concentration. These Bayesian hierarchical models were fitted using Stan version 2.9.1 via Rstan (13,14).

Results

Demographic and baseline characteristics of the 5,313 case-control pairs are presented in Table 1, both overall and by geographic region. Participants from the US had higher circulating concentrations of total 25(OH)D on average compared with participants from Asia, or Europe and Australia. Table 2 shows the histological subtype of lung cancer cases, as well as the distribution of time between baseline blood collection and diagnosis.

Figure 1 shows the observed 25(OH)D concentrations by day of blood draw, as well as the model-estimated mean concentrations, for each cohort separately. Figure 2 shows the season-adjusted and observed 25(OH)D concentrations. ORs for risk of lung cancer by category of 25(OH)D concentration, as well as for a doubling in concentration, are presented in Table 3. No apparent dose-response association was observed between 25(OH)D and risk of lung cancer overall (OR for a doubling in concentration: 0.98, 95% CI: 0.91, 1.06), nor for specific categories of 25(OH)D concentration. Further, we found little evidence of interaction by any demographic factor or individual participant characteristic (Figure 3). There was some evidence of an inverse association between 25(OH)D and risk of lung cancer in the European cohorts (OR for a doubling in concentration 0.75, 95% CI: 0.61, 0.92), but statistical evidence for an interaction was weak (p =

0.05). Cohort-specific estimates suggest that any association among European participants was driven by MDCS and NSHDC, the two Swedish cohorts), although we found no overall evidence of interaction by cohort (Figure 4). Sensitivity analyses in which we excluded circulating cotinine from the models provided similar results to the fully adjusted models (Supplementary Figure 1). Using $25(OH)D_3$ only also yielded similar results (Supplementary Figure 2).

Discussion

Based on a comprehensive analysis of over 5,000 case-control pairs from 20 prospective cohort studies, we found no association between circulating vitamin D concentrations and risk of subsequent lung cancer diagnosis.

Our results contrast with the two previous meta-analyses that reported inverse associations between circulating vitamin D concentrations and lung cancer risk (2,3). As always with meta-analyses, it is possible that the collection of published studies does not represent the total sum of all conducted studies, whereby studies with null results are more likely to go unpublished. Although neither study reported evidence for publication bias, we note that the results were strongly influenced by one study from Copenhagen with approximately 10,000 participants and 507 incident lung cancer cases with 25-year follow-up (15), accounting for 65% of the association reported by Zhang et al (2). The second meta-analysis by Chen et al. was based on 10 cohorts (3), but there was considerable overlap between included studies in both the Chan and Zhang meta-analyses, and both included the aforementioned Copenhagen cohort study (15). It should therefore be no surprise that the pooled estimates

across these two meta-analyses suggest a similar inverse association with risk.

Median 25(OH)D concentrations in the Copenhagen cohort were 41 nmol/L overall, and 37 nmol/L among those who went on to develop a tobacco related cancer, similar to approximately the first fourth of 25(OH)D for the LC3 cohorts overall. It is of interest that Afzal et al. also observed significant associations with vitamin D for other tobacco related cancers including head and neck cancer, bladder and kidney cancer in the Copenhagen cohort (15). Given the limited adjustment for tobacco use in the Copenhagen cohort (i.e., pack-years only), and the strong association between vitamin D concentrations and smoking intensity, this would suggest the possibility that residual confounding by smoking may explain at least some of the observed inverse associations of vitamin D concentrations with lung cancer risk, as well as with other tobacco related cancers in that study (15). In our study, estimates were similar in models that were not adjusted for circulating cotinine, indicating that residual confounding is unlikely to fully explain these discrepant results. Further, the estimates from the Copenhagen cohort are consistent with the estimates that we present for the two cohorts from Sweden. The absence of evidence supporting an interaction by cohort, however, suggests that any observed inverse associations in our study are likely to be consistent with sampling variability.

Our results, based on over 5,000 incident lung cancer cases and 5,000 individually matched controls, clearly point to a lack of association for risk of

lung cancer across a broad range of vitamin D concentrations. Our study has several strengths including centralized biochemical analysis of pre-diagnostic vitamin D and detailed control for tobacco exposure among current smokers using cotinine concentrations, an objective measure of recent tobacco exposure. There is a lack of consensus surrounding terminology and vitamin D concentrations that are considered adequate or deficient. The 2011 US Institute of Medicine report considers individuals with concentrations below 30 nmol/L at risk of deficiency, and those below 50 nmol/L at risk of inadequacy (16). An alternative definition considers concentrations below 50 nmol/L as deficient, with further subclassifications of mild (25-50 nmol/L), moderate (12.5-25 nmol/L) and severe (<12.5 nmol/L) deficiency (17). Our reference exposure group included participants with vitamin D concentrations between 7 and 41 nmol/L, and therefore included a broad range of persons with vitamin D deficiency or inadequacy. Even when we further restricted the reference category to include only those participants with vitamin D concentrations below 25 nmol/L, we found no evidence for an association for adequate vitamin D concentrations with lower lung cancer risk.

In summary, we found no overall evidence for an association between circulating concentrations of vitamin D and risk of lung cancer in 20 prospective cohort studies from the Asia, Australia and Europe, and the US, but some suggestion of an inverse association in two Swedish cohorts. In light of this, we consider that vitamin D supplementation is unlikely to prove broadly effective for the primary prevention of lung cancer. Ongoing cancer prevention trials testing vitamin D supplements may eventually provide

additional evidence on whether or not increases in vitamin D concentrations translate to reductions in lung cancer risks or not.

Funding Acknowledgements

The Lung Cancer Cohort Consortium (LC3) was supported by Grant NIH / NCI (n° 1U01CA155340-01) and NHMRC (Grant ID: 1050198). DCM was supported by an IARC/Australia Fellowship funded by Cancer Council Australia, and a Cancer Research UK Population Research Fellowship. TLL was supported by The Research Council of Norway (grant number 267776/H10). The work of TLL presented in this paper was undertaken during a postdoctoral placement at the International Agency for Research on Cancer, within the framework of an agreement between the Research Council of Norway and the Norwegian University of Science and Technology. The funding organizations had no role in design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript.

Cohort specific acknowledgements:

The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts HHSN268201600018C, HHSN268201600001C, HHSN268201600002C, HHSN268201600003C, and HHSN268201600004C. The authors thank the WHI investigators and staff for their dedication, and the

study participants for making the program possible. A full listing of WHI investigators can be found at: http://www.whi.org/researchers/Documents%20%20Write%20a%20Paper/WHI%20Investigator%20Long%20List.pdf

SWHS was/is supported by R37 CA070867 and UM1 CA182910, SMHS, by R01 CA082729 and UM1 CA173640 from the US National Cancer Institute.

Declarations

Drs Ueland and Midttun reports that they are members of the steering board of the nonprofit Foundation to Promote Research Into Functional Vitamin B12 Deficiency. No other disclosures were reported.

We thank the study participants and LC3 consortium members for their contributions in making this study possible.

References

1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer. 2015 Mar 1;136(5):E359–86.
2. Zhang L, Wang S, Che X, Li X. Vitamin D and Lung Cancer Risk: A Comprehensive Review and Meta-Analysis. Cell Physiol Biochem. 2015 May 4;36(1):299–305.
3. Chen G-C, Zhang Z-L, Wan Z, Wang L, Weber P, Eggersdorfer M, et al. Circulating 25-hydroxyvitamin D and risk of lung cancer: a dose-response meta-analysis. Cancer Causes Control CCC. 2015 Dec;26(12):1719–28.

4. Feldman D, Krishnan AV, Swami S, Giovannucci E, Feldman BJ. The role of vitamin D in reducing cancer risk and progression. Nat Rev Cancer. 2014 May;14(5):342–57.

5. Holick MF. Resurrection of vitamin D deficiency and rickets. J Clin Invest. 2006 Aug;116(8):2062–72.

6. Holick MF. Vitamin D deficiency. N Engl J Med. 2007 Jul 19;357(3):266–81.

7. Dou R, Ng K, Giovannucci EL, Manson JE, Qian ZR, Ogino S. Vitamin D and colorectal cancer: molecular, epidemiological and clinical evidence. Br J Nutr. 2016 May;115(9):1643–60.

8. Ekmekcioglu C, Haluza D, Kundi M. 25-Hydroxyvitamin D Status and Risk for Colorectal Cancer and Type 2 Diabetes Mellitus: A Systematic Review and Meta-Analysis of Epidemiological Studies. Int J Environ Res Public Health. 2017 Jan 28;14(2):127.

9. Fanidi A, Muller DC, Yuan JM, Stevens VL, Weinstein SJ, Albanes D, Prentice R, Thomsen CA, Pettinger M et al. Circulating Folate, Vitamin B6, and Methionine in Relation to Lung Cancer Risk in the Lung Cancer Cohort Consortium (LC3). J Natl Cancer Inst.doi: 10.1093/jnci/djx119

10. Midttun Ø, Ueland PM. Determination of vitamins A, D and E in a small volume of human plasma by a high-throughput method based on liquid chromatography/tandem mass spectrometry. Rapid Commun Mass Spectrom. 2011;25(14):1942–1948.

11. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2015. Available from: http://www.R-project.org/

12. Stan Development Team. The Stan C++ Library, Version 2.9.0 [Internet]. 2016. Available from: http://mc-stan.org

13. Stan Development Team. RStan: the R interface to Stan. R package version 2.9.0 [Internet]. 2016. Available from: http://mc-stan.org

14. Afzal S, Bojesen SE, Nordestgaard BG. Low Plasma 25-Hydroxyvitamin D and Risk of Tobacco-Related Cancer. Clin Chem. 2013 May 1;59(5):771–80.

15. Institute of Medicine (US) Committee to Review Dietary Reference Intakes for Vitamin D and Calcium. Dietary Reference Intakes for Calcium and Vitamin D [Internet]. Ross AC, Taylor CL, Yaktine AL, Del Valle HB, editors. Washington (DC): National Academies Press (US); 2011 [cited 2017 Feb 28]. (The National Academies Collection: Reports funded by National Institutes of Health). Available from: http://www.ncbi.nlm.nih.gov/books/NBK56070/

16. Stroud ML, Stilgoe S, Stott VE, Alhabian O, Salman K. Vitamin D - a review. Aust Fam Physician. 2008 Dec;37(12):1002–5.

Figures captions

Figure 1: Circulating log 25(OH)D as a function of day-of-year of blood draw. Scattered points are the observed concentrations, and the lines are the predicted cohort-specific mean concentrations given the calendar day on which blood was drawn. Predictions were made by regressing log 25(OH)D concentrations on sine and cosine functions of calendar day. A hierarchical model was used to allow the model coefficients to vary between cohorts.

Figure 2: Season-adjusted log 25(OH)D versus observed concentrations. Adjusted log 25(OH)D was calculated by subtracting the seasonal component of the model predicted concentration from the observed concentration. The gradient of colour varies with the magnitude of the adjustment. This adjustment ranged from a subtraction of over 0.2 log-units (for those with blood drawn during the height of summer, magenta points) to an addition of over 0.2 log-units (for those with blood drawn in the middle of winter, dark blue points).

Figure 3: Odds ratios for a doubling in 25(OH)D concentration overall, and by demographic and individual characteristics. Estimates are from conditional logistic regression models conditioned on matched case set, and adjusted for four categories of circulating cotinine. *P*-values are from likelihood ratio tests of the interaction terms between 25(OH)D and each covariate.

Figure 4: Odds ratios for a doubling in 25(OH)D concentration separately for each cohort. Estimates are from a conditional logistic regression model conditioned on matched case set, and adjusted for four categories of

circulating cotinine. The *P*-value is from likelihood ratio test of the interaction terms between 25(OH)D cohort.

Supplementary Figure 1: Odds ratios for a doubling in 25(OH)D concentration overall, and by demographic and individual characteristics, not adjusted for circulating cotinine. Estimates are from conditional logistic regression models conditioned on matched case set. *P*-values are from likelihood ratio tests of the interaction terms between 25(OH)D and each covariate.

Supplementary Figure 2: Odds ratios for a doubling in 25(OH)D3 concentration overall, and by demographic and individual characteristics. Estimates are from conditional logistic regression models conditioned on matched case set, and adjusted for four categories of circulating cotinine. *P*-values are from likelihood ratio tests of the interaction terms between 25(OH)D3 and each covariate.

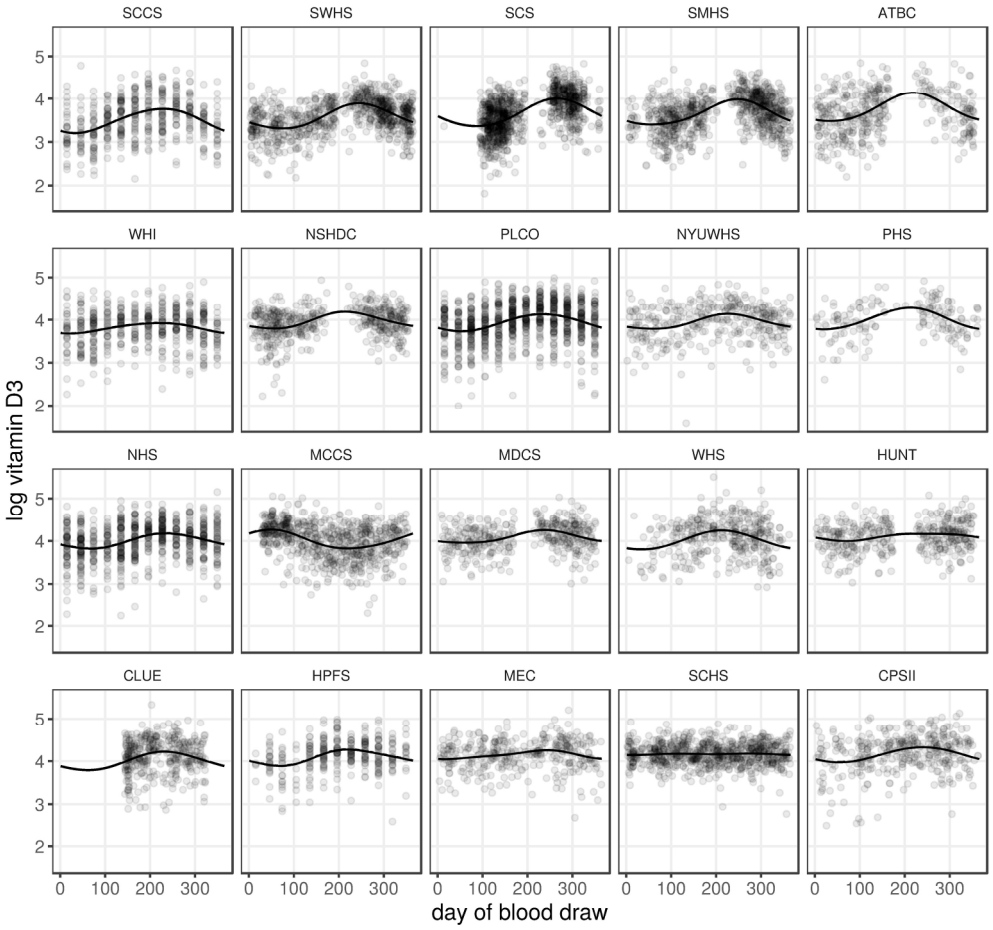


Figure 1: Circulating log 25(OH)D as a function of day-of-year of blood draw. Scattered points are the observed concentrations, and the lines are the predicted cohort-specific mean concentrations given the calendar day on which blood was drawn. Predictions were made by regressing log 25(OH)D concentrations on sine and cosine functions of calendar day. A hierarchical model was used to allow the model coefficients to vary between cohorts.

216x205mm (300 x 300 DPI)

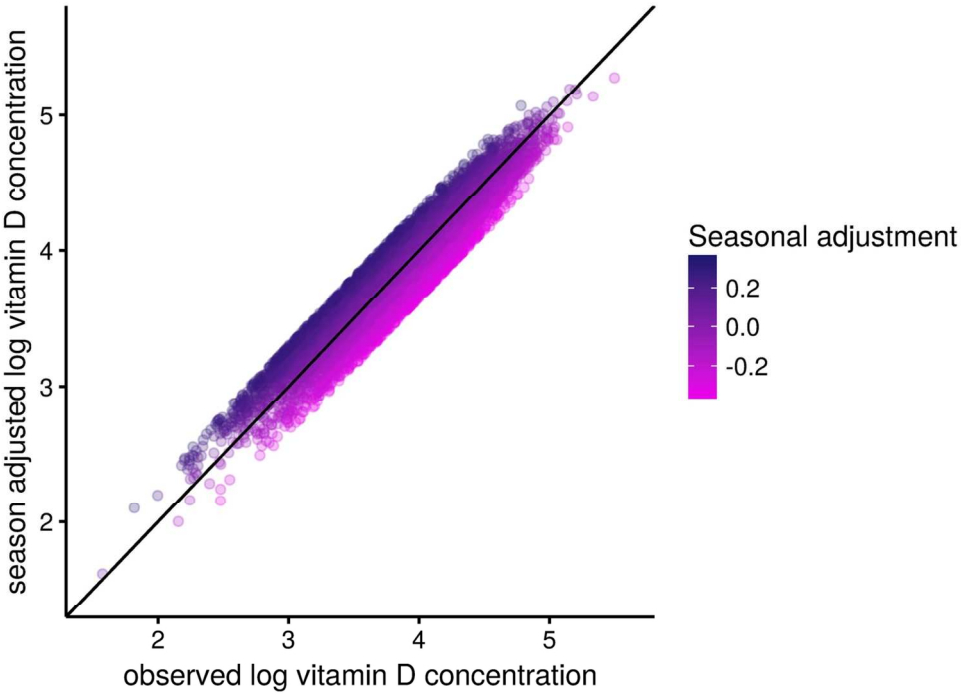


Figure 2: Season-adjusted log 25(OH)D versus observed concentrations. Adjusted log 25(OH)D was calculated by subtracting the seasonal component of the model predicted concentration from the observed concentration. The gradient of colour varies with the magnitude of the adjustment. This adjustment ranged from a subtraction of over 0.2 log-units (for those with blood drawn during the height of summer, magenta points) to an addition of over 0.2 log-units (for those with blood drawn in the middle of winter, dark blue points).

127x90mm (300 x 300 DPI)

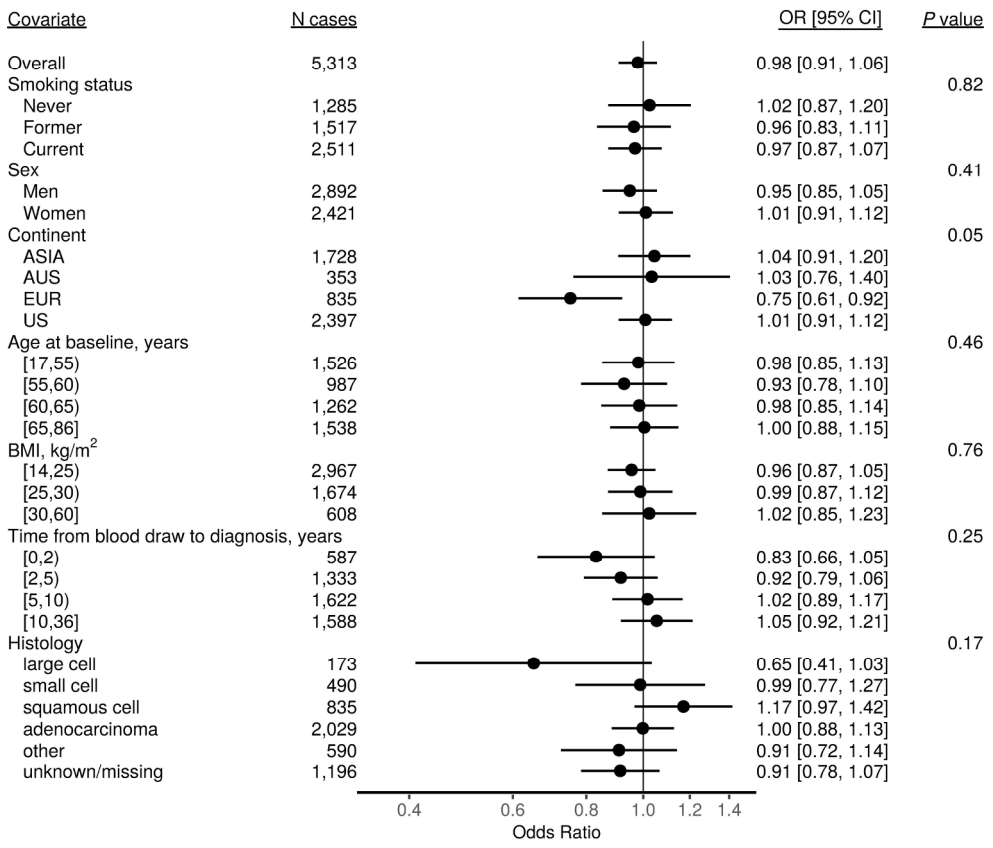


Figure 3: Odds ratios for a doubling in 25(OH)D concentration overall, and by demographic and individual characteristics. Estimates are from conditional logistic regression models conditioned on matched case set, and adjusted for four categories of circulating cotinine. *P*-values are from likelihood ratio tests of the interaction terms between 25(OH)D and each covariate.

203x180mm (300 x 300 DPI)

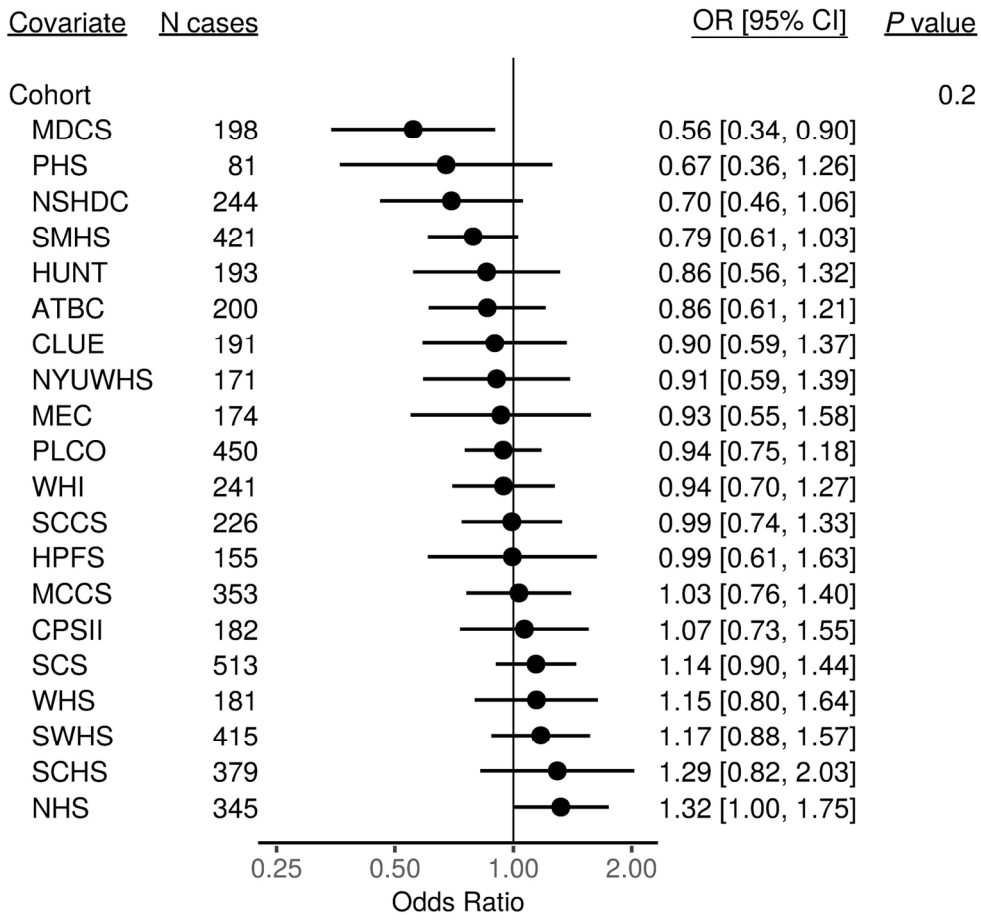
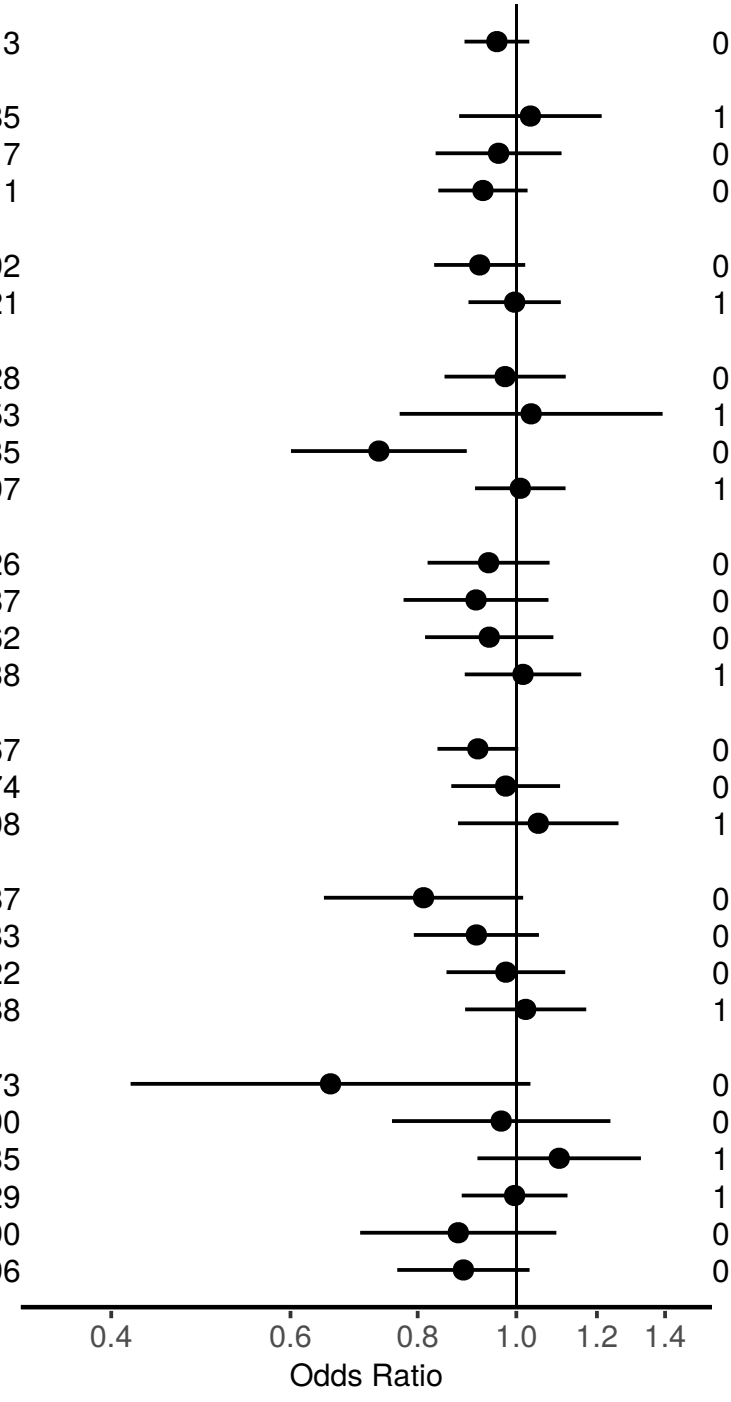


Figure 4: Odds ratios for a doubling in 25(OH)D concentration separately for each cohort. Estimates are from a conditional logistic regression model conditioned on matched case set, and adjusted for four categories of circulating cotinine. The *P*-value is from likelihood ratio test of the interaction terms between 25(OH)D cohort.

153x154mm (300 x 300 DPI)

Covariate	N cases	OR [95% CI]	*P* value
Overall	5,313	0.96 [0.89, 1.03]	
Smoking status			0.54
Never	1,285	1.03 [0.88, 1.21]	
Former	1,517	0.96 [0.83, 1.11]	
Current	2,511	0.93 [0.84, 1.03]	
Sex			0.29
Men	2,892	0.92 [0.83, 1.02]	
Women	2,421	1.00 [0.90, 1.11]	
Continent			0.04
ASIA	1,728	0.97 [0.85, 1.12]	
AUS	353	1.03 [0.77, 1.39]	
EUR	835	0.73 [0.60, 0.89]	
US	2,397	1.01 [0.91, 1.12]	
Age at baseline, years			0.27
[17,55)	1,526	0.94 [0.82, 1.08]	
[55,60)	987	0.91 [0.77, 1.08]	
[60,65)	1,262	0.94 [0.81, 1.09]	
[65,86]	1,538	1.02 [0.89, 1.16]	
BMI, kg/m^2			0.33
[14,25)	2,967	0.92 [0.84, 1.00]	
[25,30)	1,674	0.98 [0.86, 1.10]	
[30,60]	608	1.05 [0.88, 1.26]	
Time from blood draw to diagnosis, years			0.33
[0,2)	587	0.81 [0.65, 1.02]	
[2,5)	1,333	0.91 [0.79, 1.05]	
[5,10)	1,622	0.98 [0.85, 1.12]	
[10,36]	1,588	1.02 [0.89, 1.17]	
Histology			0.22
large cell	173	0.66 [0.42, 1.03]	
small cell	490	0.97 [0.75, 1.24]	
squamous cell	835	1.10 [0.92, 1.33]	
adenocarcinoma	2,029	1.00 [0.88, 1.12]	
other	590	0.88 [0.70, 1.09]	
unknown/missing	1,196	0.89 [0.76, 1.03]	



Covariate	N cases	OR [95% CI]	*P* value
Overall	5,313	0.96 [0.89, 1.04]	
Smoking status			0.7
Never	1,285	1.02 [0.87, 1.20]	
Former	1,517	0.95 [0.82, 1.09]	
Current	2,511	0.95 [0.85, 1.05]	
Sex			0.8
Men	2,892	0.95 [0.86, 1.06]	
Women	2,421	0.97 [0.88, 1.08]	
Continent			0.23
ASIA	1,728	1.03 [0.89, 1.20]	
AUS	353	1.04 [0.77, 1.41]	
EUR	835	0.80 [0.66, 0.98]	
US	2,397	0.97 [0.87, 1.07]	
Age at baseline, years			0.31
[17,55)	1,526	0.93 [0.80, 1.07]	
[55,60)	987	0.95 [0.80, 1.12]	
[60,65)	1,262	1.06 [0.91, 1.23]	
[65,86]	1,538	0.94 [0.82, 1.08]	
BMI, kg/m^2			0.89
[14,25)	2,967	0.95 [0.86, 1.04]	
[25,30)	1,674	0.98 [0.86, 1.11]	
[30,60]	608	0.98 [0.81, 1.19]	
Time from blood draw to diagnosis, years			0.15
[0,2)	587	0.87 [0.69, 1.10]	
[2,5)	1,333	0.87 [0.76, 1.01]	
[5,10)	1,622	1.08 [0.94, 1.24]	
[10,36]	1,588	1.01 [0.88, 1.16]	
Histology			0.27
large cell	173	0.75 [0.49, 1.15]	
small cell	490	0.91 [0.71, 1.18]	
squamous cell	835	1.17 [0.96, 1.42]	
adenocarcinoma	2,029	0.95 [0.84, 1.07]	
other	590	0.86 [0.69, 1.06]	
unknown/missing	1,196	0.98 [0.84, 1.14]	

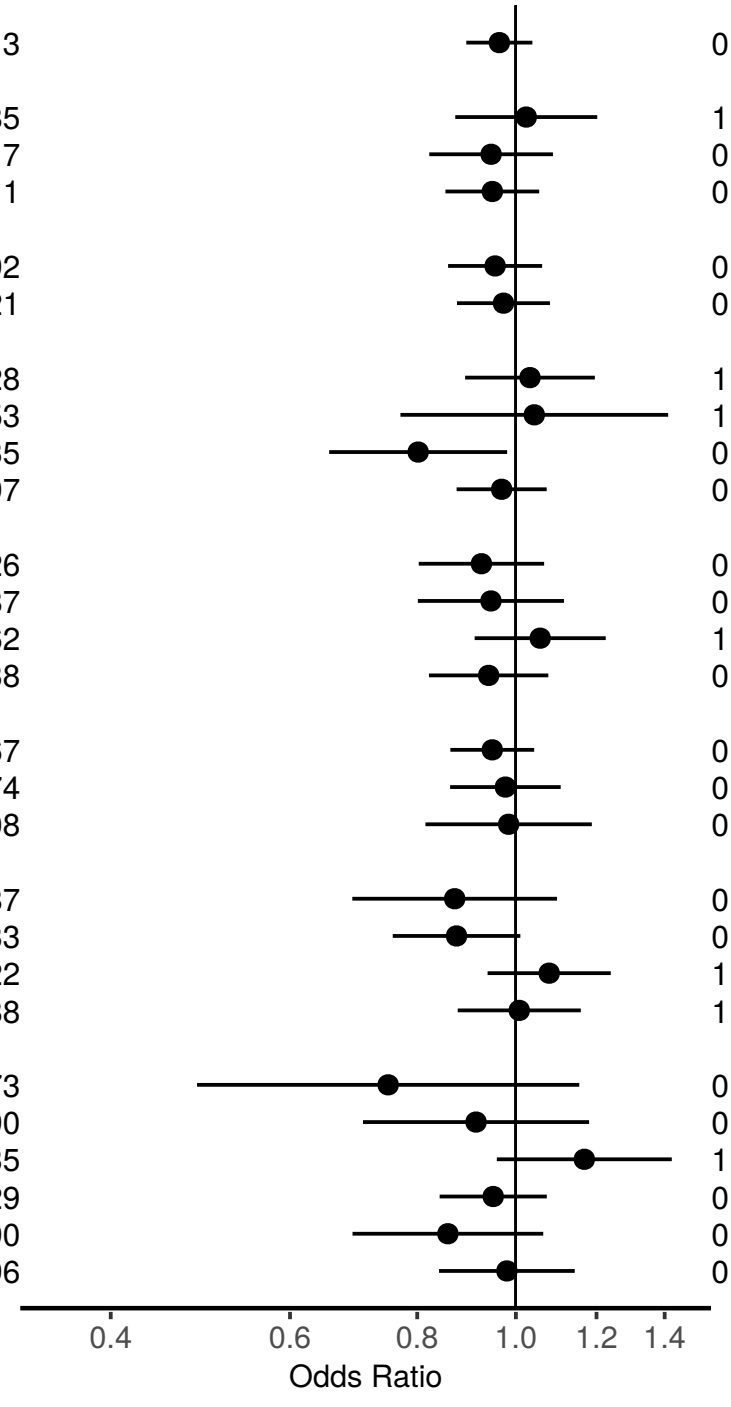


Table 1: Distribution of demographic and individual characteristics of cases and controls, overall and by continent.

						Asia				Australia				Europe				United States			
		control		case		control		case		control		case		control		case		control		case	
		n	percent	*n*	percent	*n*	percent	*n*	percent	*n*	percent	*n*	percent	*n*	percent	*n*	percent	*n*	percent	*n*	percent
	Total	5313	100	5313	100	1728	100	1728	100	353	100	353	100	835	100	835	100	2397	100	2397	100
Sex	Men	2892	54	2892	54	1214	70	1214	70	212	60	212	60	475	57	475	57	991	41	991	41
	Women	2421	46	2421	46	514	30	514	30	141	40	141	40	360	43	360	43	1406	59	1406	59
Smoking status	Never	1285	24	1285	24	560	32	560	32	49	14	49	14	107	13	107	13	569	24	569	24
	Former	1517	29	1517	29	176	10	176	10	145	41	145	41	190	23	190	23	1006	42	1006	42
	Current	2511	47	2511	47	992	57	992	57	159	45	159	45	538	64	538	64	822	34	822	34
Age at recruitment (years)	[17,55)	1547	29	1526	29	427	25	422	24	87	25	89	25	251	30	251	30	782	33	764	32
	[55,60)	995	19	987	19	368	21	355	21	57	16	64	18	198	24	191	23	372	16	377	16
	[60,65)	1247	23	1262	24	427	25	456	26	125	35	111	31	227	27	228	27	468	20	467	19
	[65,86]	1524	29	1538	29	506	29	495	29	84	24	89	25	159	19	165	20	775	32	789	33
$25(OH)D_2$ detectable	no	3963	75	3951	74	1413	82	1413	82	337	95	340	96	656	79	667	80	1557	65	1531	64
	yes	1350	25	1362	26	315	18	315	18	16	5	13	4	179	21	168	20	840	35	866	36
Season-adjusted 25(OH)D (nmol/L)	[7,42)	1328	25	1370	26	690	40	703	41	69	20	62	18	132	16	163	20	437	18	442	18
	[42,56.6)	1328	25	1322	25	511	30	494	29	99	28	105	30	229	27	255	31	489	20	468	20
	[56.6,73.6)	1328	25	1274	24	348	20	334	19	104	29	87	25	264	32	230	28	612	26	623	26
	[73.6,237]	1329	25	1347	25	179	10	197	11	81	23	99	28	210	25	187	22	859	36	864	36

Table 2: Characteristics of the lung cancer cases, overall and by continent.

		Total		Asia		Australia		Europe		United States	
		n	percent	*n*	percent	*n*	percent	*n*	percent	*n*	percent
Histology	large cell	173	3	15	1	31	9	15	2	112	5
	small cell	490	9	97	6	47	13	103	12	243	10
	squamous cell	835	16	318	18	67	19	162	19	288	12
	adenocarcinoma	2029	38	589	34	152	43	260	31	1028	43
	other	590	11	116	7	53	15	128	15	293	12
	unknown/missing	1196	23	593	34	3	1	167	20	433	18
Time to diagnosis (years)	[0,2)	587	11	258	15	35	10	50	6	244	10
	[2,5)	1333	25	457	26	45	13	96	11	735	31
	[5,10)	1622	31	592	34	101	29	266	32	663	28
	[10,36]	1588	30	421	24	172	49	423	51	572	24

Table 3: Odds ratios [95% confidence intervals] for lung cancer by group of seasonally adjusted 25(OH)D, and for a doubling in 25(OH)D.

	N controls	N cases	OR	95% CI	*p*
Categorical					
[7,25)	245	283	1.00		0.59
[25,50)	1820	1825	0.91	[0.75, 1.1]	
[50,75)	2014	1940	0.89	[0.73, 1.08]	
[75,237]	1234	1265	0.93	[0.75, 1.15]	
Categorical (fourths)					
[7,42)	1328	1370	1.00		0.93
[42,56.6)	1328	1322	0.98	[0.87, 1.1]	
[56.6,73.6)	1328	1274	0.96	[0.85, 1.09]	
[73.6,237]	1329	1347	0.99	[0.87, 1.13]	
Continuous					
Doubling in concentration	5313	5313	0.98	[0.91, 1.06]	0.58

Estimates from conditional logistic regression models conditioned on matched case set, and adjusted for four groups of circulating cotinine concentration. *P*-values are from likelihood ratio tests of the 25(OH)D terms.

Supplementary materials: Circulating vitamin D and lung cancer risk

Supplementary Table 1: Studies participating in the Lung Cancer Cohort Consortium (LC3)

Cohort name	Abbreviation	Year(s) of enrolment/blood collection	Case-control pairs
Cohorts recruited in the US			
Southern Community Cohort Study	SCCS	2002-2009	226
Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial	PLCO	1993-2001	450
ACS Cancer Prevention Study-II	CPS-II	1992	182
Campaign Against Cancer and Stroke + Campaign Against Cancer and Heart Disease	CLUE I & Clue II	1974 & 1989	191
Multiethnic Cohort Study	MEC	2001-2007	174
Women's Health Initiative I & II	WHI	1994-1998	241
NYU Women's Health Study	NYUWHS	1985-1991	171
Women's Health Study	WHS	1992-1995	181
Nurses' Health Study	NHS	1989-1990	345
Health Professionals Follow-up Study	HPFS	1993-1994	155
Physicians Health Study I & II	PHS	1982-1983 & 1997-2000	81
Cohorts recruited in Europe/Australia			
The Melbourne Collaborative Cohort Study	MCCS	1990-1994	353
Northern Sweden Health and Disease Study	NSHDS	1985-onward	244
The Malmö Diet and Cancer Study	MDCS	1991-1996	198
Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study	ATBC	1985-1988	200
The Nord-Trøndelag Health Study	HUNT	1995-1997	193
Cohorts recruited in Asia			
Shanghai Cohort Study	SCS	1986-1989	513
Singapore Chinese Health Study	SCHS	1993-1998	379
Shanghai Men's Health Study	SMHS	2001-2006	421
Shanghai Women's Health Study	SWHS	1996-2000	415

Supplementary materials: Circulating vitamin D and lung cancer risk

Supplementary Methods

Material and Methods

Study population

We invited all prospective cohort studies that in 2009 were members in the US National Cancer Institute (NCI) Cohort Consortium to participate in the study. Additional inclusion criteria included the occurrence of at least 200 incident lung cancer cases with baseline questionnaire data and either plasma or serum samples available. Twenty cohorts fulfilled those criteria and accepted to participate, resulting in a combined cohort population of over 2,000,000 participants from North America, Europe, Asia and Australia.

Brief description of the participating cohorts in the Lung Cancer Cohort Consortium

1) **US cohorts**

The Women's Health Initiative (WHI)

WHI is a long-term health study of 161,808 post-menopausal women aged 50 to 79 years at 40 clinical centers throughout the U.S. WHI comprises a Clinical Trial (CT) component (68,132 women), and an Observational Study (OS) component (93,676 women), and has included several extension studies. Some detailed descriptions of WHI have been previously presented[1 2].The CT evaluated two forms of postmenopausal hormone therapy, a low-fat dietary pattern intervention, and calcium and vitamin D supplementation in a randomized, controlled fashion, in a partial factorial design. The hormone therapy component findings led to major reductions in the use of hormone therapy worldwide, and are thought to have led to noteworthy reductions in breast cancer incidence.

In the present study lung cancer cases occurring during the follow-up of WHI cohorts since enrolment (1993-1998) among non-smoking women, were matched 1-1 to corresponding non-smoking lung cancer free controls, for serum and DNA analyte comparisons.

The Southern Community Cohort Study (SCCS)

The Southern Community Cohort Study (SCCS)[3] is a prospective cohort of African and non-African Americans which during 2002-2009 enrolled approximately 86,000 residents aged 40-79 years across 12 southern states. Recruitment occurred mainly at community health centers, institutions providing basic health services primarily to the medically uninsured, so that the cohort includes many adults of lower income and educational status. Each study participant completed a detailed baseline questionnaire, and nearly 90% provided a biologic specimen (approximately 45% a blood sample and 45% buccal cells). Follow-up of the cohort is conducted by linkage to national mortality registers and to state cancer registries.

Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO)

The PLCO study, a randomized trial aimed at evaluating the efficacy of screening in reducing cancer mortality, recruited approximately 155,000 men and women age 55 to 74 years from 1992 to 2001[4]. Screening for lung cancer among participants in the intervention arm included a chest x-ray at baseline followed by either three annual x-rays (for current or former smokers at enrollment) or two annual x-rays (for never smokers); participants in the control arm received routine health care. Screening-arm participants provided data on sociodemographic factors, smoking behavior, anthropometric characteristics, medical history, and family history of cancer, as well as blood samples annually for the first 6 years of the study (baseline [T0] and T1 through T5). Lung cancers were ascertained through annual questionnaires mailed to the participants, and positive reports were followed up by abstracting medical records or death certificates. Follow-up in the trial as of July 2009 was 96.7%.

We conducted a nested case-control study within the screening arm of the PLCO trial. As of December 31, 2004, 898 lung cancers were diagnosed among the 77,464 participants. Patients were excluded because of missing baseline questionnaire, previous history of any cancer, diagnosis of multiple cancers during follow-up, missing smoking information at baseline, missing consent for utilization of biologic specimens for etiologic studies, or unavailability/insufficient quantity of serum or DNA specimens. Hemolyzed vials were excluded. Participants for this nested case-control study were sampled from the intervention arm.

Controls were individuals free of cancer at the time of a case's lung cancer diagnosis. Controls were individually matched to lung cancer patients on sex, date of birth +/- 1 year with a possible relaxation to 5 years, race, study year of blood draw, date of blood draw +/- 1 month (with a possible relaxation to 3/6 months), time of blood draw (6AM-9AM, 9AM-12PM, other), smoking categories (smoking status at enrollment, never, former, or current smoker; cumulative amount of smoking (0 to 29, 30 to 39, 40 to 49, and 50+ pack-years), with additional matching for time since quitting (< 15 years and ≥ 15 years) for former smokers, cigarettes per day smoked, and number of days in the study.

The New York University Women's Health Study (NYUWHS)

The New York University Women's Health Study (NYUWHS) is a prospective cohort study of women enrolled at a mammography screening center in New York City. From March 1985 through June 1991, 14,274 women between the ages of 34 and 65 were enrolled in the study. Because the original focus of the study was endogenous hormones and breast cancer, women who had taken hormone medications in the 6 months preceding baseline enrolment were not eligible for the study.

At the time of enrolment, data on demographics, anthropometric measures, medical history, reproductive and lifestyle variables were collected through self-administered questionnaires after written informed consent was obtained. Incident lung cancer cases were identified through active follow-up of the cohort conducted with questionnaires mailed approximately every two to four years and record linkages with state tumor registries in New York, New Jersey, and Florida, as well as record linkage with the National Death Index (NDI). Medical records were obtained to verify reported cancer outcomes.

The American Cancer Society Cancer Prevention Study-II (CPS-II) Nutrition Cohort

The CPS-II Nutrition Cohort is a prospective study of cancer incidence and mortality among 86,404 men and 97,786 women. The CPS-II Nutrition Cohort, which is described in detail elsewhere[5], was initiated in 1992 as a subgroup of CPS-II, a prospective study of cancer mortality involving approximately 1.2 million Americans begun in 1982. Participants in the CPS-II Nutrition Cohort were recruited from CPS-II members who resided in 21 states and were between the ages of 50 and 74 years. At enrollment in 1992/1993, participants completed a self-administered questionnaire that included demographic, medical, dietary, and lifestyle information. Follow-up questionnaires were sent to all living Nutrition Cohort members in 1997, and every two years after this to update exposure information and to ascertain newly diagnosed cancers. Between June 1998 and June 2001, blood samples were collected from a subset of CPS-II Nutrition Cohort participants (21,965 women and 17,411 men).

Incident lung cancer cases were identified through self-report on a follow-up questionnaire, linkage with state cancer registries, or death certificates. Self-reported cancers were verified through medical records.

The Campaign Against Cancer and Stroke (CLUE I) and the Campaign Against Cancer and Heart Disease (CLUE II).

The CLUE studies include two large cohorts of volunteers from Washington County, Maryland that were enrolled in 1974 and 1989, respectively. CLUE I was conducted in Washington County, Maryland, in the fall of 1974. Brief health histories and blood pressures were taken and 15 ml of blood was drawn from 26,147 volunteers (23,951 were residents of Washington County) at the time of enrollment. Linkage of the records from this program to those of a private census in the summer of 1975 indicated that almost a third of the adult population of the county had participated. CLUE II was an outgrowth of CLUE I conducted from May through October in 1989. As in CLUE I, a brief health history was obtained and 20 ml of blood was drawn. A blood sample was collected from 32,894 volunteers at the time of enrollment (25,076 were residents of Washington County). Participants were also given a

food frequency questionnaire to complete at home and were asked to return it with a toenail clipping of the large toe for trace metal assays. Comparisons with published figures from the 1990 Census indicted that approximately 30 percent of adult residents had participated.

The Multiethnic Cohort (MEC)

The MEC includes over 215,000 men and women aged 45-75 years at recruitment from five different racial/ethnic groups (African Americans, Japanese Americans, Native Hawaiians, Latinos and European Americans) in Hawaii and California[6]. The cohort was assembled in 1993-1996 by mailing a self-administered, 26-page questionnaire to obtain extensive information on demographics, medical and reproductive histories, medication use, family history of various cancers, physical activity and diet. Identification of incident cancer cases is by regular linkage with the Hawaii, Los Angeles County and California SEER registries. From 1995 to 2001, blood collection was conducted from incident cases with breast, prostate, or colorectal cancers, as well as a random sample of cohort participants to serve as controls in genomic nested case-control studies (participation rate 72% and 63%, respectively). In addition, from 2001 to 2006, blood was also collected prospectively, without regard for cancer diagnosis, from willing cohort participants. Approximately 67,000 gave a blood sample (participation rate 43%). All incident lung cancer cases diagnosed before 2010 with a pre-diagnostic blood sample were considered for inclusion in this study. Each case was matched to a control based on study site, sex, age, race/ethnicity, smoking status, hours of fasting, and date and time of blood draw.

Women's Health Study (WHS)

The WHS was a randomized trial of low-dose aspirin, vitamin E, and beta-carotene in the primary prevention of cardiovascular disease and cancer beginning in 1992 among 39,876 female US health professionals aged ≥45 years[7]. Information on major clinical, lifestyle, and dietary factors was collected via self-reports on baseline questionnaires. Women also provided baseline bloods. During more than two decades of follow-up, WHS participants reporting new cases of cancer on annual follow-up questionnaires were confirmed by medical record review by the WHS Endpoints Committee. Reports of cancer were confirmed on the basis of pathology or cytology reports or, rarely, strong clinical and radiologic or laboratory marker evidence when a pathology or cytology review was not conducted. Only confirmed cases of lung cancer were included in the present analyses, which were matched with eligible controls.

Physicians' Health Study (PHS)

The PHS I began in 1982 as a randomized trial of aspirin and beta-carotene for the primary prevention of heart disease and cancer among 22,071 male, Caucasian physicians initially aged 40 to 84 years[8], followed by the PHS II trial beginning in 1997 to evaluate beta-carotene, vitamin C, vitamin E, and a daily multivitamin on the prevention of cancer, CVD, and other endpoints. The PHS II included 14,641 men, with 7,641 participants from the PHS I plus 7,000 new physicians, for a total of 29,071 PHS participants[9]. A wide range of demographic, clinical, and lifestyle factors were assessed via baseline questionnaires, along with baseline bloods. PHS participants reported major clinical endpoints, including cancer, yearly in a mailed questionnaire and postcards every six months. Self-reported, incident lung cancer cases were confirmed through medical record review by the PHS Endpoints Committee in included in the present analyses.

The Nurses' Health Study (NHS)

The Nurses' Health Study (NHS)[10,11] was established in 1976, when 121,700 married female registered nurses aged 30 to 55 years residing in 11 States in the U.S. completed and returned a self-administered questionnaire. Questionnaires have been mailed to participants in both cohorts every 2 years since baseline to collect updated information on demographics, lifestyle factors, medical history, and disease outcomes. A semi-quantitative food frequency questionnaire (FFQ) was administered to obtain information on usual dietary intake over the previous year. The reproducibility and validity of the FFQs have been established[12-16]. The FFQ was first administered in 1980 in the NHS, and were repeated almost every 4 years thereafter. For each food item, the questionnaire specified a common serving size and queried respondents on average intake during the previous year; responses in 9 categories ranged from almost never to 6 or more per day. Most nutritional variables measured by these FFQs have been developed, tested, and refined by our group over the past 30 years (https://regepi.bwh.harvard.edu/health/).

The follow-up rate has been greater than 90%. The institutional review board at the Brigham and Women's Hospital approved the study. As approved by the committee, return of the questionnaires was considered to imply informed consent. Cases of lung cancer were self-reported by the participants or identified on their death certificates and were subsequently confirmed by medical records.

Health Professionnals Follow-up Study (HPFS)

The Health Professionals Follow-up Study (HPFS)[17] is an ongoing cohort study of 51,529 U.S. male professionals who were aged 40 to 75 years at baseline in 1986. Questionnaires have been mailed to participants every 2 years since baseline to collect updated information on demographics, lifestyle factors, mwedical history, and disease outcomes. The follow-up

rate has been greater than 90%. The institutional review board at the Harvard T.H. Chan School Public Health approved this study. As approved by the committee, return of the questionnaires was considered to imply informed consent. A semi-quantitative food frequency questionnaire (FFQ) was administered to obtain information on usual dietary intake over the previous year. The FFQ was first administered in 1986 and was repeated every 4 years thereafter. The reproducibility and validity of the FFQ have been estabished[14 18]. For each food item, the questionnaire specified a common serving size and queried respondents on average intake during the previous year; responses in 9 categories ranged from almost never to 6 or more per day. Cases of lung cancer were self-reported by the participants or identified on their death certificates and were subsequently confirmed by medical records.

We would like to thank the participants and staff of the Nurses' Health Study and the Health Professionals Follow-up Study, for their valuable contributions as well as the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, WY. This work was supported by the National Institutes of Health (NIH) grants, UM1CA167552, UM1CA186107, P01CA87969, and R01CA49449. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

2) European/Australian cohorts

The Melbourne Collaborative Cohort Study (MCCS)

The MCCS is a prospective cohort study of 41,514 participants (17,045 men and 24,469 women) aged 27-88 years at recruitment[19]; 99.3% of whom were aged 40-69 years. Recruitment occurred between 1990 and 1994. Southern European migrants to Australia (including 5,411 Italians and 4,525 Greeks) were over-sampled to extend the range of lifestyle exposures and to increase genetic variation.

Subjects were recruited via Electoral Rolls (registration to vote is compulsory for adults in Australia), advertisements, and community announcements in local media. Comprehensive lists of Italian and Greek surnames were used to target southern European migrants in phonebooks and electoral rolls. Passive follow-up of the cohort has been conducted by record linkage to Electoral Rolls, electronic phonebooks, the Victorian Cancer Registry and death records; as well as national cancer and death records to identify events outside of Victoria.

At recruitment participant's height and weight were measured, blood samples collected and questionnaires covering lifestyle (diet, smoking, physical activity and alcohol consumption), demographics and medical history completed.

Incident lung cancer cases under 80 years of age up to 31st December 2010 were identified through record linkage.

The Malmö Diet and Cancer Study (MDCS)

The Malmö Diet and Cancer Study (MDCS) is a population-based prospective cohort study that between 1991 and 1996 recruited men and women aged 44 to 74 years of age living in Malmö, Sweden[20]. The main goal of the MDCS is to study the impact of diet on cancer incidence and mortality. It consists of a baseline examination including dietary assessment, a self-administered questionnaire, anthropometric measurements and collection of blood samples.

The Northern Sweden Health and Disease Study Cohort (NSHDS)

The Northern Sweden Health and Disease Study (NSHDS) encompasses several prospective cohorts, the current study involving study participants from the Västerbotten Intervention Project (VIP), a sub-cohort within NSHDS[21]. VIP is an ongoing prospective cohort and intervention study intended for health promotion of the general population of the Västerbotten County in northern Sweden. VIP was initiated in 1985 and all residents in the Västerbotten County were invited to participate by attending a health check-up at 40, 50 and 60 years of age. Participants were asked to complete a self-administered questionnaire including various demographic factors such as education, smoking habits, physical activity and diet. In addition, height and weight were measured and participants were asked to donate a fasting blood sample for future research. Incident lung cancer cases were identified through linkage with the regional cancer registry.

The Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study (ATBC)

The ATBC Study was a randomized, double-blind, placebo-controlled, primary cancer prevention trial testing daily supplementation with α-tocopherol (50 mg/day) or β-carotene (20 mg/day), or both[22]. Between 1985 and 1988, the study enrolled and randomized 29,133 50-69 year old male cigarette smokers from southwestern Finland. Study supplementation continued for 5-8 years (median 6.1 years) until death or trial closure (April 30, 1993). At baseline, participants completed questionnaires regarding general risk factors, medical history, smoking habits, and dietary intake. Height, weight, heart rate, and blood pressure were measured by trained nurses and fasting serum samples were collected. Lung cancer cases diagnosed during follow-up through 2009 were identified through linkage with the Finnish Cancer Registry, and 200 were individually matched to 200 controls for this analysis. For the current study, study participants were selected randomly from the different arms of the ATBC study.

Supplementary materials: Circulating vitamin D and lung cancer risk

The Nord-Trøndelag Health Study (HUNT)

The HUNT study is a longitudinal population based study having invited all persons aged 20-100 years living in the county of Nord-Trøndelag, Norway to three data collections, HUNT1 (1984-86), HUNT2 (1995-97) and HUNT 3 (2006-08) (http://www.ntnu.edu/hunt). Each time about 95,000 persons were invited, and respectively 88%, 69% and 54% participated. Comprehensive data on life style, health status, symptoms, diseases and anthropometrics have been collected through questionnaires, interviews and clinical examinations, and in HUNT2 and HUNT3 biological material as blood and urine additionally were collected and stored. Incident cancer cases were identifiead via linkage to the Norwegian Cancer Registry.
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 Health Authority, and the Norwegian Institute of Public Health.

3) Asian cohorts

The Shanghai Men's Health Study (SMHS) and the Shanghai Women's Health Study (SWHS)

The SMHS and SWHS are population-based cohort studies conducted in eight communities of urban Shanghai. Their designs and methods have been described elsewhere[23 24].
Briefly, the SWHS recruited 74,941 women during 1997-2000 (response rate: 93%) and the SMHS recruited 61,480 men during 2002-2006 (response rate: 74%). Similar methods and questionnaires were used in both studies. At baseline in-person interviews, information on sociodemographic, diet, lifestyle, occupation and medical history was obtained; height, body weight, and waist circumference were measured. Blood samples were collected from 75% of the study participants in both studies, processed within 6 hours, and stored at -70℃ until analysis.
The SMHS and SWHS have been followed up by annual record linkage with the population-based Shanghai Cancer Registry and Shanghai Vital Statistics Registry and in-person surveys every 2-3 years. Exposure information, including dietary intake, was updated in the in-person follow-up surveys. All possible matches from the linkages are checked manually and verified by home visits. Medical charts were obtained from the initial diagnostic hospitals to verify cancer diagnosis. Death certificate data from the Shanghai Vital Statistics Unit was used to identify the primary cause of death.
The studies were approved by the Institutional Review boards of the Shanghai Cancer Institute and Vanderbilt University. Informed consent was obtained from all participants.

The Singapore Chinese Health Study (SCHS)

Supplementary materials: Circulating vitamin D and lung cancer risk

The design of the SCHS study has been described[25 26]. Briefly, the cohort was drawn from permanent residents or citizens of Singapore who resided in government-built housing estates (86% of the Singapore population reside in such facilities). The eligible age range for cohort enrolment was 45-74 years. We restricted study subjects to the two major dialect groups of Chinese in Singapore: the Hokkiens and the Cantonese, who originated from Fujian and Guangdong provinces in Southern China, respectively. Between April 1993 and December 1998, 63,257 subjects (approximately 85% of eligible subjects) were enrolled into the cohort study. At recruitment, each study subject was interviewed in person by a trained interviewer using a structured questionnaire that emphasized current diet assessed via a validated, 165-item food frequency questionnaire. The questionnaire also requested information on demographics, lifetime use of tobacco, incense use, current physical activity, usual sleep duration, reproductive history (women only), occupational exposure, medical history, and family history of cancer.

Beginning in April 1994, a random 3% sample of cohort participants were asked to provide blood or buccal cell (if request for blood sample was denied), and spot urine samples. Eligibility for this biospecimen subcohort was extended to all surviving cohort participants starting in January 2000. By April 2005, all surviving cohort subjects had been contacted for biospecimen donation. Approximately 60% of eligible cohort participants donated biospecimens.

The cohort has been passively followed for death and cancer occurrence through regular record linkage with the population-based Singapore Cancer Registry and the Singapore Registry of Births and Deaths. Migration out of Singapore, especially among housing estate residents, is negligible. As of latest update, only 55 individuals from this cohort were known to be lost to follow-up due to migration and other reason.

A nested case-control study of incident lung cancer cases within the Singapore Chinese Health Study was used to examine the association between serum levels of vitamin B_6 and other compounds in the one-carbon metabolism pathway and risk of lung cancer. Briefly, 422 lung cancer cases were identified among cohort participants with available prediagnostic plasma samples as of 12/31/2011. For each case, one control subject was randomly selected from all eligible cohort members who were alive and free of cancer on the date of cancer diagnosis of the index case. The control subject was individually matched to the index case by gender, dialect group (Hokkien, Cantonese), age at enrolment (±3 years), date of baseline interview (±2 year), date of biospecimen collection (±6 months), and smoking status (current, former, and never smokers). For current smokers, cases and controls were further matched by number of cigarettes per day (<15, ≥15 cigarettes/day). For former smokers, cases and controls were further matched by years since quitting smoking

(<10, ≥10 years). One plasma aliquot per subject was retrieved from the biorepository and all plasma samples were sent to the laboratory (B-vital) for measurements.

The Shanghai Cohort Study (SCS)

The SCS study is a residential cohort of 18,244 men in Shanghai, China, assembled during 1986-89 when subjects were between the ages of 45 and 64 years. Approximately 80% of eligible men participated in the study. At the time of recruitment, each cohort subject was interviewed in-person by a trained nurse interviewer using a structured questionnaire that included background information, history of tobacco and alcohol use, current diet, and medical history[27 28].

At the completion of the interview, the nurse collected a 10 ml blood and a single void urine specimen from the study participant. Blood and urine samples were kept in insulated boxes with ice (0-2^{o}C). The serum was separated from blood specimen within 3-4 hours after collection. Two sets of serum (2 ml and 1 ml, respectively) and two sets of urine samples (10 ml each) per subject have been stored at -80^{o}C.

The cohort has been followed for the occurrence of cancer and death through routine ascertainment of new cases from the population-based Shanghai Cancer Registry and Shanghai Vital Statistics Units. To maximize the cancer findings and minimize the loss of follow-up, we have recontacted each surviving cohort member annually. Retired nurses visit the last known address of each living cohort member and record details of the interim health history of the cohort member. As of December 31, 2014, cumulatively 612 (3.4%) original subjects were lost to follow-up (i.e., persons we have no record of death and we have been unable to locate through our annual follow-up recontacts), and 574 (3.1%) refused to our continued follow-up interview (their cancer and vital status has been continually updated through record linkage analyses) after 26 years of follow-up since the beginning of the study.

A nested case-control study of incident lung cancer cases within the Shanghai Cohort Study was used to examine the association between serum levels of vitamin B_6 and other compounds in the one-carbon metabolism pathway and risk of lung cancer. Briefly, 516 lung cancer cases were identified among cohort participants with available serum samples as of 12/31/2006. For each case, we randomly selected one control subject from all cohort members who were free of cancer and alive at the time of cancer diagnosis of the index case. Controls were matched to the index case by age at enrolment (±2 years), date of biospecimen collection (±1 month) and neighbourhood of residence at recruitment, and smoking status (current, former and never smokers) as established previously for other studies. For former smokers, cases and controls were further matched by years since

quitting smoking (<10 vs ≥10 years). One serum vial per subject was retrieved from biorepository and all serum samples were sent to the laboratory (B-vital) for measurements.

References

1. Design of the Women's Health Initiative clinical trial and observational study. The Women's Health Initiative Study Group. *Control Clin Trials* 1998;19(1):61-109.
2. Hays J, Hunt JR, Hubbell FA, Anderson GL, Limacher M, Allen C, et al. The Women's Health Initiative recruitment methods and results. *Ann Epidemiol* 2003;13(9 Suppl):S18-77.
3. Signorello LB, Hargreaves MK, Steinwandel MD, Zheng W, Cai Q, Schlundt DG, et al. Southern community cohort study: establishing a cohort to investigate health disparities. *J Natl Med Assoc* 2005;97(7):972-9.
4. Prorok PC, Andriole GL, Bresalier RS, Buys SS, Chia D, Crawford ED, et al. Design of the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial. *Control Clin Trials* 2000;21(6 Suppl):273S-309S.
5. Calle EE, Rodriguez C, Jacobs EJ, Almon ML, Chao A, McCullough ML, et al. The American Cancer Society Cancer Prevention Study II Nutrition Cohort: rationale, study design, and baseline characteristics. *Cancer* 2002;94(2):500-11.
6. Kolonel LN, Henderson BE, Hankin JH, Nomura AM, Wilkens LR, Pike MC, et al. A multiethnic cohort in Hawaii and Los Angeles: baseline characteristics. *Am J Epidemiol* 2000;151(4):346-57.
7. Buring JE HC. The women's health study: Rationale and background. *J Myocardial Ischemia* 1992;4:30-40.
8. Final Report on the Aspirin Component of the Ongoing Physician's Health Study. *NEJM* 1989 July;321:129-35.
9. Christen WG, Gaziano JM, Hennekens CH. Design of Physicians' Health Study II--a randomized trial of beta-carotene, vitamins E and C, and multivitamins, in prevention of cancer, cardiovascular disease, and eye disease, and review of results of completed trials. *Ann Epidemiol* 2000;10(2):125-34.
10. Belanger CF, Hennekens CH, Rosner B, Speizer FE. The nurses' health study. *Am. J. Nurs.* 1978;78(6):1039-40.
11. Colditz GA, Hankinson SE. The Nurses' Health Study: lifestyle and health among women. *Nat Rev Cancer* 2005;5(5):388-96.
12. Willett WC, Sampson L, Stampfer MJ, Rosner B, Bain C, Witschi J, et al. Reproducibility and validity of a semiquantitative food frequency questionnaire. *Am J Epidemiol* 1985;122(1):51-65.

13. Wu K, Willett WC, Fuchs CS, Colditz GA, Giovannucci EL. Calcium intake and risk of colon cancer in women and men. *J Natl Cancer Inst* 2002;94(6):437-46.
14. Rimm EB, Giovannucci EL, Stampfer MJ, Colditz GA, Litin LB, Willett WC. Reproducibility and validity of an expanded self-administered semiquantitative food frequency questionnaire among male health professionals. *Am J Epidemiol* 1992;135(10):1114-26; discussion 27-36.
15. Salvini S, Hunter DJ, Sampson L, Stampfer MJ, Colditz GA, Rosner B, et al. Food-based validation of a dietary questionnaire: the effects of week-to-week variation in food consumption. *Int. J. Epidemiol.* 1989;18(4):858-67.
16. Feskanich D, Rimm EB, Giovannucci EL, Colditz GA, Stampfer MJ, Litin LB, et al. Reproducibility and validity of food intake measurements from a semiquantitative food frequency questionnaire. *J. Am. Diet. Assoc.* 1993;93:790-96.
17. Giovannucci E, Ascherio A, Rimm EB, Colditz GA, Stampfer MJ, Willett WC. Physical activity, obesity, and risk for colon cancer and adenoma in men. *Ann Intern Med* 1995;122(5):327-34.
18. Feskanich D, Rimm EB, Giovannucci EL, Colditz GA, Stampfer MJ, Litin LB, et al. Reproducibility and validity of food intake measurements from a semiquantitative food frequency questionnaire. *J Am Diet Assoc* 1993;93(7):790-6.
19. Giles GG, English DR. The Melbourne Collaborative Cohort Study. *IARC Sci Publ* 2002;156:69-70.
20. Berglund G, Elmstahl S, Janzon L, Larsson SA. The Malmo Diet and Cancer Study. Design and feasibility. *J Intern Med* 1993;233(1):45-51.
21. Hallmans G, Agren A, Johansson G, Johansson A, Stegmayr B, Jansson JH, et al. Cardiovascular disease and diabetes in the Northern Sweden Health and Disease Study Cohort - evaluation of risk factors and their interactions. *Scand J Public Health Suppl* 2003;61:18-24.
22. The alpha-tocopherol, beta-carotene lung cancer prevention study: design, methods, participant characteristics, and compliance. The ATBC Cancer Prevention Study Group. *Ann Epidemiol* 1994;4(1):1-10.
23. Shu XO, Li H, Yang G, Gao J, Cai H, Takata Y, et al. Cohort Profile: The Shanghai Men's Health Study. *Int J Epidemiol* 2015;44(3):810-8.
24. Zheng W, Chow WH, Yang G, Jin F, Rothman N, Blair A, et al. The Shanghai Women's Health Study: rationale, study design, and baseline characteristics. *Am J Epidemiol* 2005;162(11):1123-31.
25. Hankin JH, Stram DO, Arakawa K, Park S, Low SH, Lee HP, et al. Singapore Chinese Health Study: development, validation, and calibration of the quantitative food frequency questionnaire. *Nutr Cancer* 2001;39(2):187-95.
26. Yuan JM, Stram DO, Arakawa K, Lee HP, Yu MC. Dietary cryptoxanthin and reduced risk of lung cancer: the Singapore Chinese Health Study. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 2003;12(9):890-8.
27. Ross RK, Yuan JM, Yu MC, Wogan GN, Qian GS, Tu JT, et al. Urinary aflatoxin biomarkers and risk of hepatocellular carcinoma. *Lancet* 1992;339(8799):943-6.
28. Yuan JM, Ross RK, Wang XL, Gao YT, Henderson BE, Yu MC. Morbidity and mortality in relation to cigarette smoking in Shanghai, China. A prospective male cohort study. *Jama* 1996;275(21):1646-50.